Essays in Labor and Environmental Economics

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

im Frühjahrs-/Sommersemester 2021 vorgelegt von

Felix Holub

Abteilungssprecher	Prof. Volker Nocke, Ph.D.
Vorsitzende der Disputation	Prof. Michèle Tertilt, Ph.D.
Referent	Prof. Ulrich J. Wagner, Ph.D.
Koreferent	Prof. Dr. Sebastian Siegloch

Tag der Disputation	21. Juni 2021

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig angefertigt und die benutzten Hilfsmittel vollständig und deutlich angegeben habe.

Oviedo, am 27. April 2021

Felix Holub

Curriculum vitæ

2015–2021	University of Mannheim (Germany)
	Doctoral Program in Economics
2013–2015	CEMFI (Spain)
	Master in Finance and Economics
2009–2013	Maastricht University (Netherlands)
	Bachelor in Economics

Acknowledgements

I would like to express my gratitude to my advisors, Ulrich J. Wagner and Sebastian Siegloch, for accompanying me throughout my studies. They provided invaluable feedback, inspiration, and guidance while at the same time encouraging me to follow my curiosity.

I sincerely thank Michèle Tertilt and Andreas Gerster whose dedicated support was a great help for me.

I am thankful to my coauthors Moritz Drechsel-Grau, Laura Hospido, Hannah Klauber, Nicolas Koch, Nico Pestel, Nolan Ritter, Alexander Rohlf, Jose-Enrique Galdon-Sanchez, Ricard Gil, and Guillermo Uriz-Uharte. They taught me a lot about economic research.

My research also benefited from the feedback from Sebastian Findeisen, Andreas Gulyas, Stephen Kastoryano, Andreas Peichl, Christoph Rothe, André Stenzel, Michelle Sovinsky, Dimitri Szerman, Beate Thies, Rosa Ferrer, Florian Zimmermann, Jarone Gittens, Lars Nesheim, Aureo de Paula, Imran Rasul, and many others.

I am thankful for financial support from the University of Mannheim's Graduate School of Economic and Social Sciences (GESS) and the German Research Foundation (DFG) through CRC TR 224 (Project B07) and for the kind administrative support at the GESS, the Department of Economics, and the Chair of Quantitative Economics. Without the generous help of WIdO, the German Environmental Agency, the Spanish Social Security Administration, the anonymous manufacturing firm sharing its personnel data, and Fundación BBVA, not a single chapter of this dissertation could have been written.

I especially thank Moritz, not only for being an amazingly smart and optimistic coauthor of the first chapter, but also a great friend. Similarly, I am grateful to my friend Guille, in particular for motivating me during the harder days and being an encouraging mock audience.

My studies in Mannheim would have been much less fun without my fellow students. I thank Raphael, Fabian, Tomasz, Enrico, Nils, Beate, and all participants of the Environmental Economics Reading Group for making the Ph.D. such a unique experience.

I am grateful for being close to my family, who were a source of support, interest, warmth, and welcome distractions.

Lastly, and most importantly, I would like to thank Vera for being there for me during all the fun and tough days, for putting things into perspective, and for her love. Gracias!

Contents

Introduction

This dissertation studies questions in the fields of labor and environmental economics. It consists of four self-contained chapters that examine how seemingly small differences in the environments in which individuals work, live, shop, and grow up shape their pay, health, and consumption behavior. The dissertation covers two particular research areas. The first area examines the role that managers play in gender inequality. The second area concerns how air quality and its regulation affect the health of workers and young children, and the behavior of consumers. Chapter 1 contributes to the former of these research areas, Chapters 2, 3, and 4 to the latter.

Chapter 1, which is co-authored with Moritz Drechsel-Grau, sheds light on the role that bosses, and in particular male bosses, play in explaining gender gaps. We seek to answer the question of whether women are more likely than men to work for lower-paying bosses. Noting that the persistent under-representation of women in leadership could structurally disadvantage female workers, we also ask whether male bosses are bad for women

Another aspect of labor markets, labor supply, is studied in Chapter 2, which is joint work with Laura Hospido and Ulrich J. Wagner. We provide the first comprehensive causal estimates of how many days of work are lost due to air pollution. Our work also contributes first evidence on how the estimated pollution-sick leave gradients interact with features of the labor market.

Chapter 3 is joint with Jose-Enrique Galdon-Sanchez, Ricard Gil, and Guillermo Uriz-Uharte. It concerns Low Emission Zones, a common policy banning emission-intensive vehicles from entering city centers. These traffic restrictions are effective in improving air quality but they also increase transportation costs and could thereby redistribute economic activity. To examine this unintended consequence, we build on the case of Madrid and study how a Low Emission Zone affects consumer behavior.

While air quality management can distort economic activity, Chapter 4 stresses that important latent health benefits of reducing pollution are difficult to detect. In joint work with Hannah Klauber, Nicolas Koch, Nico Pestel, Nolan Ritter, and Alexander Rohlf, we argue that this is the case because health adjusts slowly over time and because health benefits are likely to take subtle forms. We use health insurance data to test whether children born just before and just after Low Emission Zone-induced air quality improvements exhibit persistent differences in medication usage.

In what follows, I briefly summarize each chapter.

1. Gender Gaps and the Role of Bosses

The first chapter investigates the contribution of managers to gender gaps and analyzes whether the over-representation of men in management positions puts women at a disadvantage. We separate out the factors explaining gender gaps relying on personnel data from one of the largest European manufacturing firms. These data contain detailed information on compensation, job descriptions, performance ratings, and workers' superiors. Adjusted pay gaps are positive, which means that men earn more than observationally equivalent women. Around 20% of the gender pay gap in base salaries and bonus pay can be explained by the differential sorting of men and women to managers. These decompositions also demonstrate that women receive better performance ratings than men, what challenges the view that the positive adjusted wage gap reflects productivity differences between genders. More importantly, gender gaps in bonus payments causally depend on the manager's gender. Accounting for worker and manager heterogeneity, bonus gaps are around 5% larger when the manager is male. This is driven by the fact that performance ratings are more favorable to men if handed out by a male manager. We present suggestive evidence that the relevance of manager gender for pay gaps is driven by discrimination rather than same-gender complementarities in productivity as manager gender tends to matter less when managers are supposedly more knowledgeable about a worker. However, independent of the root cause of these differences in evaluations by manager gender, the findings imply that a lower number of female managers increases gender gaps and thus constitutes a structural disadvantage for women.

2. Urban Air Pollution and Sick Leaves: Evidence from Social Security Data

In the second chapter, we estimate the causal impact of air pollution on the incidence of sick leaves in a representative panel of employees affiliated with the Spanish social security system. While the literature on the health effects of air pollution is extensive, little is known about the effect on daily labor supply and sub-clinical outcomes. Using over 100 million worker-by-week observations from the period 2005-2014, we estimate the relationship between the share of days an individual is on sick leave in a given week and exposure to particulate matter (PM_{10}) at the place of residence, controlling for weather, individual effects, and a wide range of time-by-location controls. We exploit quasi-experimental variation in PM_{10} that is due to Sahara dust advection in order to instrument for local PM_{10} concentrations. We estimate that the causal effect of PM_{10} on sick leaves is positive and varies with respect to worker and job characteristics. The effect is stronger for workers with pre-existing medical conditions and weaker for workers with low job security. Our estimates are instrumental for quantifying air pollution damages due to changes in labor supply. We estimate that improved ambient air quality in urban Spain between 2005 and 2014 saved at least € 505 million in foregone production by reducing worker absence by more than 5.58 million days.

3. Benefits and Costs of Driving Restriction Policies: The Impact of Madrid Central on Congestion, Pollution and Consumer Spending

The third chapter examines the economic impact of a popular air quality control policy, Low Emission Zones. Low Emission Zones are defined areas within a city that regulate traffic to reduce congestion and pollution. We empirically evaluate the effects of such a Low Emission Zone in Madrid, Spain, known as Madrid Central. Non-residents of the regulated area can access it only using electric or hybrid vehicles. While the effectiveness of such driving restrictions has been demonstrated in several settings, little is known about the indirect costs of these policies. Consumers and local commerce bear part of these indirect costs if transportation costs to the city center increase. This chapter sheds light on this aspect by examining the policy's impact on consumer behavior. First, using a difference-in-differences identification strategy, we find a decrease of 15% in both traffic and nitrogen dioxide levels. Second, we rely on a unique dataset on credit card transactions detailing spending for each pair of buyer-seller zip codes in

Madrid and its surroundings in the period 2015-2019 to analyze how the Low Emission Zone changed consumption behavior. Although we find no significant effect on overall consumption spending inside the regulated area, our findings show that consumers affected by the regulation partially substitute their consumption spending from brick-and-mortar to online shopping. This finding suggests that e-commerce may smooth the impact of changes in transportation costs due to environmental regulations.

4. Killing Prescriptions Softly: Low Emission Zones and Child Health from Birth to School

The fourth chapter examines the impact of early-life exposure to air pollution on children's health from their in utero period to school enrollment. There is ample evidence for particularly severe impacts of contemporaneous air pollution on infant health. Also, persistent effects of early-life exposure to air pollution on human capital and labor market outcomes have previously been found. In contrast, we present evidence for the persistence of health benefits from exposure to cleaner air in a fixed period, the time from conception until age one. We use administrative public health insurance records covering one third of the population of children in Germany. For identification, we exploit air quality improvements caused by the implementation of Low Emission Zones across German counties. Our results indicate that children born just before and just after the policy adoption exhibit persistent differences in medication usage for at least five years after treatment. We document that a slight improvement in air quality in a single year reduces spending for respiratory medication for children born between 2008 and 2017 by about € 26.5 million over their pre-school years. The initially latent health response materializes only gradually in lower medication usage, leaving important but subtle health benefits undetected in studies of contemporaneous health.

Chapter 1

Gender Gaps and the Role of Bosses

Joint with Moritz Drechsel-Grau

1.1 Introduction

While the gender wage gap has declined considerably, convergence has slowed down and substantial gender disparities persist. In the US, for example, the unadjusted gender gap has stagnated at around 19% since the turn of the century (BLS, 2019). The adjusted wage gap is even more persistent as, for the past three decades, women have been earning about 9% less than men after adjusting for differences in education, experience, industry and occupation (Blau and Kahn, 2017). Gender pay gaps thus continue to receive significant attention as policy makers discuss gender quotas and many firms and large organizations train managers to be more aware of gender-related biases (Chang et al., 2019a). It has been suggested that the adjusted wage gap is driven by differences in productivity, negotiation prowess, temporal flexibility, or (unconscious) discrimination (e.g. Azmat and Ferrer, 2017; Babcock et al., 2003; Goldin, 2014; Sarsons, 2018, respectively). All of these explanations could be closely related to the behavior of bosses. The direct superiors of workers affect productivity, negotiate salaries, shape work environments, and evaluate performance and thereby determine bonus payments (e.g. Lazear et al., 2015; Hoffman and Tadelis, forthcoming; Frederiksen et al., 2019).

In this chapter, we use novel personnel data of a large multinational firm in order to shed light on the role of managers for gender gaps. In particular, we ask two novel questions connecting the literature on managers to that on gender gaps. First, do women work for “worse” bosses than men? Similarly to the distribution across occupations, sorting of workers and managers may explain a part of the gender wage gap. In other words, if male workers in the same occupation work for better-paying bosses than their female colleagues, this will drive a wedge between the earnings of equally qualified men and women. Second, are male bosses bad for women? Three main mechanisms come to mind. First, men may undervalue the performance of women due to conscious or unconscious biases. Second, male managers may create work environments that make it harder for women to shine. Third, men might be more productive under male managers due to gender-specific complementarities in productivity.[1] The persistent under-representation of women in powerful positions may therefore be cause and consequence at the same time if—for some reason—female employees are systematically disadvantaged when having male superiors.

In order to separate out the different factors explaining gender disparities and investigate the importance of (male) managers for gender gaps, we bring in unique personnel data provided by one of the largest European manufacturing firms. The panel dataset cover the multinational’s entire workforce in the period 2014–2019 and has several key advantages allowing us to address these questions. First, the data contain detailed information on job characteristics, sociodemographics, compensation, and performance evaluations. This allows us to identify the performance-related component of earnings. Second, we are able to trace out the organizational hierarchy and identify every employee’s coworkers, superiors, and subordinates. Third, we can condition on time-invariant unobservable characteristics of both employees and their managers.[2]

The chapter has two sets of results. The first quantifies to which extent the sorting of male and female workers to different managers can explain gender gaps while also taking into account the contributions of other observables such as sociodemograph-

[1]The opposite could also be true if women hinder other women, as for example has been documented by Bagues and Esteve-Volart (2010)

[2]We will refer to a worker’s direct superior as manager. A manager is also a worker from the perspective of his or her manager.

ics and job characteristics. To that end, we implement a Kitagawa-Oaxaca-Blinder-decomposition for base salaries, bonus payouts, contracted bonus targets, and performance ratings (Kitagawa, 1955; Oaxaca, 1973; Blinder, 1973). For male and female workers, we run separate regressions of the outcome of interest on job characteristics, manager indicators, age, tenure, and location controls. The decomposition reveals the following three findings.

First, the raw gender gaps in base salary and bonus payouts are 12.3 log points and 22.2 log points, respectively. Men's contracted bonus targets are on average 2.8% greater than those of women. The raw gender gap in performance ratings is negative as men are two percentage points less likely than women to receive a high performance rating.

Second, 25% of the raw gender gap in base salary and 19% of the gender bonus gap are attributed to the sorting of male and female workers to different managers. The unexplained component of the gender gap is larger for bonuses (16.9%) than for base salaries (8.8%). For performance ratings, we find that a large part of the gender gap cannot be explained. However, while the contribution of standard observables such as age or job characteristics drops substantially, the impact of managers remains sizable. Worker-manager sorting increases the performance gap (in favor of men) by 2.1 log points. This means that if women were assigned to the same jobs and managers, the performance gap would be even more negative.

Third, comparing similar employees doing the same job under the same manager, we find significant residual gender gaps in base salaries (1.1%), bonus targets (2.0%), and bonus payouts (3.8%). While all of these gaps favor men, the opposite holds for performance evaluations. Women are 3.5 percentage points (14.2% relative to the mean) *more* likely to receive high a performance rating, which implies above-target bonus payouts. The performance-corrected gap in bonus payouts is thus even larger than the raw gap. We find no evidence that women simply receive better ratings because they would cost the firm less in terms of implied bonus payments. We do not find evidence in favor of the interpretation that residual pay gaps are majorly related to child care obligations.

The second set of results answers the question whether the gender gap is different under male and female managers. Intuitively, we do so by defining within-manager gender gaps and comparing their average sizes between male and female managers in a difference-in-differences framework. As we take into account unobserved worker heterogeneity, identification comes from workers working for managers of different genders over time. As the worker fixed effects fully absorb the absolute level of the gender gaps, we identify by how much the expected gender gap *changes* when the manager is male rather than female.

We find that the over-representation of male managers implies a structural disadvantage for women. In particular, male managers cause the gender gap in bonus payouts to increase by 5.1%. This is driven by a relative increase of the gender gap in performance evaluations of 2.7 percentage points comparing male to female bosses. Hence, while in general men receive lower ratings, this gap closes considerably when the manager is male.

We evaluate which mechanism is likely to drive these findings. While more productive men could work more often for male bosses, this mechanism cannot rationalize our findings as we control for unobserved worker characteristics. Another potential explanation are within-gender complementarities. However, gender gaps do not increase with the share of male coworkers in a team. Assuming that potential within-gender complementarities would also exist among coworkers, we can rule out that the productivity channel drives the results. In contrast, we find suggestive evidence consistent with (unconscious) discrimination, as manager gender tends to matter more for less knowledgeable decision makers. In particular, we split managers into groups who should be more or less informed about the true quality or the needs of their subordinates. While not statistically significant, we find that for all proxies of manager knowledge the effect of manager gender is smaller. The observation that a manager's experience, team size, spatial proximity and the time that a manager has worked with a subordinate all are correlated with a lower effect of manager gender is consistent with discrimination due to biased beliefs of less knowledgeable managers. This mechanism relates to Bohren et al. (2019), who show that decision makers resort to their biased beliefs if little information about workers is available to them.

Related literature This chapter primarily contributes to the vast literature on gender inequality in the labor market summarized among others by Altonji and Blank (1999), or more recently by Bertrand (2011) and Blau and Kahn (2017). One set of papers tries to understand (raw) gender pay gaps. Early papers focused on the role of education and human capital (Altonji and Blank, 1999). As the gender gap in human capital has vanished over time, recent studies have highlighted the role of children as well as differences in occupation and industry. Kleven et al. (2019b) use Danish data to show that the arrival of children creates a substantial long-run gender gap in earnings driven by hours worked, participation, and wage rates (see also Kleven et al., 2019a, for evidence on other countries). Blau and Kahn (2017) document that differences in occupation continue to account for parts of the gender wage gap, and Goldin (2014) finds that work environments rewarding working long hours prevent female wages from fully catching up. Based on an AKM-model (Abowd et al., 1999) in which workers are sorted to firms, Card et al. (2016) find that part of the gender wage gap can be attributed to women working for firms that pay lower premiums. We add to the literature on raw gender wage gaps by showing that the sorting of men and women to different managers in part explains the gender wage gap. To our knowledge, our work is the first to focus on the impact of worker-manager sorting on pay gaps. We bring in a data source—a large manufacturing firm's personnel data—that allows for better control over job characteristics. Research on gender gaps using personnel data dates back to Malkiel and Malkiel (1973) and is vast.[3] However, our data are unique as they identify hierarchical relations between workers and managers while also containing highly detailed information about pay, performance ratings, ranks, and occupations

A second strand of the literature investigates the reasons behind the persistence of the adjusted gender wage gap, in particular with a focus on performance and evaluations. Previous studies come from very specific settings and may therefore lead to different conclusions. In the context of academia—where performance is relatively easy to measure—Sarsons et al. (forthcoming) show that female researchers get less credit for joint work than male co-authors. Card et al. (2020) conclude that journal editors and referees hand out too few revise-and-resubmit decisions to female-authored papers relative to a citation-maximizing benchmark. Similarly, Hospido and Sanz (2019) find that all-female authored papers are less likely to be accepted for major economics conferences. Outside of academia, female surgeons and financial advisors have been found to be more heavily penalized for bad performances or misconduct (Sarsons, 2018;

[3] For example, Sorensen (1986), Kahn (1992), Ransom and Oaxaca (2005), Barnet-Verzat and Wolff (2008), Dohmen et al. (2008), Ichino and Moretti (2009), Pema and Mehay (2010) or Pekkarinen and Vartiainen (2016) study within-organization gender gaps, mainly based on public sector data.

Egan et al., 2017, respectively). Mengel et al. (2019) show that female university tutors receive systematically lower teaching evaluations. Azmat and Ferrer (2017) find that gender performance differences exist as male lawyers actually outperform their female colleagues and that accounting for this substantially alters the interpretation of the gender wage gap. What distinguishes our work is that we observe wages, bonus payouts and performance evaluations in the context of a large multinational enterprise in the manufacturing sector, i.e. a setting that is highly relevant to many workers in developed economies. In this context, we find that the over-representation of men in top positions does harm women as gender gaps in bonus payments and performance ratings increase substantially under male managers. In addition, our findings challenge the view that productivity differences can account for adjusted gender wage gaps in a wide range of occupations. We further provide suggestive evidence in favor of the biased-beliefs mechanism proposed by Bohren et al. (2019).

We also add to the literature on the effect of male leadership on gender gaps. While the gender composition at the very top of firms does not affect gender gaps (Bertrand et al., 2019; Maida and Weber, 2019), a number of papers show that gender compositions matter when the distance between superior and subordinate is smaller. Kunze and Miller (2017) and Kurtulus and Tomaskovic-Devey (2011) find that a larger share of women at higher ranks increases women's chances of being promoted. However, it is unclear whether individual interactions between workers and managers or firm-wide policies drive these observations. This work differs in that we *directly* link workers to their managers at all levels of the firm hierarchy rather than only at the very top. Using cross-sectional survey data, several authors have documented that gender gaps tend to be greater under male superiors (Ragan and Tremblay, 1988; Rothstein, 1997; Abendroth et al., 2017).[4] In the context of schools, Biasi and Sarsons (2020) find that gender pay gaps among teachers increase when principals or superintendents are male. A recent study by Cullen and Perez-Truglia (2019) shows that the gender promotion gap in a Southeast Asian bank widens when the direct superior is male. Relative to their study, we focus on how manager gender impacts within-team gaps in wages and performance evaluations in a wide range of occupations and countries. Our approach

[4]Another cross-sectional study by Halldén et al. (2018), based on Swedish survey data, finds that women earn less when their superior is female. However, the data do not allow to make claims about the gender gap.

also takes into account unobserved manager characteristics, which prove critical to the finding that managers affect gender pay gaps.

Outline The remainder of the chapter is structured as follows. Section 1.2 provides more information about the data and the firm. Section 1.3 describes our empirical strategy, Section 1.4 documents the findings, and Section 1.5 concludes.

1.2 Data and setting

The firm We use the personnel data of a large firm in the manufacturing sector.[5] The firm is among the 250 largest European firms in terms of sales and employment and an industry leader in an R&D-intensive sector. A quarter of the multinational's workforce is located in the firm's home country but it has establishments in over 50 countries. For example, around 20% of the workforce is located in the United States.

While a single firm can hardly be representative of the economy as a whole, its size, international representation, range of occupations, and diversity of skills required ensure that its internal labor market is typical for what a worker would encounter at any large firm. Key variables of our study are earnings and the share of male workers, which we can use to compare the firm to other firms in the same sector. In the US, workers in the same three-digit NAICS industry earned during the sample period around 7.2% less than workers in our data.[6] At the firm, the share of female workers in the US is about 6 percentage points higher than the sectoral average.[7]

By now, administrative data matching employees to firms are widely available. While such kind of data would be preferable to use for a more holistic view, our dataset has several unique features. As opposed to administrative data where workers would be linked by working in the same firm and same occupation, we observe precisely who works in the same team and who is each worker's responsible superior. Furthermore, the description of jobs and hierarchy provided to us by the firm goes beyond typical

[5] We are not allowed to reveal the identity of the firm.

[6] Bureau of Economic Analysis, Wages and Salaries Per Full-Time Equivalent Employee by Industry, https://apps.bea.gov/iTable/iTable.cfm?reqid=13&step=3&isuri=1&nipa_table_list=201&keyword_index=w

[7] Bureau of Labor Statistics, Labor Force Statistics from the Current Population Survey, https://www.bls.gov/cps/cpsaat18.htm

definitions of occupations. This allows us to control much more precisely for the nature of the job. We also do not only observe earnings but the detailed variables determining compensation, including performance ratings.

Personnel data We were provided with an anonymized monthly panel of all personnel records between January 2014 and March 2019. The data includes information on employees' compensation (base salaries and bonus payments), performance ratings, occupation, hierarchical rank, location, tenure, and some sociodemographic characteristics such as gender, age, or nationality.[8] Importantly, the data also indicate the identity of each worker's superior, to which we refer as *manager* or *boss*. Therefore, we can trace out the organizational hierarchy and identify employees' coworkers, superiors, and subordinates. Employees can be superiors and subordinates at the same time as the firm has many hierarchical layers.

We observe ten thousands of full-time employees.[9] The unadjusted gender gap (the average difference between male and female outcomes) in base salaries (before taxes) is 3917€, or 7.6%. One would typically control for experience, location, and job characteristics to determine the adjusted gender pay gap. Such an adjustment yields a gender gap in base salaries of 1.7%. 26 404 workers are paid a bonus. In this set of workers, the gender gap in annual bonus payout (before taxes) is 2375€, or 20.3%. Controlling for experience, location, and job characteristics results in an adjusted bonus gap of 4.6%.

Job definitions Jobs are classified into 109 different occupations, for example *Electrical Engineering*, *Scientific Technical Assistance*, *Fire Brigade*, or *Web Design*.[10] In addition, jobs are classified by hierarchical rank on a scale from one to ten, representing a range from low-skilled helpers to executives. Ranks one to three are to a large extent blue-collar production jobs. While workers of different hierarchical rank can work in the same occupation, occupations only comprise a limited number of ranks. Starting at rank four, white-collar occupations are more prevalent. We henceforth characterize a job by the combination of occupation and hierarchical rank.

[8] The number of employees reporting education and previous employers is very small.

[9] We cannot report the exact number to protect the identity of the firm.

[10] We cannot show a full list of occupations as it could reveal the identity of the firm.

Table 1.1: Gender Distribution across Hierarchical Ranks

Rank	Share [%]	Salary [€]	Bonus [€]	Any Bonus [%]	Share Male [%]	Salary Gap [€]	Bonus Gap [€]
1	0.5	18 002.58	13 036.89	33.67	63.85	−5324.59	5792.14
2	7.65	27 131.86	4007.53	53.26	66.3	5958.01	2095.94
3	14.76	32 181.96	2479.71	54.02	67.62	2395.46	454.9
4	25.54	32 771.42	3002.1	47.4	57.1	−860.15	−162.59
5	27.79	44 937.26	5931.84	44.28	57.01	654.7	138.78
6	17.25	76 957.14	14 907.22	71.98	61.8	3002.29	131.61
7	3.71	117 899.9	31 514.63	80.15	68.5	−8527.77	−135.85
8	2.41	154 211.52	57 869.84	81.39	75.94	−5702.43	−1890.47
9	0.34	227 647.17	137 310.0	74.71	87.77	−16 322.27	−24 453.82
10	0.05	335 444.2	276 739.82	69.79	72.92	59 631.59	17 396.89
All	100.0	50 090.26	10 781.41	54.27	61.16	3917.02	2374.62

Notes: The table displays mean values of base salaries and bonus payouts in € with base year 2010 for nine out of ten hierarchical ranks (1 being lowest and 10 highest). The table also displays the share of workers, the share of men, and the gender gaps (male outcome - female outcome) in base salaries and bonus payouts.

From Table 1.1 we can see that the majority of workers works on jobs of intermediate hierarchical rank. In general, higher ranks pay higher salaries and bonuses, but not all workers receive bonuses. While participating in a bonus program seems to be more common at high ranks, a number of workers opts out. At very high ranks, other long term incentives also play a greater role. The share of men at low ranks is relatively large as many production-related occupations are ranked here. Women represent around 40% of the workforce in the middle range of the hierarchy. But the higher the rank, the lower becomes the share of women. While there is no clear pattern of raw gender pay gaps at separate hierarchical levels, it seems that men face an advantage at the most common ranks in the firm.

Salaries, bonus payments and performance ratings According to the firm, conditional on individual productivity, experience, and location, employees working in the same job should be compensated equivalently. It might well be the case that a certain occupation pays more at a lower rank than another occupation at a higher rank. Each job has a salary band set by the human resource department which is only known by the worker's manager and not part of the data we obtain. Managers and workers negotiate a base salary within such a band. Usually, pay raises within the same job are only negotiated once a year, namely when an employee's performance is evaluated.

The contract also specifies a bonus target, which is the amount that will be paid out in addition to the base salary as an annual bonus. The target is expressed as a percentage of the base salary. For example, if the employee's target is 5% and the base salary is 50 000€, she can expect annual earnings of 50 000€ + 2500€ = 52 500€.

However, the bonus is supposed to incentivize effort. Workers showing satisfactory but not outstanding performance are paid as just described. Less is paid out as a bonus if the employee performs poorly, and more is paid out if the employee does especially well. This means that workers with a higher base salary, higher bonus target, or higher performance will receive a larger annual bonus. Workers' effort has a significant influence on the amount eventually paid out as a bonus. Workers know the function $f(\text{rating})$, which maps their grading into a factor multiplying base salary and bonus target. In principle, earnings can be expressed as the sum of base salary and the performance-dependent component, i.e. the product of base salary, target and a function of performance.

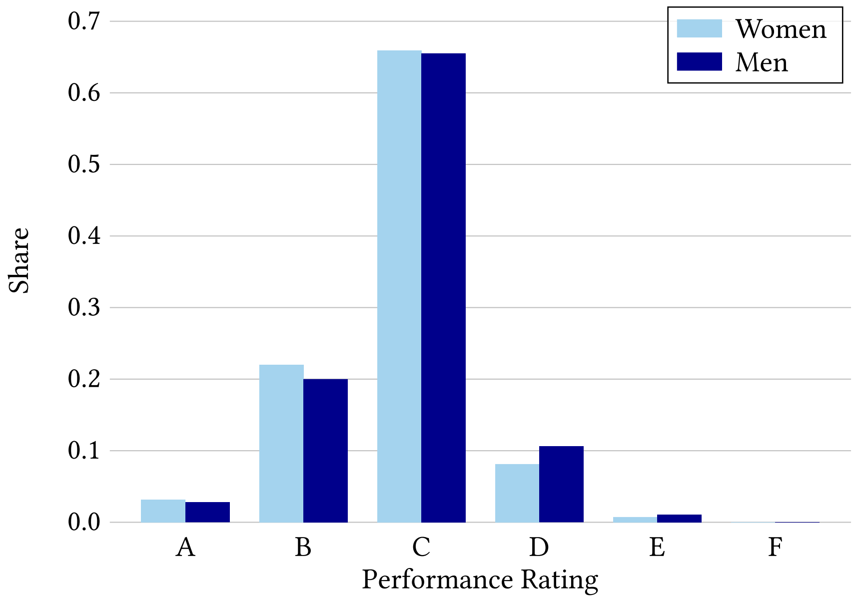
$$\text{Earnings} \approx \text{Base Salary} + \text{Base Salary} \times \text{Target} \times f(\text{rating})$$

Performance ratings are handed out by the employee's direct superior once per year, evaluating the previous twelve months. An evaluation scheme of six grades is applied across all jobs and countries. The firm considers a ranking to be high if an employee achieves at least the second-best grade. A *high* ranking will ceteris paribus result in a bonus payout greater than the contracted target. As an employee's output is hard to measure, the mapping from effort to performance ratings cannot be contracted.[11] Hence, performance ratings are potentially subject to conscious or unconscious gender biases of the manager (e.g. Bordalo et al., 2019). Through their performance evaluations, managers thus have a substantial impact on the total earnings of their subordinates.

Figure 1.2.1 displays the distribution of performance ratings for men and women. The majority of workers receives a *C*, which implies no adjustment to the contracted bonus. A high performance rating is achieved by workers receiving ratings *A* or *B*. The graph also shows that in the raw data women are more likely than men to receive a high rating.

[11] Employees for which individual output can easily be measured are sales workers. They receive a separate sales bonus in addition to the general bonus which reflects the generated revenue. We do not have access to these data.

FIGURE 1.2.1: Distribution of Ratings by Gender



Note: The figure plots a histogram of annual performance ratings received by male and female workers. The best rating is *A*, the worst rating is *F*.

Observed bonuses are not equivalent to Base Salary × Target × f(rating). There are several reasons for this. First, department-wide achievements also affect bonus payout. Second, we are not provided with the mapping for workers with changes in workplace characteristics. Third, performance ratings are very detailed, while we only use a simple approximation, an indicator for receiving a high rating. Importantly, it holds that all else equal, workers with high performance should receive a larger bonus.

Sample description We focus on full-time employees aged 25 to 60 and for which we observe gender, superior, job, age, and tenure. We aggregate the data to annual frequency for 2014-2019 as bonus payments and performance ratings are only determined once per year. Monetary variables are converted to Euros with 2010 as the base year. The resulting dataset is summarized in Table 1.2.[12]

Base salaries are on average 50 000€. We observe positive bonus payouts only for a subset of workers because not every worker receives performance pay and due to

[12]Extreme values are omitted for confidentiality reasons.

Table 1.2: Descriptive Statistics

	Mean	SD	10th Perc.	Median	90th Perc.	Observations
Salary [€]	50 090.26	53 096.49	10 077.67	40 404.73	94 234.79	178 377
Bonus [€]	10 781.41	20 324.33	1289.45	4233.43	24 812.78	96 803
High Performance	0.22	0.42	0.0	0.0	1.0	154 790
Low Performance	0.13	0.33	0.0	0.0	1.0	154 790
Bonus Target [%]	11.82	8.81	4.0	10.0	22.51	108 839
Male	0.61	0.49	0.0	1.0	1.0	178 377
Age	41.53	9.24	29.0	41.0	55.0	178 377
Tenure	9.82	9.19	1.0	7.0	24.0	178 377
Span of Control	1.16	3.77	0.0	0.0	5.0	178 377
Coworkers	9.65	20.59	1.0	5.0	19.0	178 377
New Manager in Same Job	0.25	0.44	0.0	0.0	1.0	128 521
New Manager in New Job	0.07	0.25	0.0	0.0	0.0	128 521
Male Manager	0.72	0.45	0.0	1.0	1.0	178 377
Male & Male Manager	0.49	0.5	0.0	0.0	1.0	178 377
Male & Female Manager	0.12	0.33	0.0	0.0	1.0	178 377
Female & Male Manager	0.23	0.42	0.0	0.0	1.0	178 377
Female & Female Manager	0.16	0.36	0.0	0.0	1.0	178 377
Age of Manager	44.97	8.08	34.0	45.0	56.0	178 377

Notes: Extreme values omitted for confidentiality reasons. Unbalanced panel based on ten thousands of workers and the years 2014-2019. Monetary variables normalized to Euros in 2010.

data limitations. The size of bonus payments is significant with a mean annual payout of almost 11 000€. Approximately 20% of workers receive a high performance rating, so are entitled to a bonus payout greater above their contracted bonus target. 13% of workers underperform, implying a below-target bonus payout. Some workers receive a rating even though their contracts do not include performance pay. 60% of employees are male and the average age is 42 years. The span of control measures the number of direct subordinates.

Identifying each worker's manager is key to studying managers' impact on gender gap. We do so by matching each worker to his or her direct superior. Each employee has on average 9.7 coworkers who work under the same manager. Employees are more likely to have a change in their manager due to managers rotating than due to a job change of the worker. One quarter of employees stay in their current job but work for a new manager. Managers are more likely to be male and on average three years older than workers.

The data reveal that male and female workers are sorted to different managers based on gender. The share of male workers sorted to a male manager is 80%. Women are more likely to work for female managers as the share of female workers sorted to a male manager is only 59%.

1.3 Research design

In the first part of the empirical analysis, we examine how the sorting of workers to managers can explain gender gaps. Evidence for the impact of managers on worker productivity has been provided by Frederiksen et al. (2019) or Lazear et al. (2015). For example, women may work for managers who are less productive. This could happen if male workers have stronger social networks within the firm which provide them with better information about managers. In Table 1.2 we saw that workers are sorted to managers based on gender. If female managers are less productive on average (in line with the gender gaps in performance found by Azmat and Ferrer, 2017) and manager productivity affects worker productivity, female workers are disadvantaged.

There are other explanations why managers could matter. Male employees are often less reluctant to negotiate (e.g. Babcock et al., 2003). This could drive them towards more generous managers who are open to negotiation. Similarly, the fact that women tend to shy away from competition (Niederle and Vesterlund, 2007) could drive female workers to work more often for managers who create less competitive work environments, with the effect that workers are on average less productive. Women might be also driven to managers who—at the cost of lost productivity—offer a more family friendly environment, for example by allowing for more flexible work hours or permitting working from home on a regular basis (Goldin and Katz, 2015). This explanation would imply that women might actually prefer to work for "worse" bosses, i.e. bosses creating a environment which makes workers less productive.

In the second part of the analysis, we hold these manager effects fixed and study whether there is evidence that managers affect their *within-team pay* gaps. Observing any residual gender gap within teams does not necessarily imply that managers are to blame for gender pay gaps. But if within-team gaps vary across managers with different characteristics, we can conclude that managers do affect workplace equity.

1.3.1 Explaining gender gaps

We aim to answer the question what portion of gender gaps can be explained by observable characteristics. In particular, we want to quantify the contribution of the matching of managers and workers to gender gaps. To do so, we implement traditional Kitagawa-Oaxaca-Blinder-decompositions of differences between male and female workers in log salaries, log bonus payouts, high performance indicators, and contracted bonus targets (Kitagawa, 1955; Oaxaca, 1973; Blinder, 1973).

The decomposition classifies differences between two groups into a composition component that accounts for different characteristics, e.g. tenure or occupation, and an unexplained, or wage structure, component. Such Kitagawa-Oaxaca-Blinder-decompositions are commonly used in the estimation of gender gaps (e.g. Bertrand et al., 2010; Blau and Kahn, 2017; Card et al., 2016; Juhn and McCue, 2017).

The unexplained component is often interpreted as a measure of discrimination, as it implies differences in pay or other outcomes for observationally identical workers. However, discrimination might also stem from different characteristics, i.e. the explained part of the gender gap. Women might be discriminated against by being allocated to worse-paying occupations. By controlling for the type of occupation, we take this allocation as given while it could already be the result of discriminatory treatment. It is also not clear that all unexplained differences are the result of discrimination. If men perform better than women (as found by Azmat and Ferrer, 2017), productivity differences contribute to the residual term.

The Kitagawa-Oaxaca-Blinder-decomposition is implemented as follows. Each worker i is either male m or female f and observed in year t. We estimate ordinary least squares (OLS) regressions for both genders separately.

$$Y_{it}^m = X_{it}^m \beta^m + u_{it}^m \tag{1.1}$$

$$Y_{it}^f = X_{it}^f \beta^f + u_{it}^f \tag{1.2}$$

Y_{it} is the outcome of interest of worker i in year t. X_{it} is a vector of variables observed at the worker level, including a constant. β^m and β^f are the gender-specific returns to these characteristics. u_{it} is the error term.

We obtain the estimates of returns $\widehat{\beta}^m$ and $\widehat{\beta}^f$ from OLS. We then calculate means of the outcomes and characteristics for both genders, denoted by a bar over the respective term. The difference of the means of Equations (1.1) and (1.2) is the observed gender gap.

$$\overline{Y}^m - \overline{Y}^f = \overline{X}^m \widehat{\beta}^m - \overline{X}^f \widehat{\beta}^f = (\overline{X}^m - \overline{X}^f)\widehat{\beta}^m + \overline{X}^f (\widehat{\beta}^m - \widehat{\beta}^f) \tag{1.3}$$

The residuals drop out when taking the mean. The first term of the decomposition in Equation (1.3) is the difference in male and female outcomes due to different characteristics, based on male coefficients. The second term is the unexplained difference in outcomes due to different returns for men and women. One could perform the decomposition using the coefficients on female returns as well. However, here we are interested in how the outcomes of women would change if the firm is required to treat women like men.

We decompose log base salaries, log bonus payouts, high performance indicators, and log bonus targets. Due to the richness of the firm's personnel data, we can include a wide range of variables in X_{it}. For age and tenure, we create bins for every six years of age. We include indicators for each year and each country. Job characteristics are controlled for by including the combination of occupation and hierarchical rank. As we are particularly interested in the role of bosses, X_{it} also includes an indicator for worker i's boss at time t.

The high resolution of worker-level controls comes at a cost. The matrices X^m and X^f, which respectively collect the vectors $X_{i,t}^m$ and $X_{i,t}^f$, need to have full rank. This is not the case if, for example, there is a certain job which is only done by men. In such a case the column of the matrix X^f indicating working for this manager would always be zero and hence perfectly correlated with the column indicating the constant. We therefore only include observations with a characteristic observed among men and women.

Another requirement for full rank matrices is that the different categorical variables are connected. This is identical to the condition explained by Abowd et al. (2002), that fixed effects in an AKM-model (Abowd, Kramarz, and Margolis, 1999) are only identified within a connected set. A set of observations is unconnected if a categorical variable is nested in another categorical variable. Consider an example where we include a constant, occupation, and location. Assume that engineers and accountants always work in Spain, cleaners always work in France, and there are no other occupations or countries. When comparing a worker in France to a worker in Spain it is unclear whether their pay difference is due to location or occupation. But if accountants work in both countries, they identify differences due to location. Once these differences are determined, residual differences in pay can be attributed to occupations. Note that even in this case we still require a linear restriction due to the inclusion of a constant.

For this reason, we only keep the largest connected set of categorical variables for men and women. In practice, we iterate over the algorithm proposed by Abowd et al. (2002) for different combinations of categorical variables until the matrices X^m and X^f are full rank and only include observations with a characteristic observed both among men and women.

Assuring that matrices are full rank reduces the sample size. To maintain a constant size of data, we also impose that workers participate in bonus schemes and that bonus targets and performance ratings are observed. The resulting dataset consists of 59 813 observations based on 20 048 workers and 4327 managers from 58 countries and 421 jobs. Appendix Table 1.A.1 shows that in this dataset workers earn a bit more which is due to the fact that some low-rank jobs without performance pay are excluded. Besides that, the subset of data used for estimation is very similar.

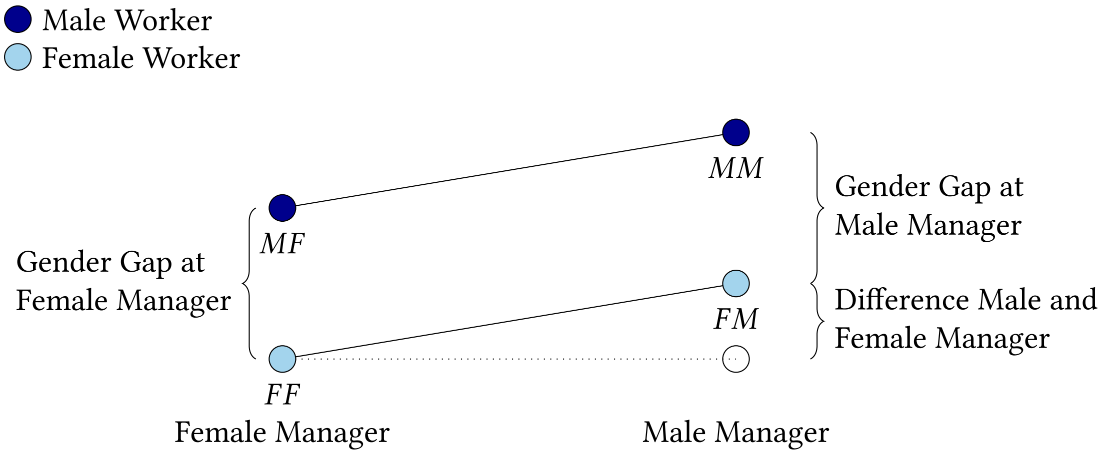
1.3.2 Understanding residual gender gaps

Residual gender gaps could be interpreted as productivity differentials between men or women. In the decomposition, managers can contribute to gender gaps through the channel that male and female workers work for different managers. But it does not tell us anything about how individual managers treat men and women who actually work for them. For example, these residual gaps could persist if the majority of bosses for some reason favor male workers. At the firm, 72% of managers are male.

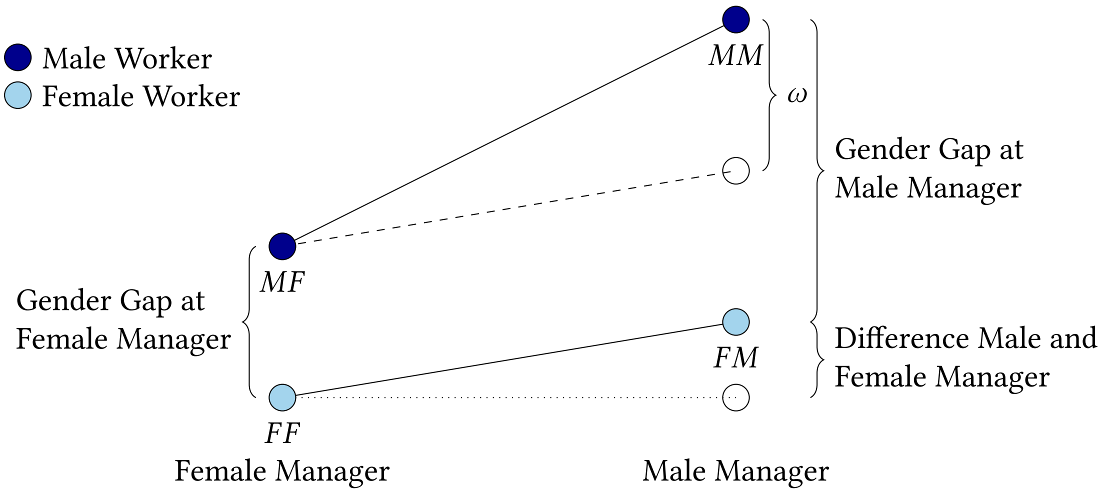
Illustration of within-manager-gender gender gaps Figure 1.3.1a is an illustration of how gender gaps could look like, separating gaps at male and female managers. The light dots indicate the earnings of women, the dark dots earnings of men. The filled dots on the left plot earnings at female managers, the filled dots on the right at male managers. The labels MF, FF, etc. also refer to the four possible combinations of worker and manager gender. For example, MF stands for male workers working for female managers. Focusing on female managers only, the difference between MF and FF is the gender pay gap. Workers at male bosses earn more ($FM - FF$), but the size of the gender gap $MM - FM$ is equal to the gap under female managers. The gender pay gap $MF - FF$ is a quantity of interest, but it is unclear why this difference exists.

FIGURE 1.3.1: Illustration of Gender Gaps

(A) Equal Gender Gaps at Male and Female Managers



(B) Different Gender Gaps at Male and Female Managers



Notes: These figure provides a graphical illustration of gender gaps under male and female managers. FF stands for a female worker working for a female manager, MF for a male worker working for a female manager, FM for a female worker working for a male manager, and MM for a male worker working for a male manager. ω is the difference in gender gaps between male and female managers, i.e. $\omega = (MM - FM) - (MF - FF)$.

In the example from Figure 1.3.1a, all managers could be discriminating against women, or women could be less productive. In such a setting we cannot draw any conclusions on the impact of managers on gender gaps.

But if we find that there is variation in gender gaps across teams managed by managers of different gender, we know that manager gender and hence managers affect gender gaps. This is depicted in Figure 1.3.1b. In this example, the gender gap at male bosses is greater than the gender gap at female bosses. If the difference in gender gaps, ω, is significant, we can conclude that managers affect gender gaps.

Estimation To estimate the change in the gender gap depicted in Figure 1.3.1b, we run a difference-in-difference estimation for outcome Y of the following form.

$$Y_{it} = \gamma_0 + \gamma_1 \times male_i + \gamma_2 \times male_{M(i,t)} + \omega \times male_i \times male_{M(i,t)} + X_{it}\beta + \epsilon_{it} \qquad (1.4)$$

$male_i$ is a dummy taking value 1 if worker i is male. Its coefficient represents the gender gap under female managers, i.e. $MF - FF$ in Figure 1.3.1b. $male_{M(i,t)}$ is a dummy taking value 1 if the manager M of worker i at time t is male. Its coefficient can be interpreted as the difference in earnings among women when working for a male instead of a female manager, i.e. $FM - FF$ in Figure 1.3.1b. $X_{i,t}$ controls for age-bin, tenure-bin, year, country, and job as before. We are interested in the difference-in-differences coefficient ω on the interaction of the dummy $male_i$ with the dummy $male_{M(i,t)}$. This product is 1 if worker and manager are male and 0 otherwise. As shown in Figure 1.3.1b, ω is the difference in gender gaps between male and female managers. A positive ω indicates that the gender gap moves in favor of men when the manager is male.

Replacing all gender dummies by $female_i = 1 - male_i$ etc., results in exactly the same estimate of ω, but now on the product of dummies $female_i \times female_{M(i,t)}$.[13] The natural interpretation would be as follows. The inverse gender gap (female outcome - male outcome) increases if the manager is female. This shows that we only can quantify

[13]This can be shown by a simple replacement of variables:

$$\begin{aligned}
&\gamma_0 + \gamma_1\, male_i + \gamma_2\, male_{M(i,t)} + \omega\, male_i \times male_{M(i,t)} + \dots \\
&\quad = \gamma_0 + \gamma_1\, (1 - female_i) + \gamma_2\, (1 - female_{M(i,t)}) + \omega\, (1 - female_i) \times (1 - female_{M(i,t)}) + \dots \\
&\quad = (\gamma_0 + \gamma_1 + \gamma_2 + \omega) - (\gamma_1 + \omega)\, female_i - (\gamma_2 + \omega)\, female_{M(i,t)} + \omega\, female_i \times female_{M(i,t)} + \dots
\end{aligned}$$

by how much the gender gap changes, but not whether male or female managers are to blame.

A major problem when estimating Equation (1.4) can be unobserved heterogeneity of workers and managers. If workers are sorted to managers based on these unobserved characteristics, the estimate of ω is biased. For example, if good male workers tend to work for good male managers, or good female workers often work for good female managers, identification is compromised. Therefore, we introduce in Equation (1.5) a worker fixed effect, α_i, and a manager fixed effect, $\Psi_{M(i,t)}$, for i's manager M at time t.

$$Y_{it} = \omega \times male_i \times male_{M(i,t)} + \alpha_i + \Psi_{M(i,t)} + X_{it}\beta + \epsilon_{it} \tag{1.5}$$

Estimating ω in Equation (1.5) yields the change in within-manager pay gaps, adjusted for worker quality, when the manager is male instead of female. Intuitively, this does the following. We residualize each worker's outcome based on the controls X_{it} and remove the worker mean. Then, we calculate the within-manager pay gap based on all workers who worked for the manager. Again, the level of the gap has no interpretation, because we subtract the mean from each worker's outcome. ω is the mean difference of the adjusted within-manager pay gap between male and female managers.

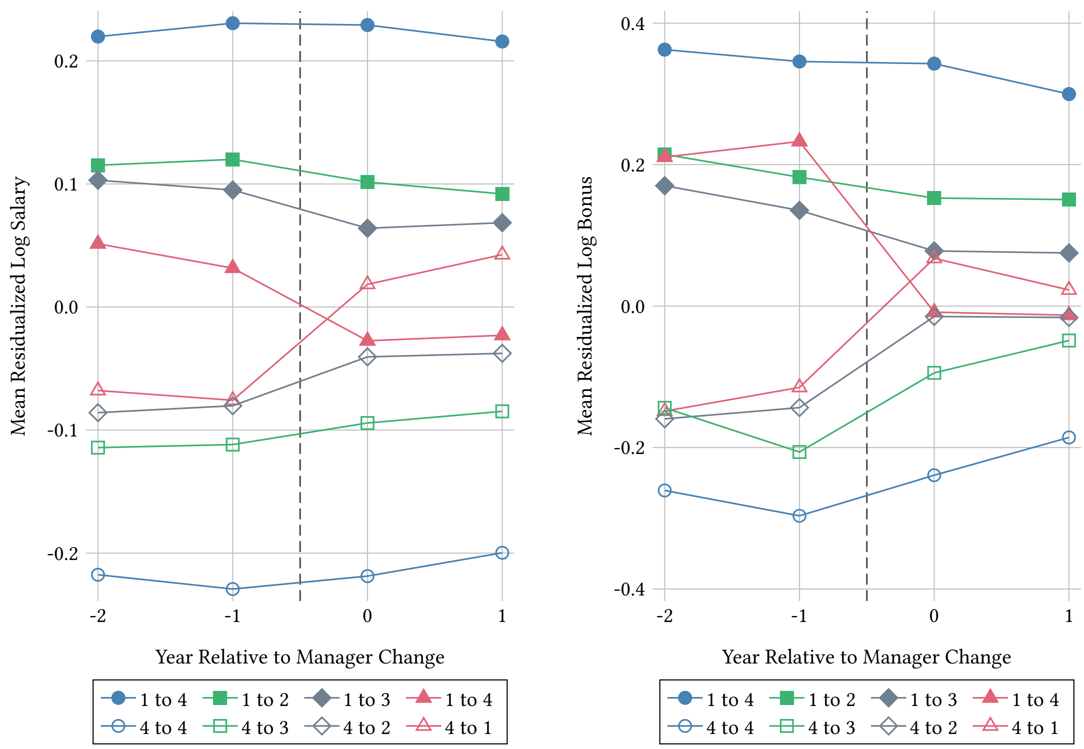
Identification Identification of ω in Equation (1.5) comes from "movers", i.e. workers who work for different managers. There are two reasons why a worker experiences a manager change. First, workers who switch their job will face a new manager.[14] Second, workers who do not change positions receive a new manager if the previous manager rotates to another job or leaves the firm. In the present setting, 32% of workers observe a change in manager in a given year, and almost 80% of these switches are due to managers rotating jobs.

We require an exogeneity assumption regarding the changes of managers. Sorting of workers to managers based on time-varying performance would bias estimations of Equation (1.5). The fact that the majority of changes is due to managers rotating limits this concern because managers would need to be assigned based on the potential future performance of workers. From our interviews with the human resources department of the firm this is extremely unlikely. Usually, managers only get to know their subor-

[14]Excluding the possibility that boss and worker jointly switch teams.

FIGURE 1.3.2: Mean Log Pay of Manager Changers by Quartiles of Mean Coworker Pay

(A) Salary (B) Bonus



Notes: These figures plot average residual base salaries and bonus payments in the two years before and after manager changes. Based on Card et al. (2013), workers are classified into 16 groups, of which eight are displayed. Workers are grouped by the quartile of their coworkers' pay before and after the manager change. The label "1 to 4" denotes workers whose coworkers' average salary was in quartile 1 before and in quartile 4 after the change of managers.

dinates after having started a new position. However, one might still worry that moves of workers correlate with time-varying performance or that workers only move if they get a better deal. While one cannot prove exogeneity, we examine whether there is evidence that sorting could bias the estimation results. As suggested by Card et al. (2013) in a setting where workers move across firms, we implement event studies on pay for workers changing managers.

To do so, we classify managers by the average of salaries or bonuses. To account for job characteristics, we first residualize these outcomes, taking into account job characteristics, location, tenure, age, and year. Then, we calculate the leave-one-out mean of the residual to avoid selection based on worker i's own productivity This is the average residual observed under i's current manager $M(i,t)$, excluding the contribution of i.

The leave-one-out means are then used to classify each worker-year observation into one of four quartiles. Next, we calculate the average residuals of workers two years before and two years after changing managers. We do so for salaries and bonuses of workers in all 16 possible transitions, e.g. workers changing from a category 4 to a category 1 manager, workers changing from a category 4 to a category 3 manager, etc. For clarity, we focus on workers who previously worked for category 4 or category 1 managers. Note that we include changes in manager due to job changes of workers and due to manager rotation.

Figure 1.3.2 shows that the groups have different pay levels before (years -2 and -1) and after changing managers (years 0 and 1). For example, salaries of workers with coworkers in the fourth quartile who move to a quartile 1 manager have lower salaries prior to a change compared to workers who change from quartile 4 to another quartile 4 manager. Moving to a manager with higher-paid coworkers, e.g. from quartile 1 to quartile 4, increases pay. Workers who stay in the same quartile have relatively constant pay, although bonus pay seems to increase quite a bit for workers switching from a quartile 1 to another quartile 1 manager. Workers who change from a quartile 4 manager to a lower quartile manager lose pay, with larger losses for more extreme changes.

Pay changes in Figure 1.3.2 look symmetric for workers moving between quartile 1 and quartile 4 managers. This suggests that a simple additive model is a reasonable approximation of base salaries and bonus payments. It implies that workers do not only change managers if higher residual pay is expected.

The pay profiles in Figure 1.3.2 also look relatively flat before and after changing manager. While there is some variation in pre-change pay, for example among workers moving from bonus quartile 1 to 4, these changes are small compared to the jumps we observe. This suggests that a static model as in Equation (1.5) should be a sufficient approximation.

Table 1.3: Gender Gap Decomposition

	Salary		Bonus Payout	
	Log points	Share explained	Log points	Share explained
Age	0.75	6.1%	0.93	4.2%
Tenure	0.16	1.3%	0.25	1.1%
Manager	3.13	25.4%	4.16	18.7%
Job	6.59	53.5%	13.57	61.1%
Year	-0.23	-1.8%	-0.83	-3.7%
Country	0.84	6.8%	0.38	1.7%
Total explained	11.24	91.2%	18.45	83.1%
Total unexplained	1.08	8.8%	3.75	16.9%
Total gap	12.32	100.0%	22.20	100.0%

Notes: The table displays Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in Log Salary and Log Bonus Pay based on 59 813 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

1.4 Results

1.4.1 Decompositions

In the first set of results, we look at the impact of men and women doing different jobs and in particular working for different managers. The decomposition of gender pay gaps requires that male and female workers identify the same set of characteristics and full-rank matrices for each gender. Therefore, we reduce the data to a dual-connected set, as described in Section 1.3. We also impose that for a given observation we observe base salary, bonus payout, performance ratings, and targets. While this excludes workers who do not have performance pay in their contract, this has the advantage that different results for the outcomes are not driven by sample composition.

Table 1.3 displays the gender pay gap decomposition based on 59 813 observations. The raw gender pay gap in salary is 12 log points (13%). The raw gap in bonuses is even larger, with a difference between men's and women's payouts of 22 log points (24%). While the raw pay gaps are large, between 80 and 90% can be explained by different observed characteristics of men and women. The residual gender gap is 1.1 log points for base salaries and 3.8 log points for bonuses.

The Kitagawa-Oaxaca-Blinder decomposition in Table 1.3 shows the relative importance of the different characteristics explaining the gap. Age and tenure differences between men and women exist but are of small magnitude. Also, the fact that the gender distribution of the workforce might not be uniform over countries is not of primary importance. Job characteristics, i.e. the combination of hierarchical rank and occupation, explain between 55 and 60% of the gender gaps in salary and bonus. This means that if women worked in the same occupations and hierarchical levels as men, and earned the same returns from these jobs as men, the gender pay gap in salaries would be reduced by more than half.

The sorting of workers to managers matters. 25% of the gender pay gap in base pay and 19% of the gap in bonuses can be explained by the fact that men work for managers who have more positive impacts on pay. This is true conditional on workers doing the same job, i.e. working in the same occupation within the same hierarchical rank. While we cannot tell why women tend to work for bosses who are less generous or have less favorable impacts on productivity, we can see that managers are not perfectly substitutable but differ in terms of productivity or styles. This also means that firms seeking to reduce gender pay gaps need to carefully consider if and how to better match women to high-impact managers. Should women prefer to work for managers whose style implies a productivity reduction, for example due to flexible work hours, female workers might in spite of a financial cost prefer the current allocation to managers (Goldin and Katz, 2015).

The role of child care A large part of gender pay gaps has been attributed to reduced working hours due to child care obligations (Kleven et al., 2019a,b). We neither observe family status nor overtime work. By limiting the data to full-time workers only and controlling granularly for the nature of the job, much of the variation in actual hours worked is already taken into account. However, to examine the role of child care in our setting, we expand the decompositions in several dimensions.

First, we repeat the pay decomposition for workers aged 44 and younger and workers aged 45 and older. Here one would expect that older workers are less likely to be affected by small kids at home. While child birth may have a lasting effect on careers of older workers as well, our approach already takes into account that parents might climb up the hierarchy more slowly. Appendix Table 1.A.2 shows that the adjusted salary gap among younger workers is actually zero, and smaller than among older workers. The adjusted bonus gap is around 5% for both groups. This suggests that differences in unobserved actual hours worked do not play a role for the gender gap in this sample. As the sample is limited to full-time workers only and due to the exact controlling for job fixed effects, most differences in work hours are probably already accounted for.

To examine this further, we recalculate pay decompositions while including all workers in the data and controlling for contracted working hours. As Appendix Table 1.A.3 shows, a significant part of the gender pay gap can be explained by differences in working hours. 34% of the pay gap in base salary of 13.7 log points is attributed to this channel. Unexplained gender pay gaps are larger in this full sample, relative to full-time workers only. This finding suggests that gender differences in working hours do matter, but are largely taken into account already by limiting observations to full-time employees.

Full-time workers might differ in their accumulated working hours. In particular, workers with children might have worked fewer hours in the past. Workers can also have spells during which they did not work at all, for example because of child birth. We treat these spells as having worked zero hours. Having worked part-time in the past could be interpreted as a measure of experience, less flexibility, or reduced likelihood of working overtime because of child care obligations. Appendix Table 1.A.4 decomposes full-time workers' pay while controlling in addition for accumulated full-time equivalent months. As we need to observe every worker's full employment history at the firm, the sample size is reduced. The unexplained gender pay gap is similar to our baseline estimate from Table 1.3. Differences in accumulated full-time equivalent months do not contribute to gender gaps. Once more, this could be because workers sort into jobs with different requirements of flexibility or because the firm requires certain experience for working in particular positions.

These extensions demonstrate that it is unlikely that the unexplained part of the gender pay gap documented in Table 1.3 can be attributed to child care obligations.

Sources of gender bonus pay gaps Bonuses depend on base salary, performance ratings, and contracted bonus targets. We have seen that a gap in base salaries exists

TABLE 1.4: Gender Gap Decomposition

	High Performance		Bonus Target	
	Percentage points	Share explained	Log points	Share explained
Age	-0.93	46.7%	0.27	9.5%
Tenure	-0.00	0.0%	-0.13	-4.6%
Manager	2.12	-106.0%	-3.44	-121.9%
Job	0.71	-35.3%	6.61	234.6%
Year	-0.06	2.9%	-0.08	-2.8%
Country	-0.30	14.9%	-2.44	-86.7%
Total explained	1.54	-76.8%	0.79	28.1%
Total unexplained	-3.54	176.8%	2.03	71.9%
Total gap	-2.00	100.0%	2.82	100.0%

Notes: The table displays Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in the probability of a high performance rating and contracted Log Bonus Targets based on 59 813 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

and can be explained by sorting to jobs and managers. Here we study whether sorting with regard to contracted bonus target or performance rating matters as well.

Table 1.4 reports that the raw gender gap in performance is −2 percentage points. The penultimate line shows that the adjusted gender gap is even more negative (−3.5 percentage points). This is mainly driven by the fact that women tend to work for managers handing out worse ratings. This means that if women had the same characteristics as men and were to earn the same returns on these characteristics as men, their performance ratings would be even *better*.

It is a striking result that gender performance gaps are negative, i.e. favor women, while the unexplained gap in salaries and bonus payouts is positive. If performance ratings are a good proxy for actual performance, these residual gaps are difficult to reconcile with the explanation that men are more productive. Previous work by Azmat and Ferrer (2017) shows that female lawyers perform worse than their male colleagues. Our results demonstrate that the use of subjective ratings can make much of a difference when comparing the performance of male and female workers.

How can it be that women earn lower bonuses, despite the fact that their ratings are better? Bonuses depend on base salaries and contracted bonus targets. As these gaps benefit men—the unexplained gap in targets is 2.0 log points and the unexplained gap in base salaries is 1.1 log points—the eventual payout still favors men.

Women do not sort to more generous or high-impact managers as can be seen from the detailed results of the decomposition in Table 1.4. Instead, men work for managers who hand out better ratings. If women worked for the same managers and benefited from them in the same way as men, the probability to receive a high performance rating would go up by two percentage points.

Table 1.4 also indicates that women are sorted to jobs with significantly lower bonus targets. However, women actually tend to work for managers who negotiate higher bonus targets. This means that while women might shy away from competition (as found by Niederle and Vesterlund, 2007), it does not imply that they work for managers where performance pay plays a smaller role.

Are women's better ratings explained by lower costs? One explanation for negative performance gaps is that managers might be more willing to hand out good ratings if they are less costly. We observed that women earn lower salaries and negotiate lower bonus targets. Managers could therefore use performance ratings to compensate women for negotiating lower salaries and targets.

To check this, we once more decompose the gender gaps in performance ratings, taking salary and target as predetermined variables and including them in the controls. We find in Appendix Table 1.A.5 that the unexplained gender gap opens even more in favor of women. This suggests that women do not simply receive better ratings because they cost the manager less.

Controlling for additional workplace characteristics We control for job characteristics by including a set of indicators for each combination of occupation and hierarchical rank. However, it could also be the case that male and female workers who work in the same job are located in different units of the firm. For example, a software engineer could work in the compensation unit of the human resources department or in the financial analysis team – and these different workplaces could provide very different pay packages. This means that by not taking into account the exact unit in which a worker is located, we could attribute such differences to the managers and thereby overestimate the role that the sorting of managers and workers plays for the gender pay gap.

In Appendix Table 1.A.6, we repeat the pay gap decomposition, but now including indicators for the unit in which a worker is employed. Each worker is matched to one of more than 200 units. The results show that a differential sorting of men and women can explain only very little of the gender pay gap. If anything, women tend to work in teams with slightly higher bonus pay. While not substantially changing our previous findings, this exercise further reduces the sample size to 50 002 observations because we finding a connected set in all categorical variables becomes more challenging.

1.4.2 Within-manager gender gaps

Having documented that the sorting of women to managers contributes to gender pay gaps, we now ask whether managers affect residual gender pay gaps.

Based on the same data used in the Kitagawa-Oaxaca-Blinder decomposition we estimate Equation (1.5) on salaries, bonus payouts, performance ratings, and bonus targets. The number of observations contributing to the estimates is slightly reduced relative to the sample in the decomposition due to the inclusion of worker fixed effects. If a worker is only observed for a single period she cannot contribute to the identification of ω, the coefficient on the term of interest, $male_i \times male_{M(i,t)}$.

In addition to this fully saturated specification, we estimate two less granular versions on the same observations. First, as specified in Equation (1.4), we do not account at all for unobserved heterogeneity at the worker and manager level. This allows us to include dummies for the worker and manager being male, respectively. In an intermediate step, we add manager fixed effects. A dummy indicating the gender of the worker can then still be estimated. Third, we estimate the fully saturated Equation (1.5).

Table 1.5 shows the results. The coefficients from the first row reports the adjusted gender gap when working under a female manager. The coefficient from the second row is the difference in outcomes of women when they work under a male instead of a female manager. The third row estimates the difference between the gender gap under male and female managers. Before turning to more granular estimations, we look at columns 1, 4, 7, and 10, which report the coefficients from estimating Equation (1.4).

The gender salary gap under female managers is 2.3 log points. Women who work for male managers earn 1.9 log points higher salaries. Men only earn 1.6 log points higher salaries, but the difference to women is insignificant. This means that the esti-

TABLE 1.5: Effects of Having a Same-Gender Superior

	log(Salary)			log(Bonus Payout)			High Performance			log(Bonus Target)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Male	0.023	0.026		0.014	0.028		-0.037	-0.034		0.008	0.008	
	(0.006)	(0.005)		(0.012)	(0.010)		(0.008)	(0.009)		(0.008)	(0.006)	
Male Mngr.	0.019			0.018			-0.005			0.003		
	(0.007)			(0.013)			(0.008)			(0.010)		
Male × Male Mngr.	-0.003	-0.003	0.002	0.041	0.017	0.051	0.014	0.006	0.027	0.005	0.007	0.005
	(0.007)	(0.006)	(0.003)	(0.014)	(0.012)	(0.015)	(0.010)	(0.011)	(0.015)	(0.009)	(0.007)	(0.005)
Job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tenure FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager FE		Yes	Yes		Yes	Yes		Yes	Yes		Yes	Yes
Worker FE			Yes			Yes			Yes			Yes
N	55,464	55,464	55,464	55,464	55,464	55,464	55,464	55,464	55,464	55,464	55,464	55,464
R^2	0.903	0.954	0.995	0.816	0.875	0.939	0.073	0.191	0.563	0.861	0.934	0.988

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

mated gender pay gap at male managers is smaller than at female managers by 0.3 log points, but this is highly insignificant.

Turning to bonuses, we find that gender gaps are 4.1% larger when the manager is male. The other coefficients imply that there is no statistical gender bonus gap under female managers and no advantage for women who work for male managers instead of female managers.

The increase in gender bonus gaps when the manager is male should be explained by higher salaries, performance ratings, or bonus targets. However, as we can see from the third row in columns 1, 7, and 10 this is not the case. The reason for this is that the high performance indicators are noise measures of actual performance ratings, which are more nuanced. We cannot reconstruct the exact mapping from more detailed performance ratings to bonuses. Also, omitted variable bias in unobserved worker or manager characteristics could invalidate the estimation approach.

In columns 2, 5, 8, and 11 we control for unobserved manager characteristics. Three observations can be made in comparison to the version without manager effects. First, the gender gap under female workers in salaries and ratings is almost unchanged. Second, the bonus gap under female workers now is higher and significant at 2.8 log points. Third, differences in gender gaps are all insignificant and smaller.

While the gender gap at female managers is an interesting statistic, we are looking for a causal interpretation of the effect of manager gender on gender gaps. To ensure that workers are comparable, we add worker fixed effects in the complete specification of Equation (1.5) in columns 3, 6, 9, and 12. This proves critical for the estimation of the coefficient of interest.

Working for a male manager increases bonuses by 5.1 log points relative to women when controlling for worker and manager fixed effects. Put differently, the within-manager gender gap increases by 5% if the manager is male. Due to symmetry, this also implies that the within-manager gender gap falls by 5%, i.e. moves in favor of women, if the manager is female. There is no evidence that the manager's gender matters for gaps in base salaries or bonus targets. Instead, the advantageous position of men at male managers is driven by changes in performance ratings. The performance gap increases by 2.7 percentage points, i.e. moves in favor of men, when the manager is male. This represents more than 10% relative to the observed probability of receiving a high rating. The finding that gender gaps open up when the manager is male relates to

Hospido and Sanz (2019) and Mengel et al. (2019) who document differences in gender gaps for different decision-maker genders in academic settings and thereby challenges findings by Card et al. (2020) who cannot document such differences.

As the point estimate for high performance ratings is only statistically significant at 10%, we examine the robustness of our findings to alternative performance rating measures in Appendix Table 1.A.7. The effect of male managers on the gender gap in the probability of receiving a very high rating (the highest possible grade) is not significant. However, Table 1.A.7 also shows that the gender gap in the probability of receiving a bad rating (one of the two lowest possible grades) shrinks when the manager is male. Finally, we translate the six grades to a numeric scale from one to six, where six refers to the best grade and one to the worst. The gender gap in this measure significantly increases by 0.02 points when the manager is male. These alternative performance outcomes lend support to our finding that working for male managers benefits male workers' ratings more than those of female workers.

Heterogeneity in the impact of manager gender on gender gaps We also allow the effect of male bosses on the gender gap to vary along other dimensions. Table 1.6 displays these results. Each panel and column refers to a separate regression. All estimates of manager gender impacts on base salaries are insignificant. The effects of manager gender on the bonus gap do not differ by worker or manager characteristics. However, some interesting patterns emerge.

Younger workers seem to be more affected by the differential gender gaps. This could be because younger workers are found at lower ranks. But as panel B shows, if anything the opposite is true. It seems that at higher ranks the gender of the manager plays a larger role for gaps. In panel C it looks like the effects of manager gender are larger when the team is mostly male. This observation supports the interpretation that male bosses cater to the needs and preferences of male workers. Panels D to F look at characteristics of managers. The results for age and hierarchical rank follow a similar pattern as for workers. Because the cultural background of managers could impact their treatment of workers of different gender, we allow the effect to vary by the region of origin. We classify managers as "western" if they have a European or Anglo-Saxon nationality. However, the point estimates look very similar. Overall, there is little statistical variation in the effect of managers, suggesting that the effect of manager gender on the gender gap is an issue across the entire firm.

Other manager characteristics It could be that manager characteristics apart from gender are the actual fundamental drivers of within-manager gaps. Note that even if this was the case, it is already clear that managers do matter. Here we examine whether gender indeed drives this finding.

We repeat the estimation of Equation (1.5) for all four outcomes considered, but in addition interact the dummy $male_i$ with manager characteristics correlated with manager gender. We allow gender gaps to vary by managers' age, tenure, origin, and hierarchical rank. If managers can experience a change in their characteristics, for example in hierarchical rank, the manager fixed effect does not absorb it. In this case, in addition to $male_i \times rank_{M(i,t)}$ we include $rank_{M(i,t)}$ as a regressor, etc.

Comparing the results from Table 1.7 with the original results from Table 1.5 shows that even if we control for the interaction between worker gender and various manager characteristics, the estimated coefficient on $male_i \times male_{M(i,t)}$ in the first row is almost identical. This suggests that omitted variable bias is not driving our finding but that indeed manager gender affects within-manager gender gaps.[15]

Table 1.7 also reveals that older managers close gender gaps in base salaries and contracted targets, relative to younger managers. The estimation also suggests that gender gaps in performance rating are greater when the manager is "western", suggesting that coming from a culturally more progressive society does not guarantee a reduced impact of managers on gaps. The finding could also mean that managers from a minority group are more concerned about the interests of other minorities.

Allowing for differential returns for men and women In the estimation of Equation (1.5), returns from male managers can differ between men and women while all other characteristics are assumed to have equal effects on men's and women's outcomes. This simplification could lead to bias if the differential returns capture differen-

[15]With a p-value of 12%, the coefficient on $male_i \times male_{M(i,t)}$ in Table 1.7 is statistically insignificant when considering gaps in the probability of receiving a high performance rating. This again might be because the grading is more complex than just an indicator for high ratings. Furthermore, while effects on base salary and contracted bonus targets remain insignificant, these can of course contribute to the effect on bonus payments in column 2.

Table 1.6: Effects of Having a Same-Gender Superior: Heterogeneity

	log(Salary)	log(Bonus Payout)
Panel A: Age		
Male × Male Mngr.	0.003	0.062
	(0.004)	(0.021)
Male × Male Mngr. × > 44 years	−0.003	−0.027
	(0.005)	(0.028)
N	55,464	55,464
Panel B: Hierarchical Rank		
Male × Male Mngr.	0.003	0.052
	(0.004)	(0.037)
Male × Male Mngr. × Medium Rank	−0.002	−0.011
	(0.006)	(0.040)
Male × Male Mngr. × High Rank	0.005	0.070
	(0.011)	(0.058)
N	55,464	55,464
Panel C: Team Composition		
Male × Male Mngr.	−0.003	0.045
	(0.004)	(0.020)
Male × Male Mngr. × Majority Male	0.007	0.014
	(0.005)	(0.024)
N	55,016	55,016
Panel D: Manager Age		
Male × Male Mngr.	0.003	0.049
	(0.004)	(0.019)
Male × Male Mngr. × > 44 years	−0.002	0.006
	(0.006)	(0.024)
N	55,464	55,464
Panel E: Manager Hierarchical Rank		
Male × Male Mngr.	−0.011	−0.177
	(0.020)	(0.167)
Male × Male Mngr. × Medium Rank	0.012	0.215
	(0.020)	(0.168)
Male × Male Mngr. × High Rank	0.013	0.256
	(0.021)	(0.168)
N	54,654	54,654
Panel F: Manager Region		
Male × Male Mngr.	0.002	0.065
	(0.007)	(0.034)
Male × Male Mngr. × Western	−0.001	−0.018
	(0.008)	(0.038)
N	55,464	55,464

Notes: Standard errors are in parentheses and clustered at the worker and manager level.

Table 1.7: Effects of Having a Same-Gender Superior: Alternative Channels

	log(Salary)	log(Bonus Payout)	High Performance	log(Bonus Target)
	(1)	(2)	(3)	(4)
Male × Male Mngr.	0.001	0.055	0.024	0.004
	(0.003)	(0.016)	(0.015)	(0.006)
Male × Mngr. Age ≥ 45	-0.006	-0.020	0.012	-0.009
	(0.003)	(0.013)	(0.013)	(0.004)
Male × Mngr. Tenure ≥ 10	-0.001	0.003	-0.000	-0.002
	(0.003)	(0.014)	(0.013)	(0.005)
Male × Mngr. Western	0.011	-0.012	0.066	0.010
	(0.009)	(0.036)	(0.034)	(0.010)
Male × Mngr. High Rank	0.002	0.014	0.015	-0.005
	(0.005)	(0.020)	(0.020)	(0.008)
Job FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Tenure FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
N	54,784	54,784	54,784	54,784
R^2	0.995	0.937	0.565	0.988

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

tial returns from other characteristics. For example, if male bosses are more likely to work in engineering, and men have, for whatever reason, higher returns than women when working in engineering, we would blame managers for these differential returns.

In Appendix Table 1.A.8 we repeat the estimation of Equation (1.5) but interact all variables except worker and manager effects with a gender dummy. The effect of managers on gender gaps is unchanged. This adds to confidence that we are indeed estimating the gender-differential effect of bosses.

Controlling for additional workplace characteristics We also examine the robustness of our findings when adding controls for work units. In the gender gap decomposition, this exercise did not affect the interpretation of the results. This is also the case here. In Appendix Table 1.A.9 we report estimates which are highly similar to the effect of manager gender on gender gaps shown previously.

Identification from workers not switching jobs In Equation (1.5), the coefficient of interest ω is identified by workers who work for different managers during the sample period. But identification could be compromised if a worker selects into a new job with a different boss based on unobserved time-varying performance of the worker. While our analysis of event studies (based on Card et al., 2013) suggests that assuming that the matches of workers and managers are as good as exogenous, we can of course not formally test whether this assumption is true. In Table 1.A.10 we therefore replicate our analysis based on a sub-sample of workers who do not face a new manager because of job changes of workers but because of new managers substituting the previous manager of a team. This means that we drop all observations where a worker who changes manager also changes her job. For this exercise we do not make use of the sample used in the decomposition where worker mobility was key in order to find a connected set and identify all fixed effects. Here, we want to restrict the mobility of workers. The results in Table 1.A.10 look very similar to our main result from Table 1.5, providing further confidence that assuming as good as exogenous matching of workers and firms is a reasonable approximation.

1.4.3 Why does manager gender matter?

Three main mechanisms could explain why the gender pay gap moves in favor of men when managers are male. First, the sorting of more productive men to male managers could result in an observed gender gap. Second, within-gender complementarities might make male workers relatively more productive than women when working under a male manager. Third, (unconscious) discrimination of workers of the other gender could drive our findings. Whatever explanation holds true, as long as women are underrepresented in managerial positions, these explanations would imply structural disadvantages for female workers We now evaluate which of these mechanisms is likely to explain the effect of manager gender.

Sorting Workers are not allocated randomly to managers. Managers might be better informed about the quality of a worker of the same gender. This could for example imply that a male manager's male workers are on average better than female workers. While such a mechanism could exist in the firm, it cannot explain our results. In Specification (1.5), we control for unobserved heterogeneity of workers. Doing so fully takes into account time-invariant ability differences. This means that our results hold true conditional on the sorting of workers to managers.

Within-gender complementarities A second explanation could be complementarities within gender. While there is ample evidence that diverse teams are more fact-focused, process facts more carefully, and are more innovative (e.g. Díaz-García et al., 2013; Herring, 2009; Levine et al., 2014; Nathan and Lee, 2013; Phillips et al., 2009), one could think that homogeneity can also benefit employee performance. For example, a competitive worker might be more productive in a competitive environment, while a cooperative worker might be more productive in a cooperative environment. If the distribution of styles differs for men and women, men should be more productive when working with men and we would observe a wider gender gap due to productivity.

While we have no measure of productivity available, we can test for the plausibility of this mechanism. Under the assumption that within-gender complementarities also would exist with respect to coworkers of the same gender, we can test for complementarities by studying the effect of male coworkers on the gender pay gap. We therefore repeat the estimation of Equation (1.5) for the share of male coworkers working for the same manager. The interaction of male-dummies for worker and manager is replaced by the interaction of a male-dummy for the worker and the share of male coworkers. We include the share of male coworkers because in contrast to manager gender it is not absorbed by the manager fixed effect. We still include worker fixed effects, manager fixed effects, and all other controls. This differs from Panel C of Table 1.6 as here we directly estimate the effect of male coworkers instead, not the heterogeneous effect of a male managers leading majorly male teams.

Table 1.8 contains only one significant coefficient, indicating that bonuses are higher when working in predominantly male teams. But, if anything, bonus payments are lower for men compared to women when working with more male coworkers. Under the assumption that within-gender complementarities also need to exist with respect to coworkers, the results imply that such complementarities do not exist with respect to manager gender. Therefore, we take this as supporting evidence that complementarities within gender are highly unlikely to substantially explain the favorable outcomes of men (women) under men (women).

Table 1.8: Effects of Having more Same-Gender Coworkers

	log(Salary)	log(Bonus Payout)	High Performance	log(Bonus Target)
	(1)	(2)	(3)	(4)
Male × Share of Male Coll.	0.001	-0.040	0.006	-0.001
	(0.006)	(0.028)	(0.027)	(0.011)
Share of Male Coll.	0.004	0.045	-0.004	0.006
	(0.006)	(0.023)	(0.022)	(0.009)
Job FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Tenure FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
N	55,016	55,016	55,016	55,016
R^2	0.995	0.939	0.564	0.988

Notes: Standard errors are in parentheses and clustered at the worker and manager level. *Share of Male Coll.* is the share of colleagues, defined as workers with the same superior, who are male. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

Discrimination and biased beliefs Managers who are uninformed about the productivity of their subordinates might be more likely to resort to (unconscious) biases when evaluating or negotiating with workers. Discrimination due to biased beliefs (Bohren et al., 2019) implies that decision-makers discriminate according to their biased priors if they are uninformed. If they are provided with previous evaluations of work quality by other decision-makers, discrimination is reduced and eventually flips if much previous information is available.

We suggest that a related mechanism can drive discrimination in our setting. Instead of collecting information from previous evaluators, one can easily think that decision-makers are less likely to discriminate according to their preexisting biased beliefs as they learn about the worker's true quality. So while Bohren et al. (2019) relate a reduction in discrimination to better information, one can also expect that a reduction in discrimination is related to better knowledge of the worker.

Besides learning about the true quality of workers, managers learning about their workers also could exhibit less discrimination for a separate reason. It could be the case that managers are less aware of the needs of workers of the other gender. For example, male managers could be less considerate of child care obligations of female workers. Over time, they could learn about their workers' needs and reshape the work environment such that all workers can show their best.

Note that this does not necessarily mean that only male managers discriminate. Female managers could be discriminating as well, so that we estimate the additional advantageous treatment of men. Theoretically, it could also be the case that male managers do not discriminate at all, and female managers favor female workers. We do not consider this a likely scenario, as previous research showed that women if anything tend to discriminate against women (e.g. Bagues and Esteve-Volart, 2010) and bias in evaluations has been found in other settings to be driven by men (Hospido and Sanz, 2019; Mengel et al., 2019)

Whether discrimination takes more direct forms, e.g. biased ratings, or indirect forms, e.g. biased work environments, we would expect in both cases that discriminating is reduced when managers are more knowledgeable about their workers. To test this, we group observations by a number of variables capturing aspects of information.

First, we split managers into a group with less than ten and a group with more than ten years of tenure at the firm. One would expect that these managers are more experienced and therefore less likely to fall back to preexisting biases, or less aware about the needs of workers of the other gender. Second, we examine whether managers of larger teams (five or more subordinates) show stronger effects of manager gender on gaps. Managers of larger teams might find it harder to observe the true effort of each worker and to cater to specific needs. Similarly, a manager who works at a different location than the subordinate and who only communicates remotely might find it harder to evaluate workers or build a connection with workers who are less alike themselves. Finally, we allow effects to change with the time a worker and manager have worked together. We do so by grouping relationships into a group with two or less and a group with more than two years of joint work. Over time, one would expect that superiors learn more and build connections with workers. Note that this measure could be biased if workers who feel discriminated against are more likely to leave their team.

Table 1.9 shows the results. Each column of each panel refers to a separate regression. Sample sizes vary because some variables are not observed for all workers and managers. For salary, the effect of manager gender on pay gaps remains statistically insignificant. For bonus payments, one can observe a pattern that could be taken as support of the findings of Bohren et al. (2019), i.e. that better information leads to less discrimination. The coefficients are not statistically distinguishable at conventional levels. However, in all cases it seems that gender effects are larger in magnitude when

Table 1.9: Effects of Information

	log(Salary)	log(Bonus Payout)
Panel A: Manager Tenure		
Male × Male Mngr.	−0.001	0.058
	(0.004)	(0.021)
Male × Male Mngr. × ≥ 10 yrs Mngr. Tenure	0.006	−0.013
	(0.005)	(0.027)
N	55,450	55,450
Panel B: Team Size		
Male × Male Mngr.	0.002	0.057
	(0.003)	(0.017)
Male × Male Mngr. × Small Team	−0.002	−0.022
	(0.005)	(0.023)
N	55,464	55,464
Panel C: Location		
Male × Male Mngr.	−0.000	0.096
	(0.007)	(0.032)
Male × Male Mngr. × Same Location	0.001	−0.054
	(0.008)	(0.034)
N	54,326	54,326
Panel D: Joint Time		
Male × Male Mngr.	0.001	0.057
	(0.003)	(0.017)
Male × Male Mngr. × > 2 Years Joint	0.001	−0.040
	(0.005)	(0.028)
N	49,338	49,338

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

information is harder to obtain. In particular, manager gender effects on pay gaps are larger when managers are less experienced, manage larger teams, work at different locations and know workers for a shorter amount of time. While none of these differences are statistically significant, all coefficients are negative and, in the case of location and joint time, close to significance with respective p-values of 12.0 and 15.8%, respectively. Proximity could also have led to the opposite effect on gender gaps if closer ties between managers and workers facilitate favoritism, an observation made by Cullen and Perez-Truglia (2019). Of course, such an effect might exist but seems to be more than netted out by managers accumulating additional information about workers.

The observation that overall the effect of manager gender on pay gaps seems to fall when the manager should be better informed supports the interpretation that some form of (unconscious) discrimination is an important driver of the results. While the observations here are in favor of discrimination by initially biased managers, they also contradict the mechanism discussed previously. If gender-specific complementarities would play a role, it is unclear why they should fall with better informed managers.

1.5 Conclusion

Using novel personnel data from a large company we examine gender gaps in wages and performance, and we show that they are affected by managers. Our analysis yielded three main findings.

First, a significant part of the gender gap can be explained by the sorting of male and female workers to different types of managers. Men are more likely to work for "better" managers, i.e. managers whose workers receive higher salaries, bonuses, and performance evaluations. The observation that women tend to work for "worse" managers means that firms seeking to improve gender equity need to find out what makes a "good" manager and why women tend to work for managers with a lower impact. While our research cannot determine which underlying manager characteristics drive wage inequality or why workers are sorted as they are, our results imply that firms should not only foster the occupational upgrading of female employees, but also consider how female workers can work for "better" bosses. However, women might actually prefer to work for "worse" managers if these offer more family-friendly environments (Goldin and Katz, 2015).

Second, we show that the gender pay gap cannot be explained by the notion that men outperform women. On the contrary, women actually receive significantly better ratings than men. Yet, on average, they earn less. This is a striking result as it implies that adjusted gender pay gaps should be even larger. It also challenges the notion that women's performance is worse due to being less ambitious or competitive (Azmat and Ferrer, 2017; Niederle and Vesterlund, 2007). In spite of the positive impact of ratings on bonus pay, bonus pay gaps still favor men because the performance effect is outweighed by differences in salaries and targets. Firms often resort to performance ratings determined by superiors if actual output cannot be quantified, as is typical in complex organizations characterized by division of labor. Future research should examine whether male and female superiors differ in what aspects of worker performance they value most when determining ratings. If firms interpret performance ratings as good proxies of actual performance, gender equity is not achieved when women earn the same as men. If anything, a negative gender performance gap means that women should earn more than men.

Third, we show that manager gender matters as male managers cause within-team bonus gaps to increase. This is driven by the fact that performance ratings are relatively more favorable towards men when the manager is male. As manager gender affects pay gaps, the over-representation of men in management position puts women at a disadvantage. Therefore, our research has important implications for the discussion of gender quotas. The basic requirement for such quotas to work is that having more female managers indeed improves gender equality. In contrast to quotas applying at the executive level only (Bertrand et al., 2019; Maida and Weber, 2019), our findings imply that quotas across all hierarchical ranks can be effective. Future research would need to consider other requirements that need to be fulfilled such that gender quotas across all ranks are indeed a suitable policy.

Digging deeper, we find suggestive evidence that discrimination due to biased beliefs could drive the findings, as manager gender tends to matter less with more knowledge about the workers. Alternatively, managers might learn about workers' needs and improve upon their initially biased work environments. This observation can inform alternative pathways for promoting gender equity. Organizing employees into smaller and more stable teams in closer physical proximity could be a feasible measure to reduce gender gaps. In addition, many firms train their staff to be more aware of gender-related biases. While the success of such diversity programs is found to vary significantly (Chang et al., 2019a), they seem to be a necessity to make managers more aware of their gender-related biases.

Appendix

1.A Additional Tables

Table 1.A.1: Descriptive Statistics

	Mean	SD	10th Perc.	Median	90th Perc.	Observations
Salary [€]	64 315.32	75 438.37	23 323.65	56 252.3	114 753.28	59 813
Bonus [€]	12 293.32	21 707.14	1393.49	4816.59	28 924.78	59 813
High Performance	0.25	0.43	0.0	0.0	1.0	59 813
Low Performance	0.1	0.3	0.0	0.0	1.0	59 813
Bonus Target [%]	12.36	8.68	4.0	11.0	25.0	59 813
Male	0.58	0.49	0.0	1.0	1.0	59 813
Age	42.65	9.04	30.0	43.0	55.0	59 813
Tenure	10.73	9.24	2.0	8.0	25.0	59 813
Span of Control	1.64	4.55	0.0	0.0	6.0	59 813
Coworkers	8.33	9.6	2.0	5.0	18.0	59 813
New Manager in Same Job	0.24	0.43	0.0	0.0	1.0	46 534
New Manager in New Job	0.07	0.25	0.0	0.0	0.0	46 534
Male Manager	0.72	0.45	0.0	1.0	1.0	59 813
Male & Male Manager	0.45	0.5	0.0	0.0	1.0	59 813
Male & Female Manager	0.13	0.34	0.0	0.0	1.0	59 813
Female & Male Manager	0.27	0.44	0.0	0.0	1.0	59 813
Female & Female Manager	0.15	0.36	0.0	0.0	1.0	59 813
Age of Manager	45.92	7.93	35.0	46.0	56.0	59 813

Notes: Extreme values omitted for confidentiality reasons. Unbalanced panel based on 20 048 workers and the years 2014-2019. Monetary variables normalized to Euros in 2010.

Table 1.A.2: Gender Gap Decomposition for Younger and Older Workers

(a) 44 and younger

	Salary		Bonus Payout	
	Log points	Share explained	Log points	Share explained
Age	0.40	6.2%	0.60	4.7%
Tenure	-0.01	-0.2%	-0.04	-0.3%
Manager	1.50	23.2%	4.39	34.9%
Job	3.87	59.9%	1.93	15.3%
Year	-0.13	-2.0%	-0.44	-3.5%
Country	0.95	14.8%	1.06	8.4%
Total explained	6.58	101.9%	7.50	59.6%
Total unexplained	-0.12	-1.9%	5.09	40.4%
Total gap	6.46	100.0%	12.59	100.0%

(b) 45 and older

	Salary		Bonus Payout	
	Log points	Share explained	Log points	Share explained
Age	0.11	0.6%	0.13	0.3%
Tenure	-0.00	-0.0%	0.16	0.4%
Manager	3.93	22.7%	4.85	13.1%
Job	13.21	76.4%	28.98	78.0%
Year	-0.29	-1.7%	-0.91	-2.5%
Country	-0.93	-5.4%	-1.42	-3.8%
Total explained	16.03	92.7%	31.78	85.5%
Total unexplained	1.27	7.3%	5.37	14.5%
Total gap	17.30	100.0%	37.15	100.0%

Notes: The tables display Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in Log Salary and Log Bonus Payout based on 48 537 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

Table 1.A.3: Gender Gap Decomposition Including Part-Time Workers

	Salary		Bonus Payout	
	Log points	Share explained	Log points	Share explained
Working hours	4.67	34.1%	1.93	7.5%
Age	0.56	4.1%	0.62	2.4%
Tenure	0.04	0.3%	-0.04	-0.2%
Manager	1.45	10.6%	4.52	17.5%
Job	5.72	41.8%	13.08	50.7%
Year	-0.25	-1.8%	-0.88	-3.4%
Country	-0.18	-1.3%	-0.54	-2.1%
Total explained	12.01	87.7%	18.69	72.5%
Total unexplained	1.68	12.3%	7.10	27.5%
Total gap	13.69	100.0%	25.79	100.0%

Notes: The table displays Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in Log Salary and Log Bonus Payout based on 66 040 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

Table 1.A.4: Gender Gap Decomposition Controlling for Accumulated Full-Time-Equivalent Months

	Salary		Bonus Payout	
	Log points	Share explained	Log points	Share explained
FTE months	-0.00	-0.0%	0.12	0.6%
Age	0.50	4.1%	0.51	2.4%
Tenure	-0.03	-0.2%	0.41	2.0%
Manager	1.37	11.2%	2.63	12.5%
Job	7.38	60.0%	12.21	58.1%
Year	-0.18	-1.5%	-0.25	-1.2%
Country	1.37	11.2%	0.42	2.0%
Total explained	10.42	84.7%	16.06	76.3%
Total unexplained	1.88	15.3%	4.98	23.7%
Total gap	12.30	100.0%	21.04	100.0%

Notes: The table displays Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in Log Salary and Log Bonus Payout based on 24 389 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

TABLE 1.A.5: Gender Gap Decomposition Controlling for Salary and Target

	High Performance	
	Percentage points	Share explained
Salary	1.62	-88.6%
Bonus Target	0.04	-2.1%
Age	-0.81	44.4%
Tenure	-0.03	1.5%
Manager	1.80	-98.7%
Job	-0.56	30.9%
Year	-0.04	1.9%
Country	-0.20	10.9%
Total explained	1.82	-99.7%
Total unexplained	-3.64	199.7%
Total gap	-1.82	100.0%

Notes: The table displays Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in the probability of a high performance rating based on 59 813 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation. Salary and bonus target are measured in logs.

TABLE 1.A.6: Gender Gap Decomposition Controlling for Unit

	Salary		Bonus Payout	
	Log points	Share explained	Log points	Share explained
Age	0.68	6.4%	0.90	4.3%
Tenure	0.13	1.2%	0.16	0.8%
Manager	2.29	21.4%	4.85	23.4%
Job	6.07	56.6%	11.72	56.5%
Year	-0.25	-2.3%	-0.83	-4.0%
Country	0.64	5.9%	0.24	1.2%
Unit	-0.15	-1.4%	-1.17	-5.6%
Total explained	9.41	87.8%	15.87	76.5%
Total unexplained	1.31	12.2%	4.89	23.5%
Total gap	10.72	100.0%	20.75	100.0%

Notes: The table displays Kitagawa-Oaxaca-Blinder-decompositions of the gender gap in the probability of a high performance rating based on 50 002 observations. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation. Salary and bonus target are measured in logs.

Table 1.A.7: Effects of Having a Same-Gender Superior: Additional Performance Measures

	High Performance	Very High Performance	Low Performance	Linearized Performance
	(1)	(2)	(3)	(4)
Male × Male Mngr.	0.027	0.009	-0.022	0.058
	(0.015)	(0.007)	(0.013)	(0.025)
Job FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Tenure FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
N	55,464	55,464	55,464	55,464
R^2	0.563	0.469	0.533	0.607

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

Table 1.A.8: Effects of Having a Same-Gender Superior: Gender-Interacted Controls

	log(Salary)	log(Bonus Payout)	High Performance	log(Bonus Target)
	(1)	(2)	(3)	(4)
Male × Male Mngr.	-0.001	0.053	0.026	0.003
	(0.003)	(0.016)	(0.015)	(0.005)
Job × Gender FE	Yes	Yes	Yes	Yes
Year × Gender FE	Yes	Yes	Yes	Yes
Age × Gender FE	Yes	Yes	Yes	Yes
Tenure × Gender FE	Yes	Yes	Yes	Yes
Country × Gender FE	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
N	55,412	55,412	55,412	55,412
R^2	0.995	0.940	0.569	0.989

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

TABLE 1.A.9: Effects of Having a Same-Gender Superior: Controlling for Unit

	log(Salary)	log(Bonus Payout)	High Performance	log(Bonus Target)
	(1)	(2)	(3)	(4)
Male × Male Mngr.	0.002	0.047	0.030	0.004
	(0.003)	(0.017)	(0.017)	(0.005)
Job FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Tenure FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
Unit FE	Yes	Yes	Yes	Yes
N	48,138	48,138	48,138	48,138
R^2	0.995	0.939	0.560	0.990

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

TABLE 1.A.10: Effects of Having a Same-Gender Superior: Keeping workers who do not switch managers due to job changes

	log(Salary)	log(Bonus Payout)	High Performance	log(Bonus Target)
	(1)	(2)	(3)	(4)
Male × Male Mngr.	0.001	0.052	0.044	-0.002
	(0.002)	(0.018)	(0.017)	(0.004)
Job FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Tenure FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
N	68,719	68,719	68,719	68,719
R^2	0.996	0.939	0.574	0.991

Notes: Standard errors are in parentheses and clustered at the worker and manager level. Age and tenure are summarized in six-year bins. Jobs are the combination of hierarchical rank and occupation.

Chapter 2

Urban Air Pollution and Sick Leaves: Evidence from Social Security Data

Joint with Laura Hospido and Ulrich J. Wagner

2.1 Introduction

Air pollution poses a major threat to public health by shortening lives (Deryugina et al., 2019) and increasing acute morbidity (Schlenker and Walker, 2016). As a negative externality of many economic activities, air pollution causes additional damage by reducing productivity on the job (Graff Zivin and Neidell, 2012) and by hindering human capital accumulation (Currie et al., 2009a; Ebenstein et al., 2016). The hypothesis that air pollution damages the economy also via reductions in labor supply was first examined by Ostro (1983) and Hausman et al. (1984). Recent research has provided credible evidence in support of this hypothesis in the context of emerging economies (Hanna and Oliva, 2015; Aragón et al., 2017), but little is known so far about this relationship in post-industrial societies where pollution levels are low and productivity is high.

In the G7 countries, the mean population exposure to fine particulate matter ($PM_{2.5}$) fell by 35% between 1990 and 2017, driven in no small part by costly environmental regulations. Over the same period, GDP per hour worked increased by 50% (OECD, 2019). These trends have important implications for cost-benefit analysis of air quality regulations. First, with air quality improving, sub-clinical health impacts such as sick leaves taken gain relative importance compared to severe health impacts which have been the main focus of the health literature so far. Second, higher labor productivity means that work days lost due to air pollution account for larger economic damages. Taken together, this calls for a better understanding of the labor-supply impacts of air pollution.

To shed light on this important issue, this chapter provides the first causal estimates of how many work days are lost due to air pollution concentrations typically observed in post-industrial economies. Our empirical analysis is based on a novel administrative dataset that links rich information on personal and occupational characteristics of Spanish workers to the frequency, length, and diagnosis, of sick leaves taken. We estimate the impact of air pollution on workers' propensity to call in sick, based on weekly variation in ambient concentrations of particulate matter (PM_{10}) across 99 cities in Spain. Our baseline model is a linear regression of the share of sick-leave days on the share of high-pollution days and weather variables. To control for non-random assignment of pollution across workers, we include city-by-year, year-by-quarter, and worker fixed-effects. Possible remaining endogeneity is addressed in an instrumental-variables regression that exploits exogenous variation in PM_{10} driven by dust storms in Northern Africa.

Our estimates imply that a 10%-reduction in high-pollution events reduces the absence rate by 0.8% of the mean (2.79%). This effect is robust across a variety of specifications and economically significant when scaled up to the entire work force. The estimates imply that the improvement in ambient air quality in urban Spain between 2005 and 2014 saved at least € 505 million in foregone production by reducing worker absence by more than 5.58 million days. In further analysis, we uncover two important sources of treatment heterogeneity. One relates to pre-existing medical conditions that we infer from a worker's sick leave record. We estimate that the health response of vulnerable workers (defined as those belonging to the top-five percentile of the distribution of sick leaves taken during the pre-estimation period) is almost three times that of healthy workers. Furthermore, our analysis of treatment heterogeneity across workers and occupations reveals that job security is an important moderating factor in workers' decisions to take a sick leave in response to a pollution shock. Specifically, we estimate that workers with a high predicted risk of losing their job respond less strongly than those with high job security. Using counterfactual analysis, we show that such interactions of behavior and labor market institutions have a large impact on estimates of the external costs of air pollution that arise from changes in labor supply.

This chapter makes several substantive contributions to the literature. First and foremost, we provide a comprehensive assessment of the impacts of air pollution on labor supply, an understudied outcome thus far in an otherwise quite extensive empirical literature on the health effects of air pollution.[1] Early research on this relationship uncovered negative correlations between air pollution and labor supply in cross-sectional data (Ostro, 1983), repeated cross sections (Hausman et al., 1984), and case studies (Hansen and Selte, 2000). More recent evidence shows that air pollution reduces hours worked among households in Mexico City, Lima, and Santiago de Chile (Hanna and Oliva, 2015; Aragón et al., 2017; Montt, 2018). Yet there is a lack of causal evidence for developed countries where exposure to ambient air pollution is much lower. Another fundamental difference is that employment contracts in developed economies are shaped by rigid labor market institutions and a comprehensive social security system. In such a setting, the typical channel of adjusting labor supply in response to a high-pollution event is by taking a (paid) sick leave. By studying this outcome for workers in Spain, a post-industrial economy with universal sickness insurance, this work fills an important gap in the literature.

Second, our analysis substantially broadens the range of health impacts considered. While the literature has mostly focused on severe health outcomes such as mortality and hospital admissions, our analysis additionally captures all health impairments that workers find troublesome enough to go and seek a sick leave from their general practitioner. Moreover, thanks to having information on the diagnosis, we shed light on the pathology behind the estimated treatment effect by identifying the medical conditions

[1] Previous research has provided credible evidence that air pollution adversely affects health in infants (Chay and Greenstone, 2003b; Currie and Neidell, 2005; Currie et al., 2009b, 2014; Arceo et al., 2016; Knittel et al., 2016) as well as in adults, based primarily on hospital records of births, deaths and emergency-room admissions (Neidell, 2009; Moretti and Neidell, 2011; Graff Zivin and Neidell, 2013; Schlenker and Walker, 2016; Deryugina et al., 2019; Currie et al., 2009b). Another strand of the literature investigates how air pollution affects the productivity of workers and students, and finds negative and statistically significant impacts in a variety of settings (Currie et al., 2009a; Graff Zivin and Neidell, 2012; Chang et al., 2016, 2019b; Ebenstein et al., 2016; Roth, 2016; Lichter et al., 2017; He et al., 2019b).

most affected when pollution levels spike, an aspect that is not yet well understood in the literature. We thus contribute new evidence that helps drawing a more complete picture of the pollution-health gradient.

Third, this chapter contributes first evidence on how the estimated pollution-sick-leave gradient interacts with features of the labor market. This aspect is new to the literature on air pollution because our outcome variable is much more driven by individual choices than severe health outcomes. Empirical evidence on this is needed because it is not evident how behavior interferes with sick leaves as a health outcome. A sick leave can be regarded as both, a health impact and an investment for improving future health. In the labor market studied by us, sick leaves might also be affected by workers pretending to be sick (moral hazard) or pretending to be healthy (presenteeism). We provide first evidence on this by showing that treatment effects systematically vary with idiosyncratic job security. We discuss the implications of this finding for cost-benefit analysis in the short and long run.

Finally, the strong segregation between employment contracts with high and low job security makes the Spanish labor market a prime example of a dual labor market. Previous research has shown that this duality lowers productivity and reduce welfare (Dolado et al., 2002; Cabrales et al., 2014; Bentolila et al., 2019). This chapter contributes to this strand of research by identifying an additional channel of inefficiency in dual labor markets: Since workers at risk of losing their jobs are found to be less likely to take a sick leave during high-pollution events, exacerbated presenteeism could adversely affect future health outcomes and lower productivity in this tier of the labor market.

2.2 Policy background

2.2.1 Sick leave pay in Spain

The vast majority of workers in Spain are entitled to temporary disability benefits. In particular, all affiliates of the social security system continue to be paid when on sick leave, provided that they see a doctor affiliated with the public health care system for treatment and that they have contributed to social security during a minimum contribution period of 180 days in the five years immediately preceding the illness.[2] Unlike in the U.S. where sick pay is granted as a fixed amount of days per year, sick pay in Spain continues either until convalescence or until a permanent disability is diagnosed.

The benefit consists of a daily subsidy, the amount of which is given by the product of a regulatory base and a replacement rate. The regulatory base is the amount that is used to determine the benefits paid by the social security system (*Instituto Nacional de la Seguridad Social*) in an insured event such as a sick leave, a permanent disability caused by a work accident, unemployment, or retirement. For each of those events, the law stipulates which and how many contribution bases must be taken into account to determine the regulatory base of the pertaining benefit.

The contribution base is calculated according to the total monthly remuneration received by the employee. To this end, all wage components including extra payments are prorated to the monthly level such that every worker has exactly twelve contribution bases per year. The replacement rate changes over the course of the sick leave. In case of common illness, no benefit is paid until the fourth day of the leave. The replacement rate corresponds to 60% from day four until day 20 of the sickness spell, and rises to 75% from day 21 onward. The maximum duration of the benefit is twelve months, renewable for another six.[3] The benefits are always paid by the employer. However, from the sixteenth day of a leave, the employer can claim reimbursement of the benefits paid by the social security administration.

In addition to social security benefits, many employers have schemes in place that complement sick pay to provide more complete coverage, especially during the first three days of an illness. Moreover, some collective labor agreements grant matching funds that add to the replacement rate during a temporary disability. As a result, the

[2] No minimum contribution period is required in the case of an accident.

[3] In the case of an accident or occupational disease, the employer pays the first day of the leave in full. After that, social security pays a replacement rate of 75%.

difference between the regular salary and the amount of the disability payment may be small or even nil.[4]

2.2.2 Air quality standards in Europe

To harmonize air quality standards across EU member states, the European Parliament and the Council have passed a series of directives (1999/30/EC, 2000/69/EC, 2002/3/EC, 2004/107/EC) over the past two decades. The most recent "Directive on Ambient Air Quality and Cleaner Air for Europe" (2008/50/EC) establishes limit values for ambient pollutant concentrations measured over different time intervals (hour, day, year). These standards are chosen in accordance to a pollutant's potential to cause health damages in the short and long run. Table 2.2.1 summarizes the standards that apply to air pollutants such as particulate matter smaller than 10 micrometers (PM_{10}), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3).[5] The pollutant of main interest in this chapter, PM_{10}, is subject to two standards. The annual mean concentration must not exceed 40 micrograms per cubic meter (μg/m^3). In addition, the occurrence of daily mean concentrations of 50 μg/m^3 or higher must remain below 36 days per year. Since our econometric identification strategy is based on short-run fluctuations in air pollution concentrations, the analysis below will focus on the 24-hour standard of 50 μg/m^3.

[4] As an example, consider a worker who earns a monthly base salary of € 1,340.54 (before taxes) which amounts to € 44.68 per day. He has been sick at home for 22 days and his collective agreement does not complement the temporary disability benefit. During days one to three of the sick leave, the worker earns € 0. During days four through 15, the company pays a benefit of 60% of the base salary, i.e., €44.68×60%×12 = €321.73. During days 16 through 20, the social security administration pays a benefit of 60%, i.e. €44.68 × 60% × 5 = €134.05. Finally, the benefit paid by the social security administration rises to 75% during days 21 and 22 (2 days), i.e. €44.68 × 75% × 2.

[5] The directive also regulates particulate matter smaller than 2.5 micrometers ($PM_{2.5}$). This pollutant is not considered in the subsequent analysis due to insufficient data.

TABLE 2.2.1: Ambient air quality standards for selected pollutants

Pollutant	Concentration (per m^3)	Averaging period	Legal nature	Exceedances each year
Particulate matter	50 µg	24 hours	Limit	35
(PM_{10})	40 µg	1 year	Limit	-
Sulphur dioxide	125 µg	24 hours	Limit	3
(SO_2)	350 µg	1 hour	Limit	24
Nitrogen dioxide	200 µg	1 hour	Limit	18
(NO_2)	40 µg	1 year	Limit	-
Carbon monoxide (CO)	10 mg	Max. daily 8-hour mean	Limit	-
Ozone (O_3)	120 µg	Max. daily 8-hour mean	Target	25 days averaged over 3 years

Notes: Abridged from European Environment Agency, http://ec.europa.eu/environment/air/quality/standards.htm

2.3 Research design

2.3.1 Sick leaves as a health outcome

Previous research into the effects of air pollution on human health has focused on polar cases. On the one hand, an extensive literature has linked air pollution to severe health outcomes such as morbidity and mortality in adults and infants. On the other hand, a recent strand of the literature has established that air pollution affects humans even under seemingly normal conditions by reducing their productivity at work and in school. Sick leaves can be regarded as an 'intermediate' health consequence that is severe enough to prevent people from following their daily routines while not necessarily leading to dramatic consequences such as hospitalization or death.

Although sick leaves cause substantial economic costs beyond the physical health impact, the literature has not yet investigated this outcome in a systematic way. This chapter provides the first, nation-wide study of the impact of urban air pollution on work absenteeism. Specifically, we study paid sick leaves taken by workers within the context of a publicly-provided sickness insurance scheme. Given the novelty of this outcome variable in the context of environmental valuation, it is important to emphasize two peculiarities of this outcome variable which bear relevance for the interpretation of the results.

First, a sick leave indicates, at the same time, a negative health shock and an investment in future health. Taking a sick leave may help to prevent a hospitalization during the pollution spell, and may contribute to better health outcomes in the future. This effect is reinforced if exposure to pollution occurs mainly at the workplace as the sick leave reduces exposure.

The second peculiarity relates to the fact that, like other insurance schemes, sickness insurance is vulnerable to moral hazard. That is, workers might pretend to be sick and take a paid leave. How to reign in moral hazard in sickness insurance is a question of great policy interest in itself, which has been the subject of a burgeoning empirical literature (e.g. Johansson and Palme, 1996, 2005; Henrekson and Persson, 2004). In our econometric analysis below, we assume that moral hazard gets absorbed into worker and time fixed-effects and hence cannot confound the impact of air pollution on sick leaves. However, moral hazard might play a role in explaining heterogeneous impact estimates. Because employers cannot observe the true health status of a worker, they may take the frequency and length of sick leaves taken as a signal about the worker's health. All else equal, employers prefer to award permanent job contracts to workers with good health. Therefore, workers with a low level of job protection have a stronger incentive to signal good health — e.g. by reducing moral hazard — than workers with strong levels of job protection. We shall test this hypothesis in Section 2.6.2 below.

2.3.2 Empirical model

Our econometric approach focuses on modeling how short-run variation in ambient pollution affects an individual's propensity to take a sick leave.[6] We specify a linear probability model (LPM) for the share of sick days that worker *i* living in city (*munici-*

[6]This is in line with the literature on the impact of pollution on health outcomes, in that it relates ambient pollution concentrations to a binary health outcome.

pio) m takes in week t,

$$\begin{aligned} SICK_{imt} &= \alpha p_{mt} & +w'_{mt}\left[\beta_1 + \beta_2 \odot w_{mt}\right] + h'_{mt}\gamma + \\ &+ & \mu_{m,year(t)} + \lambda_{quarter(t),year(t)} + \\ &+ & \sum_{a=16}^{65} \theta_a \cdot \mathbb{I}\{AGE_{imt} = a\} + \eta_i + \epsilon_{imt} \end{aligned} \tag{2.1}$$

where p_{mt} is a measure of ambient pollution concentrations in city m and week t, w_{mt} is a vector of weather variables in city m and week t, and h_{mt} is a vector containing further city-level controls for school vacations, bank holidays and flu prevalence in week t. Furthermore, the equation includes quarter-by-year effects λ and city-by-year effects μ to control for business-cycle effects and for unobserved local shocks, respectively. As worker-level controls, we include a full set of age dummies and individual fixed effects η_i.

The identifying variation in this regression comes from week-to-week changes in local pollution concentrations and sick leaves within a city and year, after netting out worker-specific effects and correcting for weather as well as other reasons for absence such as business-cycle fluctuations. Inference on the parameters in Equation (2.1) is based on robust standard errors with two-way clustering at the week and city levels.

It is widely held that air pollution is not randomly assigned across space and individuals (Graff Zivin and Neidell, 2013). In Equation (2.1), endogeneity of air pollution might arise for a variety of reasons. First, economic fluctuations that affect both employment and pollution might confound the estimates. For example, an unobserved shock to labor demand might induce both an increase in local pollution while also increasing labor supply (Hanna and Oliva, 2015). Second, to the extent that sick workers cause fewer emissions than they do at work (or on their way to the workplace), the causality might go from sick leaves to air quality. Third, individuals that are more susceptible to adverse health impacts of air pollution might choose to live in less polluted areas. Fourth, pollution exposure is likely measured with error because we use average concentrations rather than individual exposure. All of the above sources of endogeneity would bias the estimated health impact of pollution towards zero. Finally, the estimated pollution impact might also be biased away from zero if it picks up the effect of omitted pollutants that are correlated with the pollutant of interest. The direction of the overall bias is thus ambiguous.

Our research design mitigates concerns about endogeneity in the following ways. First, our focus on sick leaves discards variation from extensive-margin adjustments to labor supply which are not related to a temporary disability. Second, the high-frequency nature of the data allows us to control for a variety of time effects that mitigate simultaneity bias. Third, thanks to the longitudinal structure of the data, we are able to purge the estimates from the effects of locational sorting by individuals or firms. These features help to mitigate some causes of endogeneity, but not necessarily all of them. In particular, if pollution is subject to classical measurement error, an instrumental variable is needed to consistently estimate its causal impact on sick leaves. In the next section, we propose such an approach to instrument for PM_{10}, the pollutant of interest in this study.

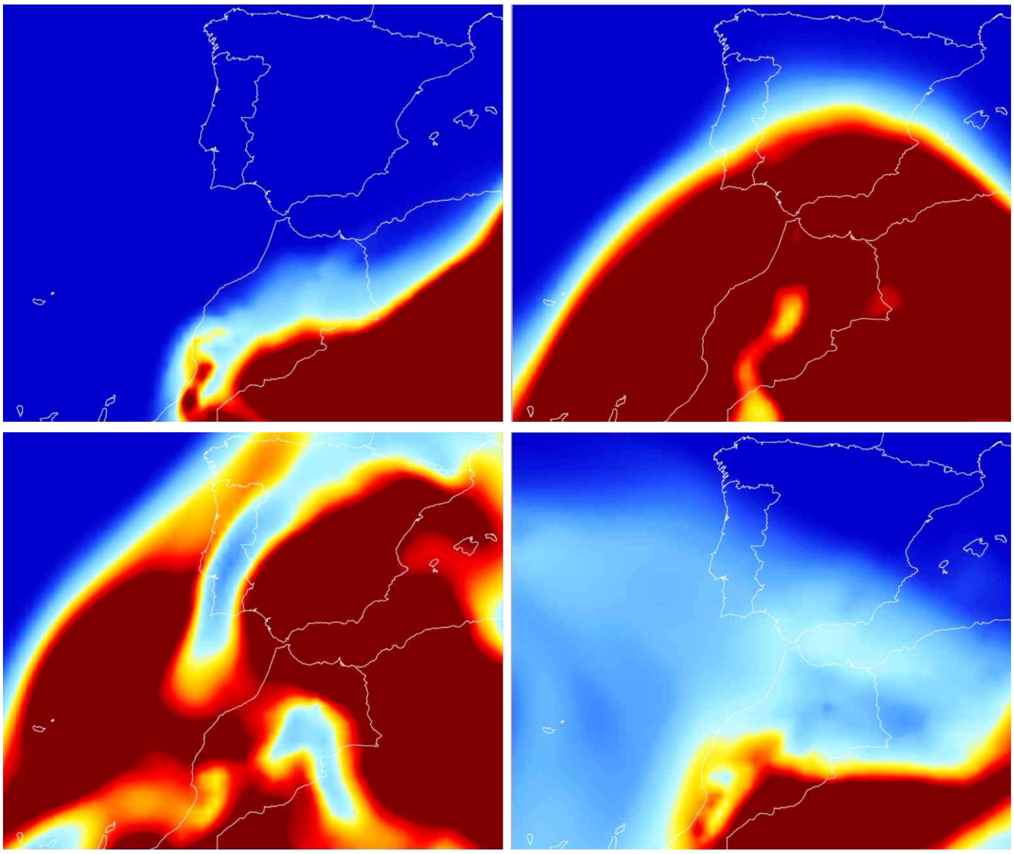
We focus on PM_{10} because no other pollutant exhibited more frequent and more severe violations of the 24-hour limits set by the EU during our study period (cf. Table 2.A.2 below). Consequently, PM_{10} is a primary target for air quality management in urban Spain. Estimating dose-response functions for PM_{10} is necessary for designing efficient pollution control regulations, and it is feasible thanks to long-range atmospheric transport of PM_{10} which shifts ambient concentrations in ways that are conditionally independent of local pollution sources.

2.3.3 Instrumental variable estimation

We address the issue of endogenous pollution in Equation (2.1) within a two-stage least squares (2SLS) estimation framework that exploits quasi-experimental variation in PM_{10} originating from Sahara dust advection. Under certain meteorological conditions, storms in the Sahara desert stir up dust into high altitudes. These dust clouds can travel very long distances and reach European territory several times a year. The arrival of Sahara dust occurs throughout all of Spain, and it is most frequently observed on the Canary Islands, due to their geographical proximity to the Sahara, where the phenomenon is popularly known as 'Calima'. For the sake of brevity, we shall use this term henceforth when referring to episodes of increased PM_{10} concentrations due to long-range transport of African dust. A Calima episode typically lasts several days and

FIGURE 2.3.1: Sahara dust intrusion in Spain

(A) June 9, 10pm — (B) June 12, 4pm



(C) June 14, 9am — (D) June 21, 10am

OD550 Dust

Notes: Visualization of Sahara dust intrusions on the Iberian peninsula during the month of June 2009. Data and images from the BSC-DREAM8b model, operated by the Barcelona Supercomputing Center, available online at: http://www.bsc.es/earth-sciences/mineral-dust-forecast-system/

is accompanied by regional weather patterns that facilitate atmospheric dust transport. For this reason, the distribution of Calima episodes is not uniformly distributed over the year but peaks during the summer months as depicted in Appendix Figure 2.B.1.

Figure 2.3.1 illustrates how Sahara dust traveled across different regions in Spain during a Calima episode in June 2009. On June 9, the dust plume was building up over North Africa (Figure 2.3.1a). Three days later, the plume had extended to cover the Canary Islands and Southern Spain, but not the Balearic Islands (Figure 2.3.1b). On

June 14, all of Spain was exposed to Sahara dust, though the intensity varied across regions (Figure 2.3.1c). Figure 2.3.1d depicts the withdrawal of the dust cloud which was completed at the end of June.

The influence of Sahara dust on ambient PM_{10} concentrations cannot be measured by regular air quality monitors. Therefore, the possible link between ambient pollution and Calima is evaluated *ex post* using data from rural background monitors and meteorological back-tracking models such as the one that generated the data underlying Figure 2.3.1. The scientific procedure (Escudero et al., 2007; Querol et al., 2013) behind this attribution is standardized across EU member states and designed to ensure a level playing field across European cities when determining whether they are in compliance with the EU standard for PM_{10} concentrations. Because Calima events substantially increase non-anthropogenic PM_{10} concentrations, the cities affected by this phenomenon are allowed to discount the measured 24-hour-mean concentration for this effect (see Appendix 2.C for more details). Offical PM_{10} discounts constitute a valid instrument for pollution because they shift local PM_{10} concentrations in ways that are plausibly orthogonal to local conditions that drive sick leaves, after conditioning on weather.

Specifically, denote by c_{mt} the weekly share of days for which the city m's applicable PM_{10} discount is strictly positive (Calima days). The first-stage equation for pollution is given by

$$\begin{aligned} p_{mt} &= \widetilde{\alpha} c_{mt} & &+ w'_{mt} \left[\widetilde{\beta}_1 + \widetilde{\beta}_2 \odot w_{mt} \right] + h'_{mt} \widetilde{\gamma} + \\ &+ & &\widetilde{\mu}_{m,year(t)} + \widetilde{\lambda}_{quarter(t),year(t)} + \\ &+ & &\sum_{a=16}^{65} \widetilde{\theta}_a \cdot \mathbb{I}\{AGE_{imt} = a\} + \widetilde{\eta}_i + \zeta_{imt} \end{aligned} \tag{2.2}$$

where all variables other than c_{mt} correspond to those in Equation (2.1). In the second stage, we estimate the outcome equation after substituting predicted pollution $\hat{p}_{mt}$ from Equation (2.2) for p_{mt} in Equation (2.1).

A crucial assumption underlying this approach is that Calima events have no direct effect on sick leaves other than via increased ambient concentrations of PM_{10}. We flexibly control for local weather to rule out the possibility that particular weather conditions associated with Calima affect the outcome variable. This strategy has proven effective even when the instrument is based on local weather conditions such as wind direction (Hanna and Oliva, 2015; Deryugina et al., 2019) or thermal inversion (Arceo et al., 2016; He et al., 2019b). Since our instrument is based on a more regional weather phenomenon, the potential for confounding weather impacts and pollution appears to be small. The exclusion restriction would also be violated if Calima changed behavior. A precondition for this is that workers are aware of Calima while it lasts. On the Canary Islands, located 1,300 kilometers to the south-west of the Iberian peninsula and just 115 kilometers off the Moroccan Atlantic coast, Sahara dust advection is frequent and sometimes visible, hence awareness must be taken for granted. Although we cannot rule out behavioral responses to Calima there *a priori*, robustness checks discussed in Section 2.5.2 below show that our results are not driven by workers from the Canary Islands. In the rest of Spain, Sahara dust events are less frequent and less intense. We found no evidence of the public being alerted to such events during our study period.

Finally, the exclusion restriction would be violated if PM_{10} originating from the Sahara has a substantially different effect on human health than PM_{10} from local sources. Specifically, if the chemical composition of PM_{10} from the Sahara differed substantially from that of non-desert PM_{10}, we should suspect that their health impacts differ, too. Perez et al. (2008, Fig. 3) compare mass-adjusted concentrations of the four group elements in PM_{10} and find that crustal elements are more frequent during Saharan dust days whereas carbon, secondary aerosols, and marine aerosols show no difference. From a study of Madrid, Barcelona and eleven other southern European cities, Stafoggia et al. (2016, p. 418) conclude that "the health effects of dust-derived PM_{10} are of the same (or similar) magnitude as those reported for anthropogenic sources of air pollution". This lends support to our assumption that the instrumental variable has no direct effect on human health except through raising overall ambient PM_{10} concentrations.

2.4 Data

For the analysis in this chapter, we merge several large datasets that are described in more detail in this section.

2.4.1 Data sources

Employment histories Our primary data come from the National Institute of Social Security which administrates both health insurance and pension benefits for more than 93% of the workforce in Spain. Since 2004, the administration maintains a research dataset, the Muestra Continua de Vidas Laborales, henceforth referred to as the MCVL (Spanish Ministry of Employment, Migration and Social Security, 2018). The MCVL is a non-stratified random sample of anonymized individual work histories, covering approximately 4% of all individuals who were affiliated with social security at some point during the reporting year. An individual record contains information on both current-year and historical employment relations, dating back to the time when the administration began to keep computerized records.

Sick leaves Information on sick leaves taken by social security affiliates is first gathered and processed by the employer's mutual indemnity association which relegates the information back to the social security administration when reimbursements are claimed. While sick leaves are not contained in the MCVL, the social security administration has provided us with a customized dataset that merges individual records of sick leaves to the MCVL during the years from 2005 to 2014. As a result, we have a panel dataset containing daily observations of sick leaves taken by 1.6 million individuals, the diagnosis code based on the International Statistical Classification of Diseases and Related Health Problems (ICD), as well as information on their employment status, wage, age, occupation, and many other characteristics. To the best of our knowledge, we are the first team of researchers to analyze this extraordinary dataset.[7] A caveat is that the dataset is not well-suited to analyze sick leaves taken by unemployed workers. This implies that the analysis to follow has little to say about the impact of air pollution on the unemployed, and on severe health consequences such as permanent disability or death.

Air pollution Data on air pollution were obtained from the Spanish Ministry for the Ecological Transition (2016). These data are updated on a yearly basis and are also used

[7] Alba (2009) and Malo et al. (2012) have used linked MCVL and sick leave data before, but only for a single year.

within the EU framework for reciprocal interchange of information and reporting on ambient air quality (2011/850/EU). The database is comprised of time series data on ambient concentrations of a variety of air pollutants with up to hourly resolution as well as meta-data on monitoring stations. For our analysis, we use readings taken by 784 air quality measurement stations across Spain between January 2005 and December 2014. Apart from location, these stations differ in terms of the set of air pollutants they monitor and the time window of measurement. The vast majority of stations remain active throughout the sample period. The meta-data report the municipality where the measurement station is located. This allows us to link them to construct a dataset of air quality across Spanish cities. When more than one air quality station is located in a municipality, the readings are averaged across stations.

Weather Meteorological data were provided by the Royal Netherlands Meteorological Institute (2019) as part of the European Climate Assessment & Dataset (ECA&D) project. Within the ECA&D project, national meteorological institutes and research institutions from 31 European countries collect daily data on twelve essential climate variables. For Spain, historical information is available from 1896 onward. The number of variables and geographical coverage has been increasing steadily until today. Based on a total of 193 geocoded weather stations, we assign to each municipality the weather conditions at the station that is closest to the municipality's centroid and has non-missing data. Hence, the assigned weather station is not necessarily located within the boundaries of the municipality.

Calima variables We downloaded data on PM_{10} discounts from the website of the Spanish Ministry for the Ecological Transition (2018). The data report daily PM_{10} discounts for 29 locations in Spain. We follow the official procedure and assign to each municipality the closest station with available data (see Appendix 2.C for more details).

Other controls Factors such as epidemics, bank holidays, and school vacations likely affect an individual's propensity to call in sick. To the extent that these factors are correlated with pollution, omitting them from the analysis might result in biased estimates. We thus collected data to control for such factors.

Flu outbreaks are monitored and recorded by the Spanish center for disease control (*Instituto de Salud Carlos III*) under the auspices of its flu surveillance system (*Sistema centinela de Vigilancia de la Gripe en España*). Weekly data for the flu incidence (number of cases per 100,000 inhabitants) are published for each autonomous community,[8] except Galicia and Murcia (Spanish Center for Disease Control, Instituto de Salud Carlos III, 2016).[9] We merge the flu data to our estimation sample at the level of the Autonomous Community. In case of missing observations, data were imputed using the national average.

The dates of school vacations and bank holidays vary at the levels of the autonomous community, province, and even municipality. We gather this information from the various regional "official bulletins" and numerous other sources. The linking is done at the pertinent geographic level.

2.4.2 Data cleaning

Our dataset contains daily records of individual sick leaves between January 1, 2005 and December 31, 2014. The raw sample is comprised of approximately four billion worker-by-day observations over the full sample period. However, some cleaning steps are necessary in order to use the sample for our purposes. This subsection describes and justifies those steps.

First, we drop all workers living in municipalities with less than 40,000 inhabitants. For these workers, the place of residence is reported only at the level of the province, which is too coarse for accurate spatial matching to pollution and weather data. For all remaining workers, we know the place of residence at the five-digit municipality code level which is required for matching. Since these muncipalities have 40,000 inhabitants or more, we shall henceforth refer to them as cities. We retain just over half of the workers in the raw sample after performing this step.

Second, we impose the following sample restrictions. We only keep worker-by-day observations of individuals aged 16 to 65 who are actively employed and for which we have information on employers and wages. Individuals who are reported to have taken

[8] Spain is not a federation, but a decentralized unitary state comprised of 17 autonomous communities and two autonomous cities.

[9] In a normal year, the monitoring is in place during the flu season, i.e. from week 40 until week 20 of the following year. In 2009, year-round surveillance was in place because of the swine flu.

any sick leave of more than 550 days are dropped, as this number exceeds the legal maximum duration. We also remove individuals with reported employment relations after death, negative-length employment durations, as well as duplicate or negative wages. Worker-by-day observations with inflation-adjusted wages in the 99.5th percentile are also excluded.

Third, we drop observations with missing pollution data. In cities where pollution measurements are derived from more than one air quality monitor, failure to account for entry and exit of monitors would lead to incoherent time series. In such cases, we drop all data from monitors reporting less than 120 days of PM_{10} in any reporting year. If this leaves two or more monitors in the data, we require that all monitors report in all years. Finally, we drop all observations on December 31 and January 1 because of the unusually high contamination levels that result from fireworks during the new year's festivities.

Following these cleaning steps, our sample contains between 231 thousand and 263 thousand workers per year who live in 99 cities spread across the Spanish peninsula and the islands (Appendix Figure 2.B.5 displays the location of cities and air quality monitors included in the estimation sample). These cities are home to 55% of the Spanish population and to 51% of all workers affiliated with the general regime of the social security system.

2.4.3 Descriptive statistics

Our sample contains more than half-a-billion daily observations for 466,174 workers aged between 16 and 65 years. To improve computational tractability, we aggregate the data to the weekly level. The first panel of Table 2.4.1 provides summary statistics at the worker level. The average propensity to take a sick leave in a given week is 2.79%. The share of female workers is 46%. Figures 2.B.2 and 2.B.3 in Appendix 2.B plot the duration of sick leaves and the frequencies of the main diagnosis codes, respectively.

The remaining panels of Table 2.4.1 report descriptive statistics on pollution variables and other covariates, gathered at the city-by-week level. The second panel summarizes the data on particulate matter. The average concentration of PM_{10} is 27 $\mu g/m^3$, which is well below the EU annual standard of 40 $\mu g/m^3$, but higher than 20 $\mu g/m^3$, the limit value recommended by the World Health Organization (WHO, 2006). Non-

Table 2.4.1: Descriptive statistics, 2005-14

Variable	**mean**	**sd**	**min**	**max**	**observations**
1. Social security data (workers)					
Age	37.7	11.5	16	65	466,174
Female share [%]	46	49.8	0	100	466,174
Absence rate [%]	2.79	15.9	0	100	⩾ 100 million
2. Particulate matter PM_{10} (city-by-week)					
Ambient concentration [μg per m^3]	26.8	13.0	0.0	188.0	38,613
Concentration due to *Calima* event [μg per m^3]	2.1	5.5	0.0	140.6	38,613
Days PM_{10} exceeds 24-hour standard [%]	7	19	0	100	38,613
Calima days [%]	15	25	0	100	38,613
3. Other air pollutants (city-by-week)					
CO [mg per m^3]	0.5	0.3	0.0	5.2	25,279
SO_2 [μg per m^3]	5.5	4.9	0.0	121.6	29,529
NO_2 [μg per m^3]	25.7	14.7	0.0	140.8	33,421
O_3 [μg per m^3]	73.1	25.8	1.0	176.6	31,305
4. Weather data (city-by-week)					
Temperature [°C]	15.9	6.4	-6.7	36.8	38,613
Wind speed [0.1 m/s]	29.4	15.2	0.0	142.9	38,613
Precipitation [0.1mm]	15.5	32.0	0.0	972.0	38,613
Cloud cover [okta]	3.9	1.8	0.0	8.0	38,613
Sunshine [h]	7.3	3.2	0.0	14.5	38,613
Humidity [%]	66.7	13.5	20.0	100.0	38,613
Pressure [hPa]	1,016.7	5.8	978.3	1,041.2	38,613
5. Flu prevalence (region-by-week)					
Flu rate per 100,000 inhabitants	49.7	89.2	0.0	953.5	11,654

anthropogenic PM_{10} contributes just 2.1 µg/m^3 to this average value. However, the maximum values show that non-anthropogenic PM_{10} matters on high-pollution days. The share of days exceeding the 24-hour standard is 7%, and the share of Calima days is 15%. This means that not every Calima day is a high-pollution day.

The third panel of Table 2.4.1 provides descriptive statistics for other air pollutants. Daily air pollution limits were also exceeded for O_3 concentrations (although those violations were less frequent and less extreme than for PM_{10}) but hardly ever for CO and SO_2 (cf. Appendix Table 2.A.2). For NO_2, annual standards were violated in one sixth of the cases.

The fourth panel summarizes the weather variables. Daily average temperature is measured in degrees Celsius, wind speed in 0.1 meters per second, precipitation in 0.1 millimeters, and cloud cover in integer-valued oktas ranging from 0 (sky completely clear) to 8 (sky completely cloudy). Sunshine is measured in hours per day, humidity in % and pressure in hectopascals. The last panel of the table reports the flu rate, in cases per 100,000 inhabitants.

2.5 Results

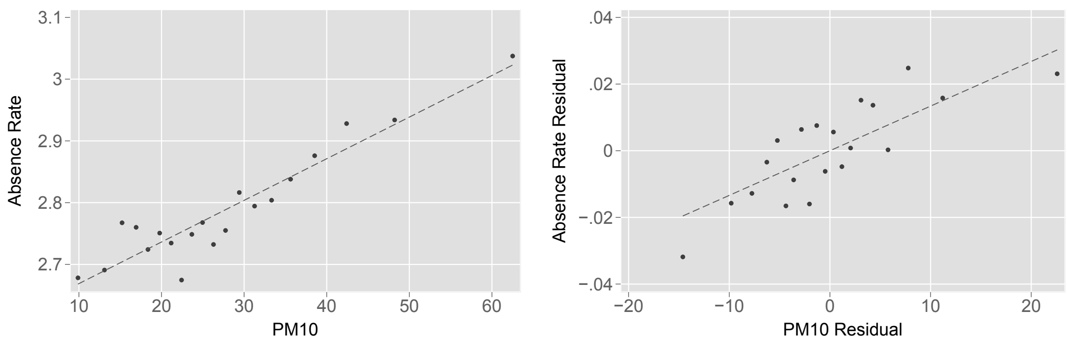
2.5.1 Baseline estimates

Figure 2.5.1 plots weekly absence rates against ambient levels of PM_{10}. Both the raw data and the residualized data from an OLS estimation of Equation (2.1) at the city-level suggest that the relationship is positive and increasing. This is also born out by the estimation results for Equation (2.1). Table 2.5.1 reports OLS and IV estimates for two alternative measures of air pollution.[10] In the first three columns, PM_{10} is measured as the share of days per week on which the daily limit concentration of 50 µg/m^3 is exceeded (PM_{10} exceedance). This identifies the treatment effect of pollution using only high-pollution events. In columns 4 to 6, pollution is measured as average concentrations in µg/m^3. If it was known with certainty that the sick leave response is directly proportional to pollution concentrations, estimating a specification linear in

[10] The estimation of OLS and IV regressions with high-dimensional fixed effects is implemented with the Julia programming language package FixedEffectModels.

FIGURE 2.5.1: Sample correlation between sick leaves and PM_{10}

(A) Binned scatter plot (B) Binned scatter plot residuals



Notes: The figures show binned scatter plots after grouping all observations into 20 bins of equal size based on the variable depicted on the x-axis. The dots represent the mean value for each bin. The dashed line shows the predicted relationship based on an OLS estimation for the underlying data. Subfigure (a) plots the relationship between PM_{10} and weekly absence rates at the city-level. Subfigure (b) plots the relationship of the same variables, after controlling for city-year fixed-effects, year-by-month fixed-effects, weather conditions, school holidays, other public holidays, and flu rates.

PM_{10} would be more efficient as it exploits variation in pollution over the entire support.

The IV regression is implemented as a two-stage-least-squares (2SLS) procedure where local PM_{10} is first regressed on the Calima variable and controls. This regression is reported in columns 3 and 6 of Table 2.5.1 for PM_{10} exceedance and PM_{10}, respectively, and shows that Calima is a strong predictor of ambient PM_{10} concentrations ($R^2 = 0.41$ and $R^2 = 0.65$, respectively). In the second stage, the absence rate is regressed on the predicted PM_{10} variable and controls.

The association between pollution and sick leaves is positive and statistically significant across all specifications, suggesting that higher levels of pollution lead to more sick leaves. The IV estimates exceed the OLS estimates by a factor of 3. This points to attenuation bias that could arise from measurement error in PM_{10} but also due to other sources of bias that were discussed in Section 2.3.3 above. In the analysis to follow, we thus focus on the IV estimator which allows for a causal interpretation of the estimated relationship. The IV estimates of our baseline regressions imply that, on average, a 10-percentage point reduction in the share of exceedances of the limit value reduces the absence rate by 0.0214 percentage points, i.e. by 0.8% of the mean absence rate (0.0279). Furthermore, a reduction in average PM_{10} concentrations by 10 µg/m^3 reduces the absence rate by 0.03 percentage points.

Table 2.5.1: Baseline estimates for PM_{10}

	(1)	(2)	(3)	(4)	(5)	(6)
	Weekly absence rate		PM_{10} exceedance	Weekly absence rate		PM_{10}
PM_{10} exceedance	0.070	0.214				
	(0.016)	(0.063)				
PM_{10}				0.001	0.003	
				(0.000)	(0.001)	
Calima			0.194			15.458
			(0.022)			(1.063)
Estimator	OLS	IV	First-stage	OLS	IV	First-stage
Mean outcome	2.791	2.791	0.085	2.791	2.791	28.038
Observations			100,739,754			
R^2	0.165	0.165	0.413	0.165	0.165	0.648
First-stage F statistic		79.861			211.595	

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

2.5.2 Robustness checks

Dynamics Our main regression equation relates sick leaves to contemporaneous pollution and weather. Previous research has documented that air pollution can have dynamic effects on worker productivity (He et al., 2019b). To investigate this, we estimate alternative specifications of Equation (2.1) which include weekly lags of PM_{10} exceedance and use the respective lags of Calima as instrumental variables. The results, reported in Appendix Table 2.A.3, show that the impact of air pollution on sick leaves can last for up to two weeks. Compared to the specification without lags, the contemporary effect of PM_{10} decreases in magnitude by about 15% but the coefficient on the first lag is almost as large. Further lags do not matter empirically.

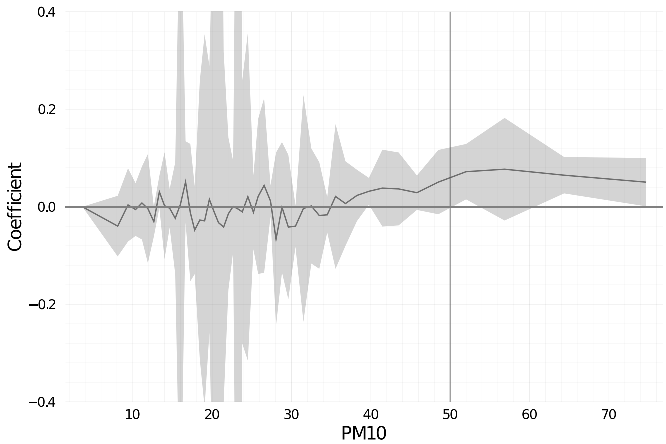
The lag distribution is open to more than one interpretation, however. When interpreted as a dynamic treatment effect, it implies that a one-off shock to pollution affects health in the current and in the next week. The total effect would then be given by the sum of both point estimates which amounts to twice the contemporaneous effect. Yet this could be confounded by high-pollution episodes that last multiple days. For example, a four-day event can fall either into a single week or extend over two subsequent weeks. In the former case, the event contributes only to identification of the coefficient on contemporaneous PM_{10}, but in the latter case the event also helps to identify the coefficient on lagged PM_{10}. Our regression model cannot disentangle the dynamic effects of two-day pollution event in the previous week from the contemporaneous effects of four-day event that extends over both the previous and current week. In the analysis to follow, we thus focus on the contemporaneous effect only. We acknowledge that this might underestimate the full health impact.

Non-linear effects of pollution We shed light on the functional form of the pollution-health relationship by re-estimating Equation (2.1) using alternative thresholds to define a high-pollution event. Appendix Table 2.A.4 presents results for OLS and IV regressions where we use the share of week days with pollution exceeding 50%, 75%, 100% (as in the baseline), or 125% of the legal limit of 50 $\mu g/m^3$ on a given day. The IV point estimates show a clear pattern in that (i) higher threshold values for pollution lead to stronger increases in the propensity to take a sick leave, and (ii) the increment in the sick leave impact increases for constant increments in the pollution threshold. The implication is that a linear functional form, which has often been used in the literature, would misrepresent the underlying pollution-health gradient in our application and would lead us to underestimate the health hazard for the right tail of the pollution distribution. The threshold-based approach we use throughout the remainder of this chapter circumvents this problem, though it averages over the impacts of high and extremely-high pollution days.

To investigate whether pollution levels below the EU 24-hour standard have a significant impact on sick leaves taken, we estimate a non-parametric version of Equation (2.1) where individual-by-week pollution exposures are sorted into 50 bins of equal size.[11] The share of days in the lowest quantile is omitted. The OLS estimates for the different pollution bins are plotted in Figure 2.5.2. The gray area represents the 95% confidence intervals of each indicator. For visual clarity, the 49 estimates and confidence intervals are connected by straight lines. The point estimates are positive and increasing for values of 35 $\mu g/m^3$ and higher, and they become statistically significant at 50 $\mu g/m^3$ just at the limit value.

[11]We classify daily city-level PM_{10} into 50 quantiles after weighing each city-day by the number of workers observed in the MCVL to account for different exposure profiles across cities. To avoid a disproportionately large bin at the top, we drop PM_{10} readings in the 99.9th percentile. The 50 indicators for each bin are then averaged at the weekly level. Hence, the resulting variable indicates the share of days in a given city and week in a particular bin. After matching the indicators to the worker-level data, we estimate Equation (2.1), replacing p_{mt} by 49 indicators.

FIGURE 2.5.2: Non-linear impacts of PM_{10} on sick leaves



Notes: The graph depicts point estimates for different bins of PM_{10} concentrations from an OLS regression of sick leave on indicators for each bin and controls. The vertical line indicates the location of the EU 24-hour standard for PM_{10}. The gray areas indicate 95%-confidence intervals.

Exclusion of Canary Islands The IV estimation rests upon variation in particulate matter due to Sahara dust advection and transport. Because of their geographic proximity to the Sahara, the Canary Islands are subject to this phenomenon more frequently and with greater intensity than the rest of Spain. This could lead to a violation of the exclusion restriction, e.g., if the local population change their behavior in response to Calima. We therefore estimate the main specification after dropping all cities located on the Canary Islands from the sample. Results reported in Appendix Table 2.A.5 show that the Canary Islands have no particular influence on the estimation results.

Non-linear regression model Our baseline specification (2.1) fits a linear probability model to an outcome variable that ranges only between zero and one. The model is thus necessarily mis-specified, but it has the enormous benefit of allowing us to implement high-dimensional fixed-effects and IV estimators in straight-forward ways. Since the realizations of the outcome variable – the share of sick leaves at the city-by-week-level – takes values between 0.38% and 7.02%, we reckon that the bias due to misspecification is not large enough to give up the simplicity and computational tractability of the linear model. We investigate this by estimating a logit model with fractions and worker fixed effects as a non-linear alternative model. This alternative has its own drawbacks as it does not accommodate instrumental variables and, in general, the inclusion of individual fixed effects gives rise to an incidental parameters problem.[12] Results reported in Appendix Table 2.A.6 show that the average marginal effect for the PM_{10} exceedance of 0.060 is smaller than but very close to the OLS baseline estimate of 0.070, while for the PM_{10} in levels, the average marginal effect of 0.001 is almost equivalent to the OLS baseline estimate.

2.5.3 Estimates by ICD-9 diagnosis group

For a subset of absence spells we observe diagnoses, coded according to the Ninth Revision of the International Statistical Classification of Diseases (ICD-9). This classification scheme subdivides diseases and health-related problems into 17 main categories, referred to as chapters. The three most commonly diagnosed ICD-9 chapters in our sample are "VI: diseases of the musculoskeletal system", "V: mental disorders", and "XVII: injury and poisoning", as reported in Appendix Table 2.B.4. Table 2.5.2 reports estimates of 17 IV regressions, using as the dependent variable in Equation (2.1) the chapter-specific absence rate. In those regressions, we drop (i) all spells for which the ICD-9 code is not reported and (ii) all spells with a diagnosis different from the diagnosis group used in the definition of the respective dependent variable, which accounts for the fact that diagnoses are mutually exclusive.[13]

The strongest effect of PM_{10} on sick leaves is found for the diagnosis chapters XVI and V. The former stands for "symptoms, signs and ill-defined conditions" and includes headaches, tachycardia (elevated heart rate), apnea (cessation of breathing), or nausea, among others. The latter (mental disorders) includes, among others, symptoms related to dementia and depression, which have been linked to air pollution in recent research (Bishop et al., 2018; Braithwaite et al., 2019; Fan et al., 2020). A positive effect also emerges for chapter VI (diseases of the nervous and sense organs) which

[12]When the panel is short (T small), noisy estimates of the fixed effects contaminate the estimates of the common parameters and the marginal effects due to the nonlinearity of the model. In particular, the magnitude of the bias of the maximum likelihood estimator would be on the order of $1/T$. However, given that we use weekly data, the average number of observations per worker in our sample is 217.7, suggesting that the incidental-parameter bias will be very small. Simulations by Fernandez-Val (2009) show that the size of the bias is already small when $T = 16$.

[13]Appendix Table 2.A.7 reports the results from an alternative regression where chapter-specific absences are set to zero whenever (i) no diagnosis is reported or (ii) an ICD-9 code does not belong to the diagnosis chapter defining the dependent variable. The results are almost identical to the ones reported in Table 2.5.2.

includes Alzheimer's disease, migraine, and eye disorders. In addition, we find a small but statistically significant negative effect for the chapter "diseases of the blood and blood-forming organs". We refrain from a causal interpretation of this point estimate because it is based on very few observations: Less than 0.3% of all sick leaves are classified in this chapter (vs. 11.6%, 13.5%, and 4.2%, respectively, for the above-mentioned diagnosis chapters; cf. Figure 2.B.4). Interestingly, we do not find statistically significant effects when looking at diagnosis chapters for respiratory diseases (which includes asthma, bronchitis, and chronic obstructive pulmonary disease) or cardiovascular diseases.[14]

[14]We can only speculate about the reasons for this. It could be that patients with asthma or respiratory symptoms are more likely to self-medicate in response to a pollution shock, in particular when they know that they have the disease. Bias could also arise due to the fact that ICD codes are not available for all sick leaves, or due to systematic differences in the reporting quality, depending on whether patients were treated in the emergency room or by their general practicioner.

Table 2.5.2: Effects of PM_{10} exceedance on ICD-9 Diagnosis Groups

ICD-9 Chapter Number and Name		Coefficient	Mean	N
I	Infectious and Parasitic Diseases	0.003	0.067	97,528,520
		(0.007)		
II	Neoplasms	0.005	0.097	97,511,232
		(0.005)		
III	Endocrine, Nutritional and Metabolic Diseases,	0.001	0.019	97,432,211
	Immunity Disorders	(0.001)		
IV	Diseases of the Blood and Blood-forming Organs	-0.001	0.007	97,417,805
		(0.000)		
V	Mental Disorders	0.037	0.302	97,730,323
		(0.010)		
VI	Diseases of the Nervous System and Sense Organs	0.016	0.096	97,523,735
		(0.007)		
VII	Diseases of the Circulatory System	0.000	0.091	97,507,524
		(0.002)		
VIII	Diseases of the Respiratory System	0.003	0.137	97,624,631
		(0.010)		
IX	Diseases of the Digestive System	0.010	0.101	97,535,500
		(0.007)		
X	Diseases of the Genitourinary System	-0.003	0.054	97,474,571
		(0.004)		
XI	Complications of Pregnancy, Childbirth,	0.002	0.080	97,498,058
	and the Puerperium	(0.003)		
XII	Diseases of the Skin and Subcutaneous Tissue	-0.001	0.026	97,441,348
		(0.003)		
XIII	Diseases of the Musculoskeletal System	0.021	0.596	98,073,192
	and Connective Tissue	(0.019)		
XIV	Congenital Anomalies	0.000	0.006	97,416,810
		(0.001)		
XV	Certain Conditions originating	0.000	0.001	97,412,186
	in the Perinatal Period	(0.000)		
XVI	Symptoms, Signs and Ill-defined Conditions	0.043	0.282	97,745,795
		(0.010)		
XVII	Injury and Poisoning	0.018	0.289	97,733,701
		(0.011)		

Notes: Each row reports the coefficient on PM_{10} exceedance estimated in an IV regression where the dependent variable is defined using sick leaves only when the diagnosis is from the given ICD-9 chapter (sick leaves with diagnosis form other chapters are dropped). Coefficients are scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

2.6 Heterogeneity of treatment effects

The richness of our data allows us to investigate the heterogeneity of treatment effects with respect to a wide variety of characteristics of workers and jobs. To do so, we define for each characteristic a categorical variable and split the sample accordingly. In this way, we accommodate heterogeneous reactions not only to air pollution but also to any other variable specified in regression Equation (2.1). First, we analyze heterogeneity with respect to worker characteristics. Next, we study how workers in different occupations react to air quality. We conclude the examination of heterogeneous treatment effects comparing workers with different initial health stocks.

2.6.1 Heterogeneity across workers

Gender When splitting the sample by gender, the point estimates are larger for female workers than for male workers (cf. Appendix Table 2.A.8). A 10-percentage point reduction in the share of PM_{10} exceedances reduces the absence rate by 0.026 percentage points for women and 0.017 percentage points for men. However, taking into account the mean absence rates for each group (3.41% for females, and 2.25% for males), the relative effect is 0.8% in both cases.

Age Appendix Table 2.A.9 reports the regression results after subdividing workers into three age groups of approximately equal size. The point estimates increase with age among young and middle-aged workers. For workers older than 45 years, the increment in the OLS point estimate is less pronounced and the higher mean absence rate in this group implies that the relative impact of high-pollution events is as low as in the group of young workers. When interacting PM_{10} exceedances with worker age we estimate that an additional year increases the pollution impact on absences by 2.8% of the main effect (cf. Appendix Table 2.A.10)

Presence of dependent children Differences in sick-leave taking across workers could be related to the presence of dependent children in the household. We explore this in Appendix Table 2.A.11 which reports the treatment effects estimated separately for workers with and without children under age twelve in the household.[15] We find that the point estimates are very similar in both groups. The differences are well within the margin of error and, relative to the mean absence rates, the treatment effect is around 0.8% in both groups.

This contrasts with the evidence provided by Aragón et al. (2017) that Peruvian workers with dependent children are more likely to reduce their labor supply during high-pollution episodes in order to take care of their children when air pollution makes them sick. While it is plausible that caregivers in our sample stay at home when their child is sick, such leaves are unlikely to be registered as a sick leave taken by the parent. This is because Spanish labor regulations grant parents at least two days of paid leave per year when a minor child is sick. Unfortunately, we cannot investigate this further as we do not have information on such leaves.

Income and skills Differences in income might induce heterogeneity in the treatment effects for a variety of reasons. Workers with higher income may attach a higher value to health, they might have access to better health care and more expensive medication, better options to avoid air pollution, or benefit from a collective agreement that complements sick pay from the social security system. The overall effect of these and other income-related factors on the estimated impact is ambiguous from an ex ante perspective.

The reporting of income is not very precise in our dataset because contribution bases are bottom and top coded. Therefore, we explore the treatment heterogeneity with respect to skill groups that are highly correlated with the salaries that workers received. The Spanish social security system classifies occupations into ten groups, three of which can be considered as high-skilled (Bonhomme and Hospido, 2017). When estimating the treatment effects separately for high-skilled and low-skilled workers, the point estimates, reported in Appendix Table 2.A.12, barely differ.

[15] Age twelve is the threshold used in Spanish labor regulations for granting work-hour reductions or leaves for child rearing.

2.6.2 Heterogeneity across occupations

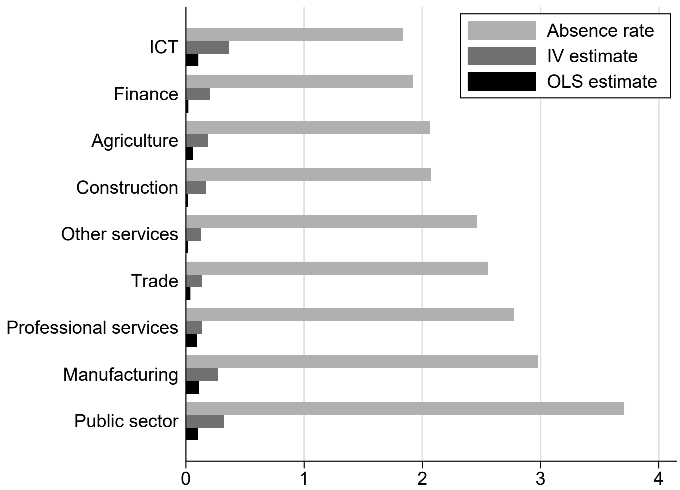
Sector affiliation We begin our investigation of treatment heterogeneity across jobs with a comparison across sectors. To this end, we split up the sample into nine sectors and estimate Equation (2.1) for each sub-sample. Figure 2.6.1 displays the mean absence rates along with the point estimates for PM_{10} exceedance (see Appendix Table 2.A.13 for detailed sector definitions and estimation results).

Three findings emerge from this exercise. First, the estimated impact of high-pollution events on sick leaves is positive in all sectors, and the relationship is statistically significant in all sectors but agriculture. Second, there is meaningful variation in the magnitude of these effects across sectors, with OLS estimates ranging from 0.02 to 0.11 (compared to 0.07 in the full sample) and IV estimates from 0.14 to 0.36 (compared to 0.21 in the full sample). Third, the mean percentage absence rate across sectors is quite heterogeneous and ranges from 1.83 in the information and communication technologies (ICT) sector to 3.71 in the public sector.

Such differences in mean absence rates could arise exclusively because of differences in job characteristics, but in reality they are likely driven also by selection. All else equal, a worker with frail health prefers a job that doesn't expose her to high levels of air pollution or that comes with generous sick leave benefits. Sorting on those characteristics is likely to induce bias in previous, cross-sectional estimates of the impact of air pollution on work days lost (Ostro, 1983; Hausman et al., 1984; Hansen and Selte, 2000). Since we have worker fixed-effects, our estimates are purged of selection bias. Nonetheless, the heterogeneity in the sectoral estimates reflects the joint influence of these and other factors on the decision to take a sick leave. For an illustration, consider the results for the ICT and the public sector. The IV point estimates for these two sectors – 0.364 and 0.321, respectively – are at the top of the range of estimated treatment effects. However, workers in ICT exhibit the lowest average absence rate whereas public sector employees have the highest. This suggests that selection on health cannot alone explain all of the heterogeneity in treatment effects. While the cross-sector comparison of treatment heterogeneities provides clues for the mechanisms driving them, pinning down such mechanisms requires further analysis.

In the remainder of this section, we undertake such an analysis focusing on the role of job security. This is motivated by the fact that both mean absence rates and

FIGURE 2.6.1: Heterogeneous pollution impacts across sectors



Notes: The chart displays mean absence rates, OLS estimates and IV estimates separately for the following sectors: Information and communication technology (ICT), financial services, agriculture, construction, other services, trade, professional services, manufacturing, and the public sector. All coefficients and absence rates have been multiplied by 100 for better readability. The full results are reported in Appendix Table 2.A.13.

treatment effects are particularly high in the public sector, where employees enjoy very high levels of employment protection – in particular, civil servants. Compared to workers with low levels of job security, public sectors workers might be more willing to take a sick leave when experiencing a negative health shock because they do not have to fear any consequences for the continuation of their tenure.[16] In contrast, in sectors where precarious employment conditions are more prevalent, such as agriculture and services, workers might take fewer leaves in order to avoid repercussions on the likelihood of remaining employed. Since we cannot rule out that confounding factors other than employment security drive the cross-sector comparison, the subsequent analyses cut the data in different ways to shed more light on this mechanism.

Employment protection in a dual labor market The Spanish labor market features stark differences in employment protection between temporary and permanent employment contracts. Previous work has shown that this dual labor market affects unemployment, job flows, productivity and welfare (Dolado et al., 2002, 2005; Cabrales

[16] In addition, until July 2012, the replacement rates for the daily subsidy were substantially more generous for workers in the public sector than in other sectors.

TABLE 2.6.1: Treatment effects by contract type

	(1)	(2)	(3)	(4)	(5)	(6)
	Temporary contract		Permanent contract		Civil servants	
PM_{10} exceedance	0.066	0.155	0.062	0.203	0.135	0.505
	(0.026)	(0.055)	(0.018)	(0.072)	(0.058)	(0.186)
Estimator	OLS	IV	OLS	IV	OLS	IV
Mean outcome	2.238	2.238	2.850	2.850	4.321	4.321
Observations	25,786,038		67,323,069		7,055,193	
R^2	0.191	0.191	0.181	0.181	0.180	0.180
First-stage F statistic		85.831		78.619		66.814

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

et al., 2014; Bentolila et al., 2019). It is conceivable that differences in employment protection could affect the generosity of sick leave benefits (e.g., at the level of collective bargaining) or the worker's willingness to take advantage of those benefits.

We investigate this by splitting the sample into three groups with increasing degree of employment protection: temporary contracts, permanent contracts, and civil servants. The results for each group are reported in Table 2.6.1. The IV point estimates imply that, on average, a 10-percentage point reduction in the share of exceedances of the limit value reduces the absence rate by 0.0155 percentage points for workers with a temporary position, and 0.0203 percentage points for workers with a permanent contract. Given the mean absence rates for each group (2.2% for temporary, and 2.9% for permanent), the reaction is similar for both groups (0.7%). However, the implied reaction for civil servants is larger both in absolute terms (0.0505 percentage points) and relative to the mean (1.2%), corroborating the view that the propensity to take a sick leave in this group is not only higher on average but also in response to pollution shocks.

Unemployment risk Next, we use rich data on individual-level attributes in order to tease apart the potential effect of unemployment risk from other factors driving a worker's decision to take advantage of sick leave benefits when pollution spikes. Unemployment risk might matter for the following reason. If a worker fears that the frequency of her sick leaves will be taken into account when her employer decides on

TABLE 2.6.2: Treatment effects by unemployment risk - interaction

	(1)	(2)
	Weekly absence rate	
PM_{10} exceedance	0.069	0.213
	(0.016)	(0.063)
PM_{10} exceedance × unemployment risk	-0.020	-0.091
	(0.010)	(0.032)
Estimator	OLS	IV
Mean outcome	2.798	2.798
Observations	100,199,483	
R^2	0.165	0.165
First-stage F statistic		34.881

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Standardized within-sector unemployment risk is also interacted with weather variables, city-year dummies and year-quarter dummies. Robust standard errors in parentheses are clustered by city and by week.

whether or not to retain her in the event of a mass layoff, it may be rational for her to not take a sick leave even when sick. This is a testable prediction because we can estimate individual unemployment risk from our data and relate this to the propensity to take a sick leave.

In order to predict individual unemployment risk we fit a logit model for the probability of losing a job in a given month for each of the nine sectors above. All regressions control for gender, year, age, age squared, tenure, tenure squared, experience, experience squared, nationality, and province of residence.[17] Using fitted regressions for each sector, we predict the individual unemployment risk for the first week in which a worker is observed. We then assign workers to one of two groups, depending on whether this prediction lies above or below the sector's median risk. The assignment is made once for the worker's entire time series. In so doing, we capture time-invariant heterogeneity in a worker's risk of job loss while avoiding possible feedbacks of air quality on employment risk. We derive two insights from this classification (summa-

[17]This approach is inspired by the labor literature. For example, Card (1996) uses predicted wages as a proxy for skills in order to study the impact of unions on workers with different skill levels. As the MCVL contains only limited information on unemployment status, we use missing employment entries in the dataset as proxies for job loss. Since some workers exit the MCVL sample at the end of the year, we include a dummy for the month of December which allows us to control for this effect when predicting job loss. See Appendix 2.D for details.

rized in results reported in Appendix Table 2.A.14). First, the average propensity to take a sick leave is lower among those facing a higher risk of losing their job. Second, the increase in this propensity in response to a high-pollution event is higher in the low-risk group than in the high-risk group, an effect that holds in both relative and absolute terms.

We further analyze this using continuous variation in predicted unemployment risk. Table 2.6.2 reports regression results from a variant of Equation (2.1) that includes the interaction of PM_{10} exceedance with the individual unemployment risk, standardized relative to other workers in the same sector. The IV coefficient implies that a one-standard deviation decrease in individual unemployment risk is associated with an impact of pollution on sick leaves that is 43% higher.

Our findings imply that estimates of the health damages of air pollution derived from sick leave data may be underestimating the true health effects if workers in jobs with low job security go to work despite being sick. In Section 2.7 we estimate the magnitude of this divergence for all of Spain.

2.6.3 Heterogeneity with respect to health status

A long-standing interest in research on the pollution-health gradient has been with the impact on particularly vulnerable individuals. In fact, much of the literature approaches the topic with a focus on vulnerable populations such as infants (Currie et al., 2014) or the elderly (Deryugina et al., 2019).

Thanks to having rich data on workers' health records, we can go beyond the current state of the literature and identify vulnerable individuals in a representative sample of workers aged 16 to 65. To this end, we first construct a straightforward indicator of health as the share of days missed due to the sick leaves taken during the first twelve months that a worker is observed in the sample. We then estimate a variant of Equation (2.1) where PM_{10} exceedances are interacted with this health measure.[18]

Table 2.6.3 reports the results. We find that the interaction term is positive and statistically significant in the IV estimation but not for OLS. Two statistics from the distribution of the absence share are helpful to further interpret these results. First, the

[18]In this estimation, we drop the first 18 months of a worker's record to ensure that the spells used to define vulnerability do not extend into the estimation period (a sick leave can last up to six months) and thereby have a direct effect on the estimation results.

Table 2.6.3: Treatment effects by health status

	(1)	(2)
	Weekly absence rate	
PM_{10} exceedance	0.074	0.157
	(0.019)	(0.084)
$\times$ absence share	0.579	2.488
	(0.354)	(0.710)
Estimator	OLS	IV
Mean outcome	2.915	2.915
Observations	74,706,090	
R^2	0.176	0.176
First-stage F statistic		38.477

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

healthiest 76.5% of workers have an absence share of zero. Hence this group responds to high-pollution events according to the main effect. Since the main effect is positive and significant in the OLS regression, but not the interaction effect, attenuation bias due to measurement error seems to be particularly severe for workers with poor health.

Second, we define as vulnerable workers the ones in the top five percent of the absence share distribution who missed at least 12.3% of work days due to illness. The IV estimates imply that the impact of high-pollution events on vulnerable workers is about 3 times the effect on healthy workers. Therefore, the mean effect masks large effects for particularly vulnerable individuals, a result that has been established thus far only for populations either younger or older than the working-age subjects we consider in our analysis. In further results reported in Appendix Table 2.A.15, we find that the discrepancy between healthy and vulnerable workers is weaker when defining vulnerability based on the length of previous sick leaves.[19] This is consistent with the interpretation that especially workers with chronic conditions but not excessively long sickness spells are harmed by air pollution.

[19] In particular, the effect size ratio between lower 75% and top 5% of the distribution is 2.5 when the health indicator is based on the maximum length and 1.2 when it is based on the mean length.

Our results complement the available evidence to support the claim that air pollution imposes a disproportionately large health burden on vulnerable groups of society. Our focus on sick leaves and workers also highlights the perhaps less appreciated fact that universal sickness insurance with a benefit scheme for temporary disability helps to alleviate unequal distributional impacts of air pollution by granting affected workers both access to treatment and time to heal.

2.7 Aggregate benefits of air quality improvements

The estimation results allow us to compute a lower bound on the benefits of improving urban air quality in Spain. We proceed in two steps. First, we calculate the reduction in sick days caused by a specific improvement in air quality. To translate this into a monetary benefit, we then multiply this number by the average daily producer wage across workers. Under the assumption that workers are paid their marginal product, this approximates the value of incremental production enabled by the reduction in sick days. While improving air quality yields sizable additional benefits by reducing mortality, human suffering, and medical treatment costs, we focus on foregone production because this component of the social costs of air pollution is directly linked to our outcome variable and has not yet been quantified in previous research.

We start by considering a counterfactual intervention that enforces strict compliance with the 24-hour standard for PM_{10} throughout the sample period. The average share of worker days on which PM_{10} exceeded the EU limit value was 8.11% between 2005 and 2014. The IV point estimate implies that reducing this share to zero would lower the absence rate by 0.017 percentage points (0.621% of the mean). Specifically, we calculate the increase in work attendance as

$$\hat{\Delta} = 0.0811 \cdot 7 \cdot 0.00214 \cdot 6{,}867{,}199 = \text{ 8,343 worker days per week}$$

where the product of the first two terms yields the expected reduction in high-pollution days per week, the third term is the IV coefficient and the last term is the average number of social security affiliates in our sample of 99 cities over the period. We multiply this number with the daily producer wage in constant 2018 Euros, averaged across

TABLE 2.7.1: Productivity benefits of reducing PM_{10} exposure in urban Spain

	I. Enforcement of 24-hour standards at				II. Reduced exposure	
	$50\ \frac{\mu g}{m^3}$		$37.5\ \frac{\mu g}{m^3}$		to > $50\ \frac{\mu g}{m^3}$ since 2005	
	Weekly absence [days]	Annual production [€ million]	Weekly absence [days]	Annual production [€ million]	Cumulative absence [days]	Cumulative production [€ million]
Baseline	-8,343 ± 4,814	39.49 ± 22.79	-13,254 ± 8,204	62.74 ± 38.83	-5.58m ± 3.22m	505 ± 291
Adjusted for job security	-11,657 ± 6,266	55.18 ± 29.66	-18,636 ± 10,938	88.21 ± 51.78	-7.80m ± 4.19m	706 ± 379

Notes: Panel I reports counterfactual productivity benefits of strictly enforcing 24-hour standards for PM_{10} in terms of reductions in (i) work days lost per week and (ii) incremental production. Counterfactual air quality improvements are calculated relative to the observed PM_{10} concentrations in each city contained in the sample, weighted by the number of social security affiliates in that city. Enforcing a limit of 50 µg/m^3 is binding on 8.11% of worker days in the sample whereas a limit of 37.5 µg/m^3 is binding on 20.73% of days. Panel II reports the cumulative benefits of reductions in worker exposure to daily PM_{10} concentrations of 50 µg/m^3 or more that have actually occurred in urban Spain since 2005.

workers and years[20] to obtain annualized benefits of

$$\hat{\Omega} = 52 \cdot \hat{\Delta} \cdot €91.0 = €39.5\text{m}.$$

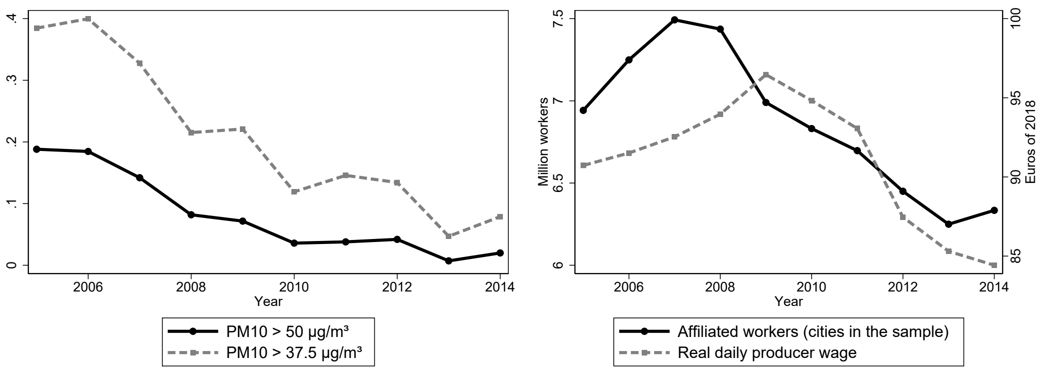
To account for sampling error we include 95% confidence bands for all counterfactual calculations reported in Panel I of Table 2.7.1.

Our IV estimate is based on a binary indicator for high-pollution events. In the above exercise, this implies an assumption that PM_{10} concentrations below 50 µg/m^3 are not harmful to human health. Yet the non-parametric OLS estimates displayed in Figure 2.5.2 suggest that pollution increases absence rates even at lower PM_{10} concentrations of less than 40 µg/m^3. We thus repeat the calculations for a scenario that enforces strict compliance with 75% of the PM_{10} limit value throughout the sample period. This new standard would be binding for 20.73% of worker-days. Using the IV estimate for exceedances above 37.5 µg/m^3 (reported in Table 2.A.4), we compute a reduction in the absence rate of 0.027 percentage points and an increase in work attendance by 13,254 days per week. The annual benefits in terms of foregone produc-

[20] Averaging wages and employment across years is intended to dampen the impact of economic fluctuations on the benefit estimates. As shown by Figure 2.7.1 below, such fluctuations were substantial during the ten-year sample period.

FIGURE 2.7.1: Trends in Pollution, Employment, and Wages 2005-2014

(A) Share of high-pollution days (B) Employment and producer wages



Notes: Figure a) displays the share of worker days with PM_{10} concentrations exceeding the EU 24-hour limit of 50 µg/m^3 (solid line) and the share of worker days with concentrations in excess of 75% of the limit value (dashed line). The figure is based on our sample of 99 Spanish cities with at least 40,000 inhabitants. PM_{10} concentrations are weighted by the number of social security affiliates in each city and year. Figure b) displays the number of workers affiliated with the General Social Security Regime on our sample of 99 Spanish cities with at least 40,000 inhabitants (solid line) and the daily producer wage, expressed in constant 2018 Euros (dashed line).

tion amount to € 62.74 million, which is a 59% increase compared to merely enforcing the EU limit value. Hence, the additional benefits of improving air quality beyond the current EU standard for PM_{10} would be economically significant.

According to our results, high-pollution events cause an increase in sick leaves, but this effect is significantly lower among workers facing a high risk of job loss relative to the sector average. If those workers go to work despite being sick we underestimate the burden of air pollution. To gauge how sensitive our benefit estimates are to such presenteeism, we repeat the above calculations after eliminating idiosyncratic unemployment risk. That is, we replace each worker's estimated unemployment risk by the smallest risk observed in the sector and predict the sick-leave response based on the IV estimates reported in Table 2.6.2 which allow for differential responses to pollution depending on unemployment risk. The results for the two counterfactual scenarios are reported in the bottom row of Table 2.7.1. This exercise shows that correcting for presenteeism leads to benefit estimates that are 40% higher than in the baseline.

With these concepts in mind, we turn to evaluating the benefits of actual air quality improvements that have taken place in urban Spain between 2005 and 2014. Figure 2.7.1a shows a strong decline in worker exposure to PM_{10} concentrations exceeding the EU 24-hour limit (or 75% of that limit). In our sample of 99 cities, the share of worker-weighted days with PM_{10} concentrations in excess of 50 μg/m^3 declined from 18.8% in 2005 to just under 2% in 2014.[21] The cumulative benefit of this development is a reduction in work days lost by 5.58 (± 3.22) million which would have resulted in foregone production worth € 505 million (± € 291 million). We calculate these numbers by computing for each year from 2006 onwards the reduction in PM_{10} exposure relative to 2005 and summing the associated benefits across years in the sample. It is important to note that this calculation is affected quite strongly by major economic fluctuations that occurred during the sample period, depicted in Figure 2.7.1b. The years 2005 to 2007 were the final years of a massive construction boom in Spain, with strong growth in both employment and producer wages. The financial shock and grand recession of 2008 triggered a prolonged economic crisis in Spain which caused dramatic declines in those variables from 2009 onward. If we used 10-year averages of employment and wages to compute benefits as in the counterfactual simulations above, we would likely overestimate the value of foregone production because air quality improvements were largest during the crisis years. We account for this by using annual values.

As above, one can adjust the estimated benefits of air quality improvements to account for presenteeism in the impact estimates. This yields a reduction in work days lost by 7.80 (± 4.19) million and a corresponding increase in production worth € 706 million (± € 379 million). Irrespective of which number one prefers, this exercise shows that the productivity-related benefits of air quality improvements that occurred in Spain between 2005 and 2014 were both economically and statistically significant.

2.8 Conclusion

We have conducted the first nation-wide study of the impact of air pollution on work absenteeism. Our application is based on a representative panel of urban workers affiliated with the Spanish social security system. Using naturally occurring variation in Sahara dust advection to instrument for ambient concentrations of PM_{10} we have estimated causal relationships that give rise to a number of policy implications.

First, higher PM_{10} concentrations lead to more work absences, in particular for levels of pollution that exceed the EU-mandated 24-hour limit value of 50 μg/m^3. This

[21] This improvement is in line with a broader pattern observed in the G7 countries, where the mean population exposure to $PM_{2.5}$ fell by 35% between 1990 and 2017 (OECD, 2019).

effect is economically significant even when sick leaves are valued only with the foregone production during work days lost, a lower bound on the full social costs which also include cost components related to loss of life, human suffering, and medical treatment. The implication is that policies aimed at reducing particulate matter have non-negligible health benefits that are not accounted for in a sizable literature focusing on severe morbidity and mortality outcomes. Our findings also underline that air pollution damages human health for a broad range of symptoms and diagnosis chapters.

Second, our findings corroborate that air pollution control is needed in particular to protect vulnerable members of society, a policy implication first derived in previous research on infants and the elderly. Our analysis has contributed new evidence for the working population by showing that workers in bad health suffer disproportionately from air pollution. From a public policy perspective, protecting the vulnerable is imperative on the grounds that there are no good substitutes for health. In such a setting, monetary compensation is a blunt instrument for mitigating unequal welfare effects of environmental externalities.

Third, we establish that workers who suffer an adverse health shock due to air pollution are less likely to take a sick leave the higher their individual risk of job loss. As a consequence, such workers might fail to seek adequate treatment for their diseases. This would likely decrease their productivity in the short run and might also have detrimental health consequences in the long run.

We note several caveats. Our study is based on data from the social security system which covers most but not all workers in Spain. Our results may thus not be representative of self-employed, unemployed and permanently disabled workers. Further data limitations imply that our results may not be representative for workers in rural parts of the country.

A cautionary note is due in regards to the use of OLS vs. IV estimates when deriving policy implications. Although our use of high-dimensional fixed effects is well-suited to control for non-random assignment of air pollution in an OLS regression, mismeasured pollution exposure likely causes attenuation bias in our application. Our discussion of the results has predominantly relied on the IV estimates because they mitigate this bias. However, their validity requires that PM_{10} originating from Sahara dust advection affects sick leaves only by shifting local PM_{10} concentrations, conditional on weather and other controls. Based on the evidence available to us, we have argued that violations of the exclusion restriction are unlikely to drive our results. We note, however, that our interpretation of the IV coefficient as a dose-response relationship for urban PM_{10} concentrations is based on the premise that the pathological effects of PM_{10} blown in from the Sahara are identical to those of urban PM_{10}. Previous research on this topic documents small but measurable differences in the chemical and mineralogical composition of those variables (Perez et al., 2008), but also suggests there are no differential health impacts (Stafoggia et al., 2016). We are aware that this does not constitute ultimate proof of identical health effects. Nonetheless, it lends empirical support to an assumption that allows us to derive valuable, additional insights from our empirical results. In making this assumption, we trade-off the remaining scientific uncertainty against the benefits of having point estimates that are not only purged of attenuation bias but also interpretable as dose-response relationships for PM_{10}.[22] Should future research produce new evidence that makes the exclusion restriction untenable, this would call for a reassessment of this trade-off and might suggest the adoption of an alternative IV strategy. Even in such a scenario, however, our empirical framework still provides intent-to-treat estimates suitable for evaluating work days lost due to long-range transport of desert dust, which might become more relevant as the problems of global climate change and desertification worsen.[23]

A final caveat to be addressed here concerns the measurement of foregone production due to sick leaves. We have proposed two alternative measures. The first measure only counts sick leaves that were actually taken. This conservative approach likely underestimates the true value of foregone production because it fails to account for the detrimental effects of presenteeism on productivity in the short run (sick workers perform worse on the job; Neidell, 2017), in the medium run (through delayed convalescence), and in the long run (not treating minor conditions can lead to more severe health conditions, as is the case with asthma). Our second measure additionally counts sick leaves that would have been taken if all workers had the maximum level of job protection. This addresses the issue of presenteeism, but it might entail overestimation of work days lost if workers with high job protection are more prone to moral hazard (e.g., because high job security lowers the cost of shirking). Whether or not

[22]The latter benefit would not arise if we used alternative IVs suggested in the literature which shift several air pollutants at a time.

[23]We report intent-to-treat estimates for the main specifications in Appendix Table 2.A.16.

this matters empirically depends on the margin of shirking. If moral hazard affects the extensive margin, i.e., through more false sick leaves, this should not affect the treatment effect we estimate because the higher baseline absence would be absorbed by worker fixed-effects. In contrast, if workers are able to shirk by taking more sick days for each sick leave, this would provide a competing explanation for observing stronger sick-leave responses to air pollution in jobs with low employment risk. Disentangling moral hazard and presenteeism is beyond the scope of this chapter and left as a topic for future research.

Overall, our study has shown that administrative data on sick leaves is very useful for closing an important gap in the empirical valuation of the health impacts of air pollution. We expect that researchers will continue to use this valuable data resource in future work, and find effective ways of dealing with the above-mentioned issues that arise with this new outcome variable.

Appendix

2.A Additional Tables

Table 2.A.1: Correlation of pollution measures

	PM_{10}	SO_2	CO	O_3	NO_2
PM_{10}	1				
SO_2	0.212	1			
CO	0.383	0.283	1		
O_3	-0.119	-0.201	-0.353	1	
NO_2	0.439	0.191	0.503	-0.479	1

Table 2.A.2: Distribution of days per week by pollutant and percent of limit value

	PM_{10}	CO	SO_2	NO_2	O_3
$[50; 75)$	26.2	0.0	0.1	22.9	41.0
$[75; 100)$	11.5	0.0	0.0	14.9	23.3
$[100; 125)$	4.3	0.0	0.0	8.8	4.2
$[125, \infty)$	3.1	0.0	0.0	9.0	0.3

Notes: Each column reports the percentage share of days per week with ambient concentrations for different quartiles of the daily limit value stipulated by the EU. Limit values refer to either 24-hour averages (for PM_{10} and SO_2) or maximum 8-hour averages (for CO and O_3). For NO_2, the bins refer to daily averages evaluated against the annual limit value, as the EU has not defined a daily limit value. All limit values are reported in Table 2.2.1.

TABLE 2.A.3: Estimates for PM_{10} with lags

	(1)	(2)	(3)
	Weekly absence rate		
PM_{10} exceedance	0.203	0.172	0.178
	(0.063)	(0.061)	(0.063)
Lag: PM_{10} exceedance		0.153	0.146
		(0.066)	(0.064)
Lag 2: PM_{10} exceedance			0.030
			(0.068)
Estimator	IV	IV	IV
Mean outcome	2.785	2.785	2.785
Observations		97,999,958	
R^2	0.165	0.165	0.165
First-stage F statistic	77.619	52.400	0.794

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

TABLE 2.A.4: Non-linear effects of PM_{10}

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weekly absence rate							
PM_{10} >50%	0.025	0.112						
	(0.010)	(0.042)						
PM_{10} >75%			0.058	0.131				
			(0.013)	(0.042)				
PM_{10} >100%					0.070	0.214		
					(0.016)	(0.063)		
PM_{10} >125%							0.053	0.371
							(0.026)	(0.107)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Mean outcome	2.791	2.791	2.791	2.791	2.791	2.791	2.791	2.791
Observations				100,739,754				
R^2	0.165	0.165	0.165	0.165	0.165	0.165	0.165	0.165
First-stage F statistic		131.104		252.482		79.861		38.481

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

TABLE 2.A.5: IV estimates for PM_{10} excluding Canary Islands

	(1)	(2)	(3)	(4)
	Weekly absence rate			
PM_{10} exceedance	0.073	0.209		
	(0.016)	(0.067)		
PM_{10}			0.001	0.003
			(0.000)	(0.001)
Estimator	OLS	IV	OLS	IV
Mean outcome	2.769	2.769	2.769	2.769
Observations	98,057,059			
R^2	0.165	0.165	0.165	0.165
First-stage F statistic		71.840		192.776

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by municipality and by week.

TABLE 2.A.6: Logit estimates for PM_{10}

	(1)	(2)
	Weekly absence rate	
PM_{10} exceedance	0.026	
	(0.008)	
	[0.060]	
PM_{10}		0.0004
		(0.0002)
		[0.0009]
Estimator	Logit	Logit
Mean outcome	2.791	2.791
Observations	100,739,754	
Efron's R^2	0.169	0.169

Notes: Coefficients scaled by a factor of 100 for better readability. Mean marginal effects evaluated at the mean in brackets, multiplied by 100 to be comparable to the linear regressions. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors for the coefficients in parentheses are clustered by city and by week.

Table 2.A.7: Effects of PM_{10} on ICD-9 Diagnosis Groups

ICD-9 Chapter Number and Name		Coefficient	Mean
I	Infectious and Parasitic Diseases	0.003	0.065
		(0.006)	
II	Neoplasms	0.004	0.094
		(0.004)	
III	Endocrine, Nutritional and Metabolic Diseases, Immunity Disorders	0.001	0.019
		(0.001)	
IV	Diseases of the Blood and Blood-forming Organs	-0.001	0.007
		(0.000)	
V	Mental Disorders	0.037	0.294
		(0.009)	
VI	Diseases of the Nervous System and Sense Organs	0.015	0.093
		(0.006)	
VII	Diseases of the Circulatory System	-0.000	0.088
		(0.002)	
VIII	Diseases of the Respiratory System	0.001	0.134
		(0.009)	
IX	Diseases of the Digestive System	0.009	0.098
		(0.006)	
X	Diseases of the Genitourinary System	-0.003	0.053
		(0.004)	
XI	Complications of Pregnancy, Childbirth, and the Puerperium	0.002	0.077
		(0.003)	
XII	Diseases of the Skin and Subcutaneous Tissue	-0.001	0.025
		(0.003)	
XIII	Diseases of the Musculoskeletal System and Connective Tissue	0.017	0.581
		(0.018)	
XIV	Congenital Anomalies	0.000	0.006
		(0.001)	
XV	Certain Conditions originating in the Perinatal Period	0.000	0.001
		(0.000)	
XVI	Symptoms, Signs and Ill-defined Conditions	0.041	0.274
		(0.010)	
XVII	Injury and Poisoning	0.017	0.281
		(0.011)	
Observations		100,739,754	

Notes: Each row comes from a separate regression. Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by municipality and by week.

Table 2.A.8: Treatment effects by gender

	(1)	(2)	(3)	(4)
	Female		Male	
PM_{10} exceedance	0.077	0.259	0.062	0.173
	(0.028)	(0.088)	(0.016)	(0.053)
Estimator	OLS	IV	OLS	IV
Mean outcome	3.410	3.410	2.249	2.249
Observations	46,995,784		53,743,967	
R^2	0.165	0.165	0.163	0.163
First-stage F statistic		76.222		82.814

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

Table 2.A.9: Treatment effects by age

	(1)	(2)	(3)	(4)	(5)	(6)
	≤ 34		> 34 & ≤ 45		> 45	
PM_{10} exceedance	0.052	0.143	0.078	0.218	0.080	0.277
	(0.018)	(0.049)	(0.020)	(0.070)	(0.034)	(0.105)
Estimator	OLS	IV	OLS	IV	OLS	IV
Mean outcome	2.045	2.045	2.485	2.485	3.899	3.899
Observations	35,276,558		32,718,271		32,744,309	
R^2	0.141	0.141	0.179	0.179	0.209	0.209
First-stage F statistic		82.294		82.243		73.378

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

TABLE 2.A.10: Treatment effects interacted with age

	(1)	(2)
	Weekly absence rate	
PM_{10} exceedance	0.070	0.212
	(0.017)	(0.065)
PM_{10} exceedance × age	0.002	0.006
	(0.001)	(0.003)
Estimator	OLS	IV
Mean outcome	2.791	2.791
Observations	100,739,754	
R^2	0.165	0.165
First-stage F statistic		37.167

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

TABLE 2.A.11: Treatment effects by household type

	(1)	(2)	(3)	(4)
	No kids below 12		Kids below 12	
PM_{10} exceedance	0.069	0.230	0.086	0.176
	(0.017)	(0.067)	(0.024)	(0.075)
Estimator	OLS	IV	OLS	IV
Mean outcome	2.866	2.866	2.561	2.561
Observations	75,836,283		24,902,425	
R^2	0.184	0.184	0.182	0.182
First-stage F statistic		77.394		86.277

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

Table 2.A.12: Treatment effects by skill level

	(1)	(2)	(3)	(4)
	High skilled		Low skilled	
PM_{10} exceedance	0.076	0.226	0.066	0.205
	(0.021)	(0.084)	(0.019)	(0.064)
Estimator	OLS	IV	OLS	IV
Mean outcome	2.022	2.022	3.042	3.042
Observations	24,822,994		75,916,027	
R^2	0.150	0.150	0.171	0.171
First-stage F statistic		70.057		82.737

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

TABLE 2.A.13: Treatment effects by sector

	(1) Agricultural	(2) Manufacturing	(3) Construction	(4) Trade	(5) ICT	(6) Finance	(7) Professional serv.	(8) Public	(9) Other services
PM_{10} exceedance (OLS)	0.062	0.111	0.019	0.038	0.105	0.021	0.095	0.099	0.019
	(0.164)	(0.029)	(0.028)	(0.021)	(0.117)	(0.215)	(0.031)	(0.038)	(0.407)
PM_{10} exceedance (IV)	0.183	0.271	0.171	0.135	0.364	0.201	0.138	0.321	0.125
	(0.232)	(0.120)	(0.077)	(0.077)	(0.173)	(0.051)	(0.071)	(0.115)	(0.053)
Observations	786,455	13,151,585	7,833,216	28,884,828	4,099,741	4,329,242	15,294,028	21,820,250	4,530,216
Mean outcome	2.062	2.976	2.074	2.554	1.834	1.920	2.776	3.708	2.458
R^2	0.272	0.186	0.194	0.173	0.176	0.175	0.214	0.178	0.199
First-stage F statistic	52.863	78.149	94.335	79.729	79.602	72.283	80.189	71.511	70.033

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week. The MCVL reports NACE industry classifications based on the Statistical Classification of Economic Activities in the European Community. We form nine sectors based on the NACE level 1 codes as follows: *Agriculture* refers to A: agriculture, forestry and fishing. *Manufacturing* refers to B: mining and quarrying, C: manufacturing, D: electricity, gas, steam and air conditioning supply, and E: water supply; sewerage, waste management and remediation activities. *Construction* refers to F: construction. *Trade* refers to G: wholesale and retail trade; repair of motor vehicles and motorcycles, H: transportation and storage, and I: accommodation and food service activities. *ICT* refers to J: information and communication. *Finance* refers to K: financial and insurance activities and L: real estate activities. *Professional services* refers to M: professional, scientific and technical activities and N: administrative and support service activities. *Public* refers to O: public administration and defence; compulsory social security, P: education, and Q: human health and social work activities. *Other services* refers to R: arts, entertainment and recreation, S: other service activities, T: activities of households as employers, and U: activities of extraterritorial organisations and bodies.

Table 2.A.14: Treatment effects by unemployment risk - split sample

	(1)	(2)	(3)	(4)
	Low risk		High risk	
PM_{10} exceedance	0.095	0.278	0.043	0.152
	(0.019)	(0.083)	(0.018)	(0.053)
Estimator	OLS	IV	OLS	IV
Mean outcome	3.126	3.126	2.459	2.459
Observations	50,853,937		49,345,546	
R^2	0.172	0.172	0.154	0.154
First-stage F statistic		79.815		78.924

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

Table 2.A.15: Treatment effects by health status

	(1)	(2)	(3)	(4)	(5)	(6)
	Weekly absence rate					
PM_{10} exceedance	0.074	0.157	0.076	0.163	0.085	0.189
	(0.019)	(0.084)	(0.020)	(0.088)	(0.020)	(0.090)
× absence share	0.579	2.488				
	(0.354)	(0.710)				
× absence max length			0.411	1.816		
			(0.283)	(0.604)		
× absence mean length					0.465	2.866
					(0.478)	(0.749)
Estimator	OLS	IV	OLS	IV	OLS	IV
Mean outcome	2.915	2.915	2.915	2.915	2.915	2.915
Observations	74,706,090					
R^2	0.176	0.176	0.177	0.177	0.176	0.176
First-stage F statistic		38.477		38.498		38.493

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

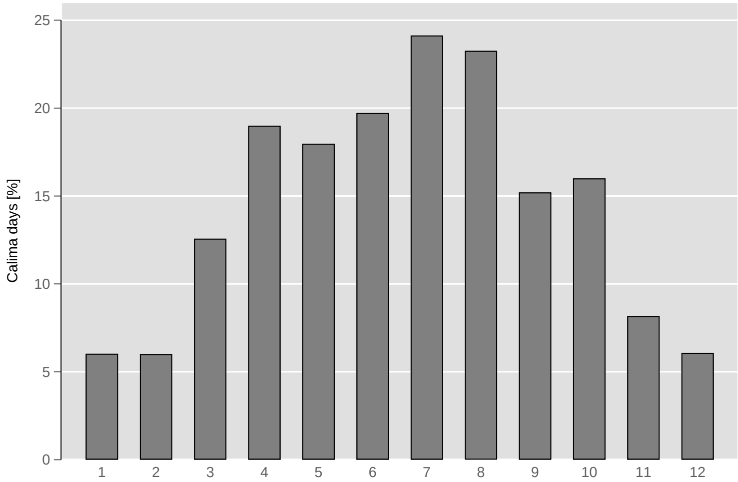
Table 2.A.16: Reduced-form regression for main specifications

	(1)	(2)	(3)	(4)	(5)
	Weekly absence rate				
Calima	0.0412	0.042	0.026	0.028	0.032
	(0.0142)	(0.014)	(0.015)	(0.015)	(0.016)
× unemployment risk		-0.018			
		(0.006)			
× absence share			0.438		
			(0.133)		
× absence max length				0.434	
				(0.144)	
× absence mean length					0.644
					(0.171)
Mean outcome	2.791	2.798	2.915	2.915	2.915
Worker-by-week observations	100,739,754	100,199,483		74,706,090	
R^2	0.165	0.165	0.176	0.177	0.176

Notes: Coefficients scaled by a factor of 100 for better readability. All regressions are estimated by OLS and control for individual fixed effects, age fixed effects, city-year fixed effects, year-quarter fixed effects, flu prevalence and include linear and squared terms of eight weather variables. Robust standard errors in parentheses are clustered by city and by week.

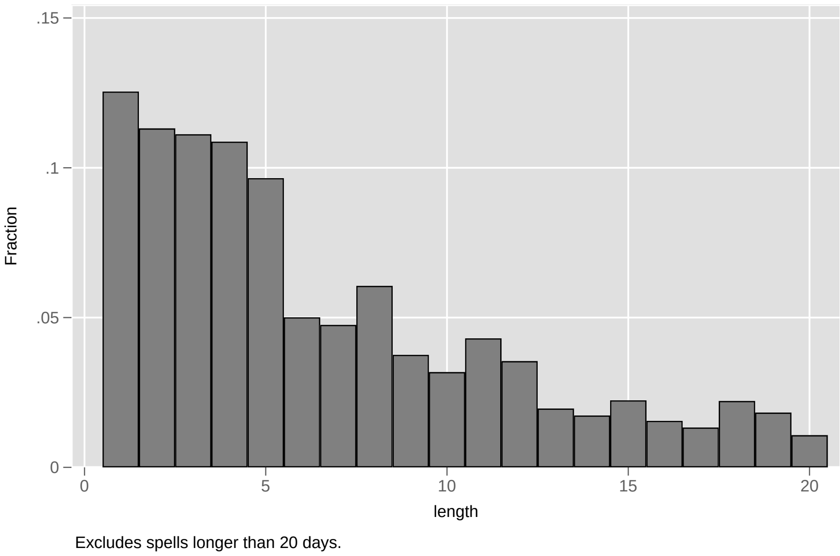
2.B Additional Figures

FIGURE 2.B.1: Duration of sick leaves in days



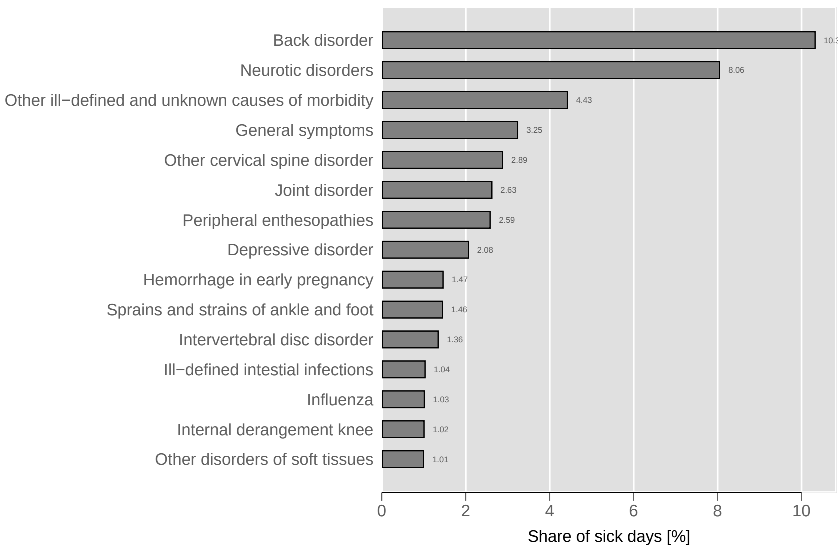
Notes: Each bar represents the average percentage share of days in a given month classified as a Calima episode, based on data from the Spanish Ministry for the Ecological Transition (2018).

Figure 2.B.2: Duration of sick leaves in days



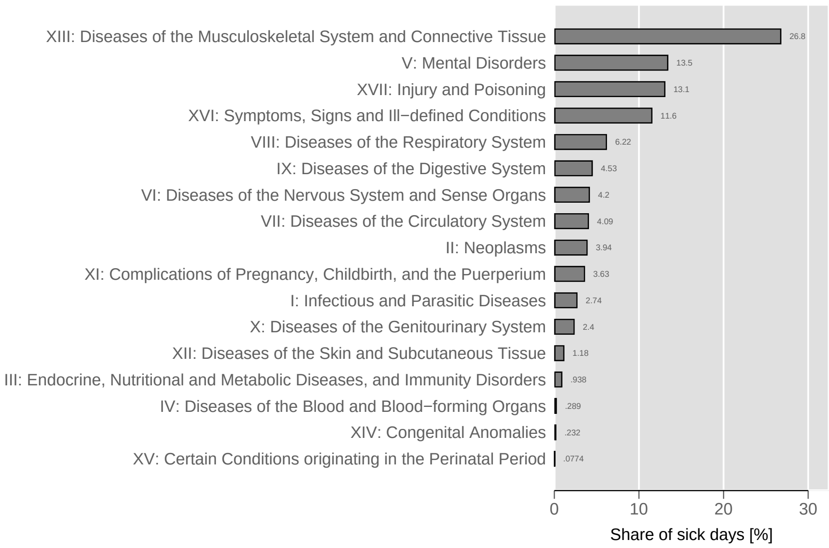
Notes: The figure shows a histogram of the duration of sick leaves in days. Spells longer than 20 days are not depicted. *Source:* Own representation based on data from the Spanish Ministry of Employment, Migration and Social Security (2018).

Figure 2.B.3: Most frequent diagnosis codes



Notes: The figure shows the shares of the 15 most common ICD-9 diagnosis codes for absence spells with a reported diagnosis. *Source:* Own representation based on data from the Spanish Ministry of Employment, Migration and Social Security (2018).

FIGURE 2.B.4: Distribution of ICD-9 diagnosis chapters

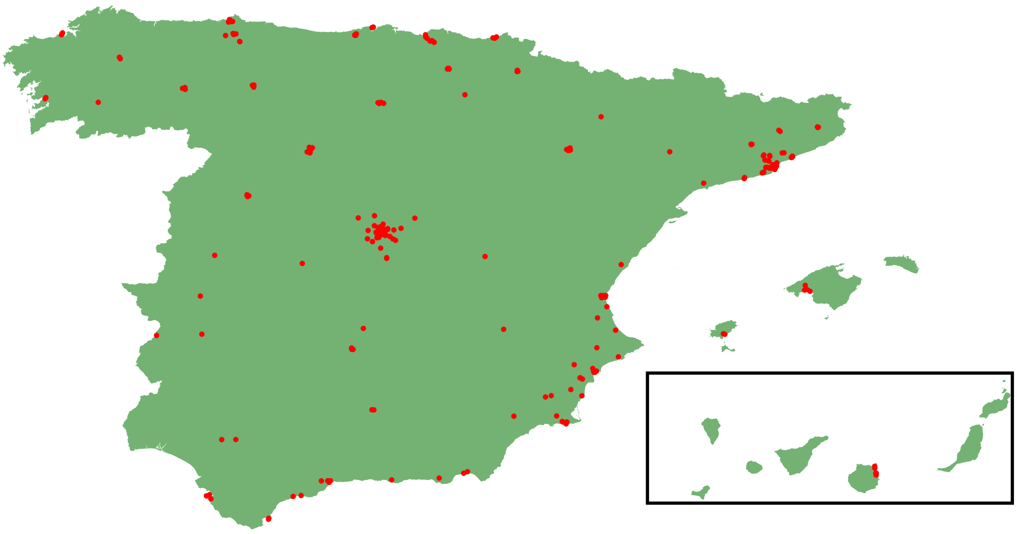


Notes: The figure shows the shares of all ICD-9 diagnosis chapters for absence spells with a reported diagnosis. *Source:* Own representation based on data from the Spanish Ministry of Employment, Migration and Social Security (2018).

FIGURE 2.B.5: Geographic coverage

(A) Cities in the sample

(B) Location of air quality monitors



Source: Own representation based on data from Database of Global Administrative Areas (GADM) https://gadm.org/

2.C Calima monitoring

Non-anthropogenic sources, for example wildfires, can influence air quality negatively. Therefore, when determining whether current PM_{10} levels violate EU limit values, observed air pollution may be adjusted downward to account for such effects. In Spain and Portugal, the most important natural factor augmenting PM_{10} is Saharan sand dust advection, known as "Calima". The magnitude of a Calima event, the PM_{10} discount, may be subtracted from actual PM_{10} readings before determining whether an EU limit value has been exceeded.

A consortium of research institutions (including the Spanish National Research Council's Institute of Environmental Assessment and Water Research, the Spanish State Meteorological Agency, the Spanish Research Center in Energy, Environment and Technology as well as the Nova University Lisbon and the University of Huelva) generates annual reports which identify exceedances of the daily limit value of PM_{10} caused by Calima events and quantify the PM_{10} discounts. The magnitude of Calima events is quantified using 29 rural background air quality monitors which are located in nine different geographic regions covering Spain and Portugal and measure PM_{10} on a daily level.

The researchers of the consortium identify Calima episodes in each of the nine regions and on each day based on sources like the European Center for Medium-Range Weather Forecasts (ECMWF) or the dust forecast of the Barcelona Supercomputing Center (DREAM). If a day and region is deemed to be affected by a Calima event, PM_{10} discounts are determined using measures from the rural background air quality stations.

The discount can be interpreted as the deviation of PM_{10} on a given Calima day from concentrations observed on days unaffected by Calima. The exact PM_{10} discount at the rural background station is calculated as follows. Based on daily measurements of ambient PM_{10}, one determines the 40th percentile of PM_{10} in a moving time window of 30 days, where the current day lies in the center of the window and all days affected by Calima events are dropped[24]. The PM_{10} discount is the difference between the current day's observation and the 40th percentile. In the case that the difference is negative,

[24] The 40th percentile is used because it was found to be the best predictor of local PM_{10} conditional on the day not being affected by sand dust.

the PM_{10} discount is set to zero. A city then can reduce its compliance-relevant PM_{10} reading by the PM_{10} discount reported by the closest rural background station, but only on days which were deemed to be affected by Calima events. In addition, we define a daily Calima indicator which takes value one if on a given day a city's applicable PM_{10} discount is strictly positive, and zero otherwise.

2.D Estimation of unemployment risk

We obtain estimates for individual unemployment risk by relating a worker's observed characteristics to the estimated impact of these characteristics on job loss. This approach is inspired by the labor literature. For example, Card (1996) uses predicted wages as a proxy for skills in order to study the impact of unions on workers with different skill levels. In contrast, we aim to quantify an individual's probability of becoming unemployed.

To do so, we first need to define a measure of job loss. As the MCVL contains only limited information on unemployment status, we proxy for job loss by missing employment entries in the dataset. That is, we define job loss by the absence of employment information in the following month. Other than unemployment, there are two principal reasons for why workers may have a missing employment entry in the dataset. First, workers who retire will by definition have no employment entry. Therefore, the estimation excludes all workers aged 65. Second, some workers exit the MCVL sample at the end of the year. To control for this, we include a dummy for the month of December in the estimation of job loss probability.

We define unemployment risk in two steps. In the first step, we regress observed job loss on worker characteristics. In the second step, we use the estimated coefficients to quantify the probability of job loss for each worker.

For step 1, we estimate a logistic regression model of the probability of job loss of worker i living in city (*municipio*) m at the end of year-month t:

$$\begin{aligned} LOSS_{imt} =& f(x'_{it}\beta^s + \theta^s \cdot \mathbb{I}\{month(t) = December\} \\ &+ \mu^s_{year(t)} + \psi^s_{province(m)} + \epsilon_{imt}) \end{aligned} \tag{2.3}$$

Since we expect that unemployment risk varies widely across sectors, we estimate this equation separately for each one of nine sectors s (agriculture, manufacturing, construction, trade, information and communication, finance, professional services, public, and other services; see Figure 2.A.13 for more detailed definitions). Workers who switch sectors will therefore also switch the estimation subset. The vector x_{it} contains dummies for a temporary contract, being female, and for having Spanish nationality. It further includes quadratic polynomials in the worker's age, tenure in the current firm, and total labor market experience. The coefficient θ captures the effect that workers might leave the MCVL at the end of the year for reasons unrelated to job loss. Furthermore, the equation includes year effects μ and province effects ψ.

We quantify each worker's individual unemployment risk by predicting her probability of job loss based on the sector-specific coefficient estimates in Equation (2.3). The prediction is evaluated using the first week of data available for the worker, and plugging the respective values of x_{it}, year, and province, into the fitted model for the sector s in which the worker is working. We constrain $\hat{\theta}$ to be zero to obtain a measure of unemployment risk also for observations drawn from the month of December. This results in a single measure of unemployment risk for each worker. We do not re-estimate this measure for all the other weeks, but treat unemployment risk as a fixed worker-specific measure. Hence, even if a worker changes sector, she will keep her initial unemployment risk. Before using this variable for further analysis, we drop all observations that were used to predict worker-level unemployment risk from the sample.

Chapter 3

The Benefits and Costs of Driving Restriction Policies: The Impact of Madrid Central on Congestion, Pollution and Consumer Spending

Joint with Jose-Enrique Galdon-Sanchez, Ricard Gil, and Guillermo Uriz-Uharte

3.1 Introduction

By now, there is wide consensus both within and outside economics that air pollution is harmful to people's health. While part of the literature establishes the causal link between air pollution and health outcomes (e.g. Chay and Greenstone, 2003b), much ongoing research also studies the consequences of air pollution beyond health outcomes. In fact, bad air quality has been associated with less cognitive development (Bharadwaj et al., 2017), lower educational and schooling attainment (Ebenstein et al., 2016), crime (Carrillo et al., 2018), or lower productivity (Chang et al., 2016), among others.[1] Given the evidence, it is not surprising that high air pollution levels across the globe have driven the implementation of a wide array of policies and regulations at different levels of government.

This chapter contributes to the policy debate on the benefits and costs of such environmental policies. A typical policy to improve local air quality are traffic restrictions, which are an example of drastic command-and-control regulations with unevenly distributed costs and benefits. While traffic restriction regulations have been found to be effective and reduce outpatient visits (Simeonova et al., 2019), ambulance calls (Zhong et al., 2017), hospitalizations and mortality (He et al., 2019a), and pharmaceutical expenditures (Rohlf et al., 2020), we know little on the indirect effect of these policies on economic activity. Indeed, a reduction in economic activity may change the perception of these pollution-reducing policies by the public. On the one hand, measuring the costs on economic activity allows regulators and policy makers to determine the net gain of implementing these policies. On the other hand, the implementation of these policies may affect different stakeholders differently by spatially redistributing economic activity and potentially generating a division between winners and losers. In other words, fixing a local pollution hot spot might require measures that impose drastic costs borne by few individuals but generate benefits for many others. This chapter contributes to the discussion of the costs of environmental regulation by evaluating the impact of a driving ban implemented in downtown Madrid, known as Madrid Central, on traffic congestion, air pollution, and economic activity.

Madrid Central ("MC" hereafter) is a Low Emission Zone in the city center of Madrid aiming to reduce air pollution through a decline in traffic congestion, and to raise air quality to European Union standards. To achieve this goal, the regulation restricts entry of cars in the center of the city of Madrid (a zone that we will refer to as "MC area") to people living elsewhere. This policy raises a stark tradeoff. Lower emissions in the city center will be a direct benefit of these regulations. However, by restricting access by car, transportation costs are likely to increase for those consumers living outside the MC area, potentially discouraging consumption and retail sales in

[1] The list of outcomes potentially affected by air pollution goes beyond those listed here and reaches out to infant mortality and other health outcomes in the developing world (Currie and Neidell, 2005; Deryugina et al., 2019; Greenstone and Hanna, 2014; Greenstone and Jack, 2015; Hammitt and Zhou, 2006; Neidell, 2004).

businesses within the MC area. Our work empirically examines and documents this tradeoff between cleaner air and lower retail sales in two distinct sections.

First, using data from the European Environmental Agency and the city of Madrid on air quality and vehicle traffic, we assess the direct effect of the regulation on traffic congestion and air pollution in downtown Madrid relative to other areas within the city and its greater metropolitan area. We use difference-in-difference specifications to estimate the effect of MC on congestion and pollution where the MC area zip codes are treated and the period post-MC is the treatment period. Our findings show significant decreases in traffic volume and air pollution in the MC area zip codes relative to other areas in Madrid. In particular, we find that the number of cars per hour in the MC area decreased by 14.7% and the concentration of nitrogen dioxide (NO_2), a harmful pollutant, decreased by 13.6% in the MC area. Moreover, we do not find any evidence of spillovers to areas adjacent to MC in terms of NO_2 but a reduction in traffic in close-by areas.

Second, we use data on credit card transactions to evaluate changes in retail spending within the MC area before and after the implementation of MC. These data on consumer spending span from the first week of 2015 to the tenth week of 2019, while MC was introduced in the first week of December 2018. The data set is unique in that it details the date of each transaction, the zip code of residence of the credit card owner (buyer's zip code) and the zip code of the selling establishment (seller's zip code). We aggregate this information weekly for each buyer zip code–seller zip code pair. As a result, we can effectively measure "trade flows" between all zip codes within the metropolitan area of Madrid before and after the introduction of MC.

We use a triple differences identification strategy in a gravity model, exploiting the fact that MC only has a direct impact on those buyers who live outside the MC area and make all or part of their purchases in the center of Madrid. Following this strategy, we are able to estimate the impact of MC on consumers traveling to downtown Madrid to do their shopping, (1) relative to the shopping of these same consumers in other areas of the city not directly affected by MC, and (2) relative to the shopping in the MC area of consumers living within the MC area, as they are effectively exempt from the MC regulation. The exceptional granularity of our data allows us to mitigate threats to identification by estimating a very demanding specification that controls for time-varying supply and demand shocks in specific areas of the city.

We find an 8.9% decrease in the value of brick-and-mortar spending and a 12.1% increase in the value of online spending of buyers residing in zip codes outside the MC area in establishments within the MC area. Moreover, it appears that the increase in online spending is, statistically speaking, offsetting the decrease in brick-and-mortar spending. This finding opens the possibility of a policy debate linking environmental and e-commerce regulation that favors e-commerce adoption by consumers, retail establishments and small and medium-sized enterprises.

Additionally, our triple differences strategy accounts for the possibility that MC may not only have affected transportation costs for a group of consumers but also increased the attractiveness of Madrid's city center for all shoppers. Our specification allows for seller-specific shocks and compares the behavior of consumers living inside and outside the MC area. This precludes the identification of the potential increase of the attractiveness of the MC area for all consumers. We can examine this pathway in a simple difference-in-difference specification, comparing sales inside and outside the regulated area. We find no impact of MC on sales when using data aggregated at the seller-zip code level. Together with the previous results, this suggests the attractiveness of the MC area does not decrease.

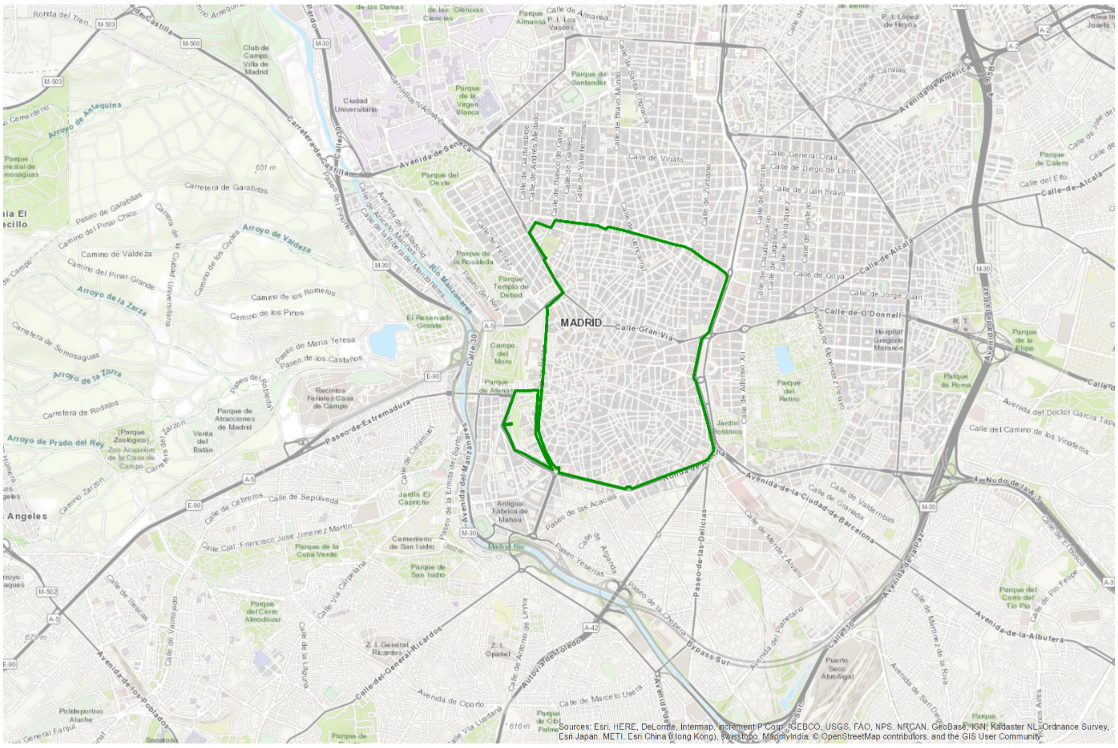
Although a large number of papers investigate the health and air quality benefits of different versions of driving bans and low emission zones, only a few papers study the effects on economic outcomes such as labor supply decisions and local commerce. Most recently, Blackman et al. (2018) and Blackman et al. (2020) use the contingent valuation method based on surveys in Mexico City and Beijing to estimate the costs faced by drivers due to driving restriction programs. Viard and Fu (2015) is the paper closest to ours. The authors show that traffic restriction policies in Beijing reduced the number of hours of labor supplied by affected workers. Besides these works, research on the impact of driving bans on economic outcomes is almost nonexistent. Our work differs from those in a number of ways. First, we present a well-founded and comprehensive empirical analysis of the impact of a traffic restriction on economic activity. Second, our credit card transaction data allow us to measure economic activity in a robust manner as trade flows between zip codes within Madrid. Third, identification relies on a well-defined triple difference strategy where we utilize geographical variation in the application of the policy within the city of Madrid. Fourth and last, our data allow us to separate brick-and-mortar from online transactions. Therefore, we are also able to demonstrate that the diffusion and adoption of e-commerce may dilute part of the potentially negative impact of pollution-reducing policies on retail sales.

Our findings also contribute to a sound literature on the causes and consequences of air pollution, as well as that on the optimization and evaluation of pollution-reduction policies (see reviews and papers by Parry et al., 2007; Graff Zivin and Neidell, 2013; Currie and Walker, 2011, 2019). We examine a particular type of policy aiming to reduce traffic congestion and air pollution by limiting the number of cars allowed to circulate in a heavily congested part of a city. Madrid is not the first city to implement a program of this nature, and consequently ours is not the first study evaluating the efficiency and efficacy of such environmental policies. Examples of cities that have implemented similar traffic-related policies and their respective studies are Mexico City (Eskeland and Feyzioglu, 1997; Davis, 2008; Salas, 2010; Gallego et al., 2013; Oliva, 2015), Quito (Carrillo et al., 2016), Santiago de Chile (Gallego et al., 2013; Rivera, 2017; Barahona et al., 2020), San José – Costa Rica (Osakwe, 2010), London (Leape, 2006; Quddus et al., 2007), Bogotá (Zhang et al., 2017), Stockholm (Simeonova et al., 2019), Taipei (Chen and Whalley, 2012), Beijing (Chen et al., 2013; Viard and Fu, 2015; Zhong et al., 2017), as well as a number of other Chinese (Lin et al., 2011; Li, 2016; Ye, 2017; Li et al., 2019) and German cities (Wolff, 2014; Gehrsitz, 2017).

We view our findings as novel within the existing literature, and important for policy evaluation and future policy design. On the one hand, our results confirm that pollution-reducing policies aiming at traffic control can be effective. On the other hand, our analysis considers the role of e-commerce attenuating potential backfire of some of these policies on economic activity. An implication of our results is that combining environmental friendly policies with regulation that helps retail and small and medium-sized enterprises transition from brick-and-mortar to e-commerce could be socially beneficial. Our results are also informative about the role that e-commerce may play in shaping consumption spending and competition patterns in modern cities.

The structure of this chapter is as follows. Section 3.2 describes in detail the regulation. In Section 3.3, we describe the data. Section 3.4 evaluates whether the regulation was effective in reducing traffic congestion and pollution in the MC area. Section 3.5 examines changes in consumption spending patterns due to the introduction of MC. Section 3.6 concludes.

FIGURE 3.2.1: Map of Madrid Central within the City of Madrid



Notes: The green line marks the part of Madrid's city center subject to the regulations of Madrid Central.

3.2 Madrid Central

The city council of Madrid, Spain, enacted a city-specific traffic regulation, known as Madrid Central, on November 30, 2018. This regulation restricted access by car to an area of 472 hectares located in the Madrid city center.[2] Figure 3.2.1 shows the extension of the affected area, which is the historic center of Madrid as well as the main commercial and leisure district of the city.[3]

When MC went into effect, local authorities noticeably restricted entry by car to the affected area, so access may only be granted under exceptional circumstances. These exceptions are based on the emission category of vehicles. All vehicles are classified in

[2] The city of Madrid has a total surface of 60,400 hectares.

[3] See Boletin Oficial Ayuntamiento de Madrid (2018) or https://www.madrid.es/UnidadesDescentralizadas/UDCMovilidadTransportes/AreaCentral/01InfGral/Ac%20Jta%20Gob%2029%20oct%202018_MC.pdf for details

five different categories according to their emission level (A, B, C, ECO and ZERO in descending order of emissions).[4] Accordingly, the city elaborated a list of exceptions that we list as follows:

(i) Residents of the MC area can enter the MC area without restrictions. If they were to buy a new car, it would need to belong to category B or cleaner to enter without restrictions. All cars of category B or cleaner can enter if they park in a public or private garage.[5]

(ii) Access of delivery vehicles is subject to time restrictions.

(iii) Commercial and industrial vehicles with a parking permit for residential areas of the MC area are allowed to access the MC area. New permits are only handed out for vehicles of category B or cleaner.

(iv) People with reduced mobility are not subject to restrictions.

(v) ZERO emission cars are not subject to restrictions.

(vi) ECO emission cars can enter to park for a maximum of two hours.

(vii) Taxis and ride-hailing vehicles can enter if they are of category B or cleaner.

(viii) Public transport vehicles are not subject to restrictions.

As a result, the population most affected by these regulations is the non-residents of the MC area. That segment of the population cannot access the MC area at all with their own vehicles if they belong to category A, and can only access to park in a garage if they belong to category B or cleaner. This implies, for instance, that non-residents are not allowed to park in the street or access the MC area to pick up or drop off passengers if their vehicles are not classified as ECO or ZERO.

The first day of implementation of the MC regulations was November 30, 2018. During its first month, large traffic signals indicated the perimeter of the MC area and the prohibition of entry. Moreover, local police monitored traffic and informed those drivers in violation of the new regulation without imposing any fines. In January 2019,

[4] Category ZERO refers to electric and hybrid vehicles with a range of more than 40 km. Category ECO refers to hybrid vehicles with a range of less than 40 km and gas vehicles. Category C refers to gasoline vehicles registered after 2006 (EURO 4, 5 and 6) and diesel vehicles registered after 2014 (EURO 6). Category B refers to gasoline vehicles registered after 2000 (EURO 3) and diesel vehicles registered after 2006 (EURO 4 and 5). Category A comprises all other vehicles.

[5] Owners and tenants of private garages need a permit. The plates of vehicles accessing public garages are automatically registered.

the local authorities introduced an automatic monitoring system based on cameras installed at all access points of the MC area. The system registered license plates and informed violating drivers by postal mail of the infraction, without imposing any fines. From March 16, 2019, violations were fined 90€. Our data and analysis cover this initial period up to March 16, 2019.[6] Once we have described in this section the regulation and timing of MC, we proceed with our estimation of the impact of MC on traffic congestion, air pollution, and retail sales.

3.3 Data

To perform our analysis, we combine two different sources of data. First, we use data on traffic, local air pollution and meteorological conditions. Second, we gain access to proprietary data on credit card spending at the transaction level that we can aggregate up to pairs of zip codes for buyer-seller locations within the metropolitan area of Madrid. While the former data allow us to quantify the direct benefits from the driving ban on traffic congestion and air pollution, the latter data will help us quantify indirect costs of the driving ban on consumer spending.

We collected these data not only for the MC area but for the whole metropolitan area of Madrid. Since there is no legal definition for the metropolitan area of Madrid, we define the metropolitan area of Madrid as the area that includes: (1) all zip codes within the city of Madrid, and (2) all zip codes at least partially inside a buffer of 5 km around the perimeter of Madrid.[7] We divide the city of Madrid into the MC area and the rest of the city. Overall, the full metropolitan area comprises 126 zip codes (56 within the city of Madrid and 70 outside). As the credit card data span from the first week of 2015 to the tenth week of 2019, we obtain all other data for the same period.

[6] This chapter investigates whether the policy had a dissuasive effect during its initial phase and finds a clear and robust reduction in car traffic as a result of the introduction of MC during our period of analysis. While also interesting, our work does not examine whether compliance increased after March 16, 2019, when fines became enforceable. Future research should pursue and answer that particular question.

[7] We check the robustness of our findings to alternative definitions.

3.3.1 Traffic and pollution data

We obtain traffic data from the Madrid Department of Traffic Technology published through the city's open data portal.[8] The majority of data comes mostly from traffic lights, but also from other types of sensors. The raw data are reported in 15-minute intervals. First, we drop erroneous observations and outliers in the 99.9th percentile. Then, we aggregate each monitor's readings to the daily level if traffic is observed at least 80 times during a given day. Finally, we aggregate all daily monitor data to the weekly level, conditional on observing every day of the week. The resulting dataset is an unbalanced panel of 4,152 traffic monitors across the city of Madrid. Traffic outside Madrid city is unobserved.

Traffic is measured by the number of vehicles per hour, the share of time (in %) a certain road section is occupied by a vehicle, and the share of road capacity utilized (in %). Summary statistics in Table 3.3.1 show that traffic is denser in the city center of Madrid where, on average, 28% of the road capacity is used during the week, compared to 20% outside of the city center. Because highway M-30 is a major ring road that helps intercity traffic bypass the center of Madrid as well as connect commuting traffic to reach the city center, a significant number of traffic monitors are purposely located on this major road, which explains the high number of vehicles observed at monitors outside the city center.

Because EU regulation defines limit values on NO_2 and other pollutants,[9] cities are obliged to install air quality monitoring stations. The European Environmental Agency (EEA) collects measures from all member countries and makes them publicly available. There are 33 stations reporting NO_2 levels across the metropolitan area.[10] Importantly, one of these 33 stations is located inside the MC area. We use information from this station to estimate treatment effects, considering the rest of the stations as the control group.

The limit value for the mean annual NO_2 concentration specified by the EU regulation is $40\,\mu g/m^3$. As any reading of a station whose daily average is higher than

[8]Retrieved from https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=33cb30c367e78410VgnVCM1000000b205a0aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default

[9]Directive 2008/50/EU. See https://ec.europa.eu/environment/air/quality/standards.htm

[10]Appendix Figure 3.B.1 shows a map with locations of all pollution monitoring stations in Madrid, represented by pink circles. Blue crosses in the map indicate the location of weather stations.

40 µg/m^3 contributes to the potential violation of this regulation, we create an indicator that takes value one if a station's daily average NO_2 reading exceeds 40 µg/m^3. We aggregate all daily NO_2 readings at the weekly level.[11] Table 3.3.1 summarizes weekly and annual mean NO_2 levels and the percentage share of days with NO_2 exceeding 40 µg/m^3. One can see that both, at the station inside the MC area and at the stations outside that area, NO_2 levels are very high according to EU standards. The daily average concentration inside the MC area is 47 µg/m^3, while it is 38 µg/m^3 outside the MC area. We also calculate the share of station-by-year observations that violate the limit value imposed by EU regulation. Table 3.3.1 shows that, during the sample period, the station inside the MC area exceeds the limit value every year. Moreover, other stations outside the city center also violate the threshold. This happens in 47% of all observations.

It is worth noting that meteorological conditions can heavily affect air quality. For example, sunlight is a key component in the decomposition of NO_2. It is therefore important to control for local weather conditions when studying determinants of air quality (Auffhammer et al., 2013). For this reason, we use data from the European Climate Assessment Dataset (ECAD), which provides daily measures of several meteorological variables across Europe. We match the pollution measurement data collected by each pollution monitoring station in the city to its closest available weather measurements from the ECAD dataset (represented with a blue cross in Appendix Figure 3.B.1). We consider data on daily mean temperature, precipitation, cloud cover, humidity, pressure, wind speed, and wind direction. All these weather variables could influence the complex chemistry of air quality and are commonly used in the literature on air quality. Again, we aggregate all readings to the week-level. To account for the effect of weather on driving, we repeat this matching procedure for linking weather data to traffic monitors. Table 3.3.2 shows summary statistics on key meteorological variables. Due to the matching algorithm of weather conditions to air quality observations, the unit of observation in Table 3.3.2 is the pollution monitor station level.[12] In our data, temperature is measured in degrees Celsius, precipitation in tenths of millimeters, cloud

[11]The aggregation of pollution and meteorological control variables could introduce measurement error. However, results on air pollution using daily observations including day-of-week fixed effects are consistent with the weekly estimations. Daily results are available from the authors upon request.

[12]Descriptive statistics of weather data at the traffic monitor level are reported in Appendix Table 3.A.1.

TABLE 3.3.1: Descriptive statistics on traffic and pollution levels

	Mean (1)	SD (2)	Min (3)	Max (4)	Obs (5)
Inside Madrid Central area					
Traffic (101 stations)					
Vehicles per hour	334.76	291.8	0	1,715.28	15,548
Time occupied [%]	10.65	9.73	0	98.51	15,544
Utilized capacity [%]	27.66	10.8	3.51	61.16	15,548
Pollution (1 station)					
NO_2 [$\mu g/m^3$]	47.43	12.13	26.96	95.69	216
$NO_2 > 40\,\mu g/m^3$	0.65	0.3	0	1	216
Yearly NO_2 [$\mu g/m^3$]	47.84	2.46	44.39	49.99	5
Yearly $NO_2 > 40\,\mu g/m^3$	1	0	1	1	5
Outside Madrid Central area					
Traffic (4051 stations)					
Vehicles per hour	454	509.41	0	4,354.98	604,808
Time occupied [%]	6.51	7.28	0	98.33	604,557
Utilized capacity [%]	19.92	10.84	0	99.56	603,920
Pollution (32 stations)					
NO_2 [$\mu g/m^3$]	38	16.96	3.82	133.44	6,971
$NO_2 > 40\,\mu g/m^3$	0.4	0.35	0	1	6,971
Yearly NO_2 [$\mu g/m^3$]	40.05	10.95	14.93	76.14	160
Yearly $NO_2 > 40\,\mu g/m^3$	0.47	0.5	0	1	160

Notes: The table shows descriptive statistics based on weekly station-level data.

cover in okta,[13] daily sunshine in hours, pressure in hectopascal, humidity in percentage terms, wind speed in tenths of meters per second and wind direction is indicated by eight equally sized bins.

3.3.2 Consumption spending data

The final database contains data at the credit card transaction-level from a large European bank.[14] The original data set is unique in that it details the date of each trans-

[13] 0 okta indicates no clouds and 8 okta full cloud cover.

[14] For simplicity, we refer to credit card transactions, but these include both credit and debit card transactions. The raw data includes all credit card transactions of consumers living within the metropolitan area of Madrid that are made, either online or offline, in establishments within the metropolitan area of Madrid with a credit card of the bank providing the data. Approximately, the data covers 15% of all transactions in the area, and can be considered as a representative sample of the credit card purchasing behavior in the overall population of the area. Galdon-Sanchez et al. (2020) provide a detailed description of the database.

Table 3.3.2: Descriptive statistics on weather conditions

	Mean (1)	SD (2)	Min (3)	Max (4)	Obs (5)
Temperature [°C]	15.68	7.69	1.23	30.87	7,187
Precipitation [0.1mm]	10.69	18.21	0	158.29	7,187
Cloud cover [okta]	3.42	1.77	0	7.71	7,187
Sunshine [h]	8.19	2.97	0.97	13.91	7,187
Pressure [hPa]	1,017.38	6.2	994.9	1,035.49	7,187
Humidity [%]	57.96	15.62	22.29	92.86	7,187
Wind speed [0.1 m/s]	22.56	10.36	1.71	80.14	7,187
0° ≤ Wind direction < 45°	0.22	0.22	0	1	7,187
45° ≤ Wind direction < 90°	0.14	0.18	0	0.86	7,187
90° ≤ Wind direction < 135°	0.09	0.13	0	0.86	7,187
135° ≤ Wind direction < 180°	0.05	0.1	0	0.57	7,187
180° ≤ Wind direction < 225°	0.12	0.16	0	1	7,187
225° ≤ Wind direction < 270°	0.2	0.2	0	1	7,187
270° ≤ Wind direction < 315°	0.11	0.15	0	1	7,187
315° ≤ Wind direction < 360°	0.07	0.12	0	1	7,187

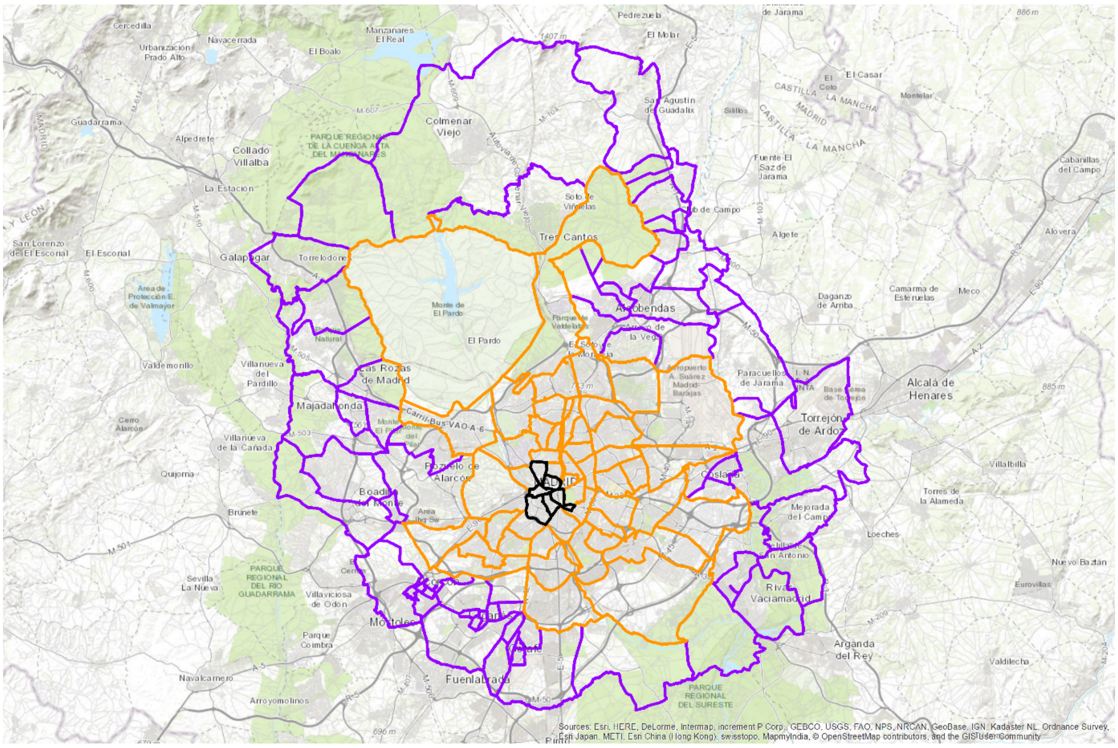
Notes: The table shows descriptive statistics on weather conditions at each pollution monitoring station, where weekly weather is obtained from the closest weather monitor.

action, the zip code of residence of the credit card owner (buyer-zip code) and the zip code of the selling establishment (seller-zip code).[15] Due to our confidentiality agreement with the bank providing the data, we aggregate transaction information at the weekly buyer-seller zip code level from the first week of 2015 to the tenth week of 2019. Figure 3.3.1 shows all 126 zip codes in Madrid. Six zip codes belong to the MC area, 50 zip codes to the rest of the city of Madrid, and 70 zip codes are outside the city of Madrid but inside the metropolitan area of Madrid. Those zip codes (even partially) inside the MC area appear in black, zip codes outside the MC area and inside the city appear in orange, and purple zip codes are those outside the city of Madrid but inside the greater metropolitan area.

Table 3.3.3 presents summary statistics for the main variables of our analysis aggregated at the weekly seller-buyer zip codes level. The average value of "trade flows" between zip codes is 2,087€ coming from on average 54 transactions. The average value of a transaction is 29€. A unique feature of our data is that we are able to separate transactions into two types: brick-and-mortar transactions (B&M in the tables) and online transactions. This is an important feature because it allows us to test the

[15] A zip code in our context is comparable to a 5-digit zip code in the US.

FIGURE 3.3.1: Map of zones and postal codes



Notes: The black lines mark zip codes subject to the regulations of Madrid Central. The orange lines mark unregulated zip codes inside the city of Madrid. The purple lines mark unregulated zip codes outside the city of Madrid but in the greater metropolitan area.

transportation cost mechanism given that transportation costs increase for brick-and-mortar transactions and they do not for online transactions. Introducing this additional level of heterogeneity enriches the substitution patterns between zip codes within and outside the MC area. On the one hand, when consumers' demand for brick-and-mortar transactions is elastic, higher transportation costs will prompt consumers residing outside the MC area to substitute their former purchases in the MC area for purchases in other areas. On the other hand, those consumers with inelastic demand for products from a specific treated zip code may substitute to online transactions. This second scenario is more likely when the retailer sells a differentiated good and, therefore, it is costly to find a suitable brick-and-mortar transaction substitute outside the MC area. The majority of flows are from 51 brick-and-mortar transactions with a total value of 1,955€, on average. The total value of online transactions is on average 132€, coming

Table 3.3.3: Descriptive statistics on sales

	Mean (1)	SD (2)	Min (3)	Max (4)
Total value	2,087.29	9,644.52	0	606,469
B&M value	1,954.69	9,488.51	0	606,469
Online value	132.59	633.51	0	126,363
Total transactions	53.87	268.20	0	16,696
B&M transactions	51.13	264.92	0	16,605
Online transactions	2.73	14.08	0	701
Total transaction value	28.73	39.44	0.09	9,227.47
B&M transaction value	27.59	38.74	0.09	9,227.47
Online transaction value	16.18	72.46	0.12	31,539.1
Share of obs with 0 total value	0.22			
Share of obs with 0 B&M value	0.23			
Share of obs with 0 online value	0.71			

Notes: The table shows descriptive statistics on sales at the weekly seller-buyer level.

from three online transactions. Additionally, 23% and 71% of weekly seller-buyer zip codes observations are zero for brick-and-mortar and online transactions, respectively.

Table 3.3.4 presents basic summary statistics related to selling and purchasing patterns across zip codes in Madrid. The top half of Table 3.3.4 details summary statistics at the seller-zip code level. We can see how the share of revenue coming from online sales changes across zip codes in different areas. While zip codes in the MC area produce 85.6% of their revenue from brick-and-mortar sales, the percentage increases for zip codes in the rest of the city of Madrid and outside the city (90% and 95%). Moreover, the mean value of brick-and-mortar and online transactions also changes across zip codes. Finally, the last two rows in the top half of the table show the share of sales that establishments in the MC area are selling to different areas of Madrid. Not surprisingly, we see that brick-and-mortar sales are tilted towards consumers in the local zip code. Zip codes in the MC area sell, on average, 5.62% of their sales to each of the zip codes in the MC area but only 0.21% to each of the zip codes outside the city of Madrid. Geographical proximity also matters for online sales (as documented by Blum and Goldfarb, 2006). On average, 1.84% of the online sales from zip codes in the MC area go to each of the six zip codes in this area, 1.34% go to each of the 50 zip codes in Madrid city and only 0.4% to each of the 70 zip codes outside of the city of Madrid.

The bottom half of Table 3.3.4 reports statistics on consumer behavior by buyer-seller-zip code dyad. Consumers living in the MC area carry 45.2% of their brick-

TABLE 3.3.4: Descriptive statistics on consumption

	MC Area (1)	Madrid City (2)	Outside Madrid City (3)
Number of zip codes	6	50	70
Seller-zip code statistics			
Share of revenue coming from B&M sales	85.6%	90%	95%
Mean value of B&M sales	38.28	36.75	41.1
Mean value of online sales	63.2	44.58	54.79
Mean share of B&M sales by zip codes in Madrid Central to each of the zip codes in MC, Madrid City, or outside Madrid City	5.62%	1.12%	0.21%
Mean share of online sales by zip codes in Madrid Central to each of the zip codes in MC, Madrid City, or outside Madrid City	1.84%	1.34%	0.40%
Buyer-zip code statistics			
Share of B&M purchases in MC	45.2%	8.9%	4.3%
Share of online purchases in MC	20.3%	18.70%	14.80%
Share of B&M purchases in local zip code	27%	28.50%	38.70%
Share of online purchases in local zip code	13.10%	15.10%	14.3%

Notes: The table shows descriptive statistics on selling and purchasing patterns at the weekly level.

and-mortar purchases and 20.3% of their online purchases in establishments inside their area. These shares decrease monotonically with the distance to MC. Consumers in other zip codes of the city of Madrid make, on average, 8.9% of their brick-and-mortar purchases and 18.7% of their online purchases in establishments within the MC area. For consumers living outside the city, these numbers decrease to 4.3% of brick-and-mortar purchases and 14.8% of online purchases. The last two rows show how much consumers living in different areas of Madrid spend within their local zip code. As the share of brick-and-mortar sales that consumers make in their local zip code is concerned, we see that consumers outside the city tend to spend more (38.7% of their total brick-and-mortar expenditures) than consumers elsewhere. By contrast, we do not see large differences across areas in the propensity to buy online in the local zip code (13.1% for consumers in MC, 15.1% for consumers in the city, and 14.3% for consumers outside the city).

3.4 The effect of Madrid Central on traffic and air quality

The main goal of the regulation of MC is to reduce traffic in the city center of Madrid and thereby lower air pollution. In this section, we study whether the policy achieved that goal. MC focuses on the reduction of NO_2, a pollutant mainly emitted by vehicles, as the city of Madrid repeatedly violated NO_2 limit values defined by European Union environmental regulation. After defining our empirical strategy, we show our results of the impact of MC on traffic and air pollution.

3.4.1 Research design

We estimate the effect of MC on traffic or NO_2 levels using the following regression equation.

$$Y_{swy} = \beta MC_{swy} + \delta X'_{swy} + \mu_{sw} + \tau_{wy} + \epsilon_{swy} \tag{3.1}$$

Y_{swy} stands for the traffic or pollution outcome of interest at the traffic or air quality monitor station s, week w, and year y. It is important to note that the traffic and air quality monitors are not identical. The variable MC_{swy} is a dummy that takes value one if station s is inside the MC area in a year-week in which MC is in effect. The vector X'_{swy} includes controls for meteorological conditions at the location of station s, week w, and year y. Therefore, the coefficient δ captures the effect of weather on air pollution levels.[16] For example, these would control for the case that the introduction of MC coincided with the wind blowing from a direction that induces lower pollution levels in the MC area. Moreover, we include station-week fixed effects μ_{sw} to control for season-specific patterns at each monitoring station. This set of fixed effects controls for instance for the case that during the Christmas season many shoppers go to the city center, increasing traffic and pollution levels. The variable τ_{wy} controls non-parametrically for time trends and year-week-specific shocks. This variable controls, for example, for the celebration of specific events attracting many visitors to the city and affecting pollution levels. The error term ϵ_{swy} is potentially serially correlated, so we cluster standard errors at the station level. By using this specification, we aim to consistently estimate the effect of MC on air pollution, captured by β, while controlling for possible confounding factors.

Our estimation strategy requires common trends in treated and untreated stations once we account for all control variables. This could fail, for instance, if people living in the MC area were substituting their old cars for electric vehicles at a faster pace than people in other areas of Madrid were. To account for this, we also allow for station-specific trends. Our estimates could still be compromised if there were other policies introduced at the same time as MC, affecting traffic or pollution levels in specific areas of the city. If, for instance, a metro line covering the city center opens at the same time as the introduction of MC, we could wrongly attribute the metro's positive effect on air quality to MC. We are not aware of any policy change or intervention of this type during the time span of our data set.[17]

[16] This includes second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators.

[17] In January 2019, the City Council of Madrid reduced the speed limit on highway M-30 in order to decrease pollution levels. As this route does not cross the MC area, if anything, we would expect the policy to decrease pollution levels in the control group.

Table 3.4.1: Effects on traffic levels

	Vehicles per hour (1)	Time occupied [%] (2)	Utilized capacity [%] (3)	Log Vehicles per hour (4)	Log Time occupied (5)	Log Utilized capacity (6)
Madrid Central	−48.500	−1.936	−5.679	−0.147	−0.178	−0.228
	(12.130)	(0.850)	(0.776)	(0.025)	(0.050)	(0.038)
Location-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	456.300	6.571	20.190	5.648	1.493	2.905
N×T	597,221	596,895	596.328	597,031	592,874	571.518
N	3948	3948	3948	3948	3927	3823

Notes: Standard errors are clustered at the station level. The variable Madrid Central takes value 1 when a station is located within the MC area and the MC regulations are in place, and 0 otherwise. The weather controls are second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators.

3.4.2 Results

Table 3.4.1 presents the results of estimating Equation (3.1) for the three measures of traffic, in levels and logs, with standard errors clustered at the station level. We find large effects of MC on traffic. The average number of cars dropped by 48.5 (column 1), or 14.7% (column 4). MC reduced the frequency of road segment usage by cars by 1.9 percentage points (column 2), or 17.8% (column 5). A decrease in these two measures implies that roads are used less, in fact, road capacity utilized under MC decreased by 5.7 percentage points (column 3), or 22.8% (column 6). These estimates are not only statistically significant at the 1% level while clustering at the monitoring station level, but also are economically significant magnitudes.

Because those that cannot enter the restricted area may park in areas close by, or not drive to the center at all, MC might generate spatial spillover effects (either positive or negative) in traffic levels to nearby areas. In fact, our initial regression specification may be overestimating the decrease in traffic in the restricted area. To account for the spillover, we include a dummy variable in Equation (3.1) that takes value one if station s is inside a 1.5 km buffer around the MC area in a year-week in which MC is enforced.[18] Appendix Table 3.A.2 shows that the net spillovers are positive, i.e. that

[18]Results are robust to defining alternative buffers around the MC area. These are available upon request.

traffic is also reduced in streets close to the regulated area. As expected, the magnitude of the reduction is smaller than inside the MC area. One can also see that not accounting for positive spillover effects leads to an underestimation of the absolute effect on traffic inside the MC area.

Table 3.4.2 presents the results on air quality. We cluster the standard errors in all specifications at the air quality monitor level. In column 1, we use the log of the average weekly level of NO_2 as the dependent variable. Our findings suggest a decrease of 16% in NO_2 in the restricted area due to the introduction of MC. Defining the three closest stations inside the 1.5 km buffer around the MC area (see Appendix Figure 3.B.1) as its immediately adjacent area, the results in column 2 show that (i) the estimated reduction in pollution levels in the MC area remains unchanged, and (ii) there is no evidence of net spillovers to adjacent areas. This result could be due to the fact that the reduction of traffic in the surroundings is not strong enough or because the composition of cars outside the regulated area is unaffected. We repeat the same exercise in column 3, considering spillovers to any station within the city of Madrid. The effect on the MC area is now slightly smaller in magnitude, but we find no evidence of spillovers neither towards adjacent areas nor to areas in the rest of the city. These estimates cannot be compared to the results on traffic, as traffic outside the city of Madrid is unobserved.

In columns 4 to 6, we show results of running the same specification with a different dependent variable, the share of days of a week in which NO_2 levels exceed 40 $\mu g/m^3$. Our findings here are consistent with those in columns 1 to 3, suggesting a decrease of 12 percentage points in the days of a week in which NO_2 levels exceed 40 $\mu g/m^3$. This represents a 25% reduction relative to the sample mean. These results are confirmed by Pseudo-Maximum Likelihood Poisson regressions in Appendix Table 3.A.3, where the outcome is the number of days in a given week in which the limit value was exceeded.

3.4.3 Robustness checks

Our results appear to remain unchanged both qualitatively and quantitatively when including station-specific trends (Appendix Table 3.A.4). In a separate specification, we also include the air quality monitoring stations located in Barcelona as a control

Table 3.4.2: Effects on NO_2 levels

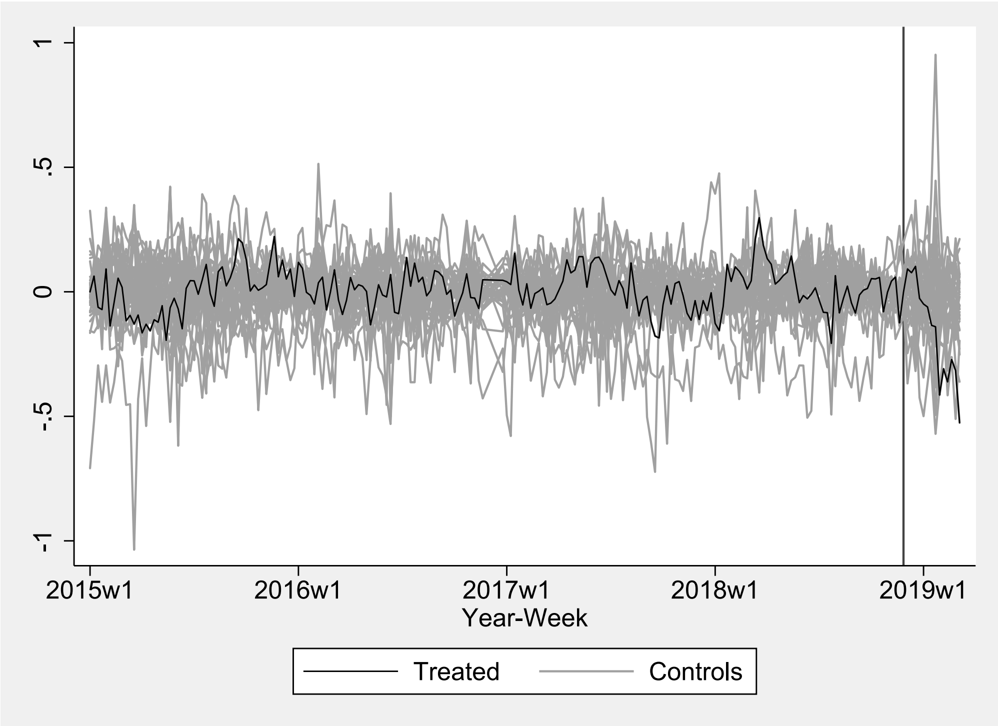
	Log NO_2			$NO_2 > 40$		
	(1)	(2)	(3)	(4)	(5)	(6)
Madrid Central	−0.162	−0.163	−0.136	−0.121	−0.122	−0.111
	(0.017)	(0.018)	(0.038)	(0.015)	(0.016)	(0.031)
Surroundings		−0.008	0.018		−0.015	−0.004
		(0.032)	(0.046)		(0.029)	(0.030)
City of Madrid			0.036			0.015
			(0.042)			(0.034)
Station-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	3.545	3.545	3.545	0.408	0.408	0.408
N×T	7187	7187	7187	7187	7187	7187
N	33	33	33	33	33	33

Notes: Standard errors are clustered at the station level. The variables take value 1 when a station is located in the indicated area and the MC regulations are in place, and 0 otherwise. The weather controls are second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators.

group and find that results remain mostly unchanged, except for a small reduction in the impact on the share of days exceeding 40 μg/m³ (Appendix Table 3.A.5).

Finally, as only a single station is treated and the number of clusters is relatively small (33), we also implement the Synthetic Control Method to estimate the impact of MC on air quality in downtown Madrid (Abadie et al., 2010). The station located inside the MC area is matched to a number of monitors outside the MC area based on pre-treatment data of air quality. Each control monitor receives a certain weight, such that the weighted mean of the control monitors' readings predicts air quality at the treated monitor. The algorithm chooses weights to minimize the mean squared error of these predictions. While one could try to find optimal weights by predicting every single observation of air quality at the treated monitor prior to intervention, we only choose a subset of NO_2 readings to be matched. From the beginning of 2015 to mid-2018, i.e. the 25th week of 2018, we only consider air quality from every 20th week to avoid overfitting. After that, we consider all readings until the 47th week of 2018. Treatment begins in the 49th week of 2018. In addition, we also match on the pre-treatment average of NO_2. We do not make use of weather controls as additional matching variables since, by construction, most stations face almost exactly the same weather conditions. Be-

FIGURE 3.4.1: Synthetic control method for pollution levels



Notes: The figure plots synthetic control estimates. The black line marks the treated station, the gray lines the controls. All lines plot the difference between a actual measurement and a prediction based on controls stations. The vertical line indicates the introduction of Madrid Central.

fore running the algorithm, we deseasonalize each station's data. The matched stations provide good predictions of pre-treatment NO_2 concentrations at the station inside the MC area with an R-squared of 0.88.

Figure 3.4.1 shows the effect on the treated station (in black). It seems that, at the beginning, MC was not yet effective. However, after some weeks, it decreased NO_2 levels by close to 50%. We cannot calculate standard errors, but repeat the analysis with a placebo treatment for each other monitor (in gray). Comparing the results from these stations, we see that the 50% drop can be interpreted as an unusually large deviation. Abadie et al. (2010) suggest that an effect is significant if the estimated effect of the treated unit is unusually large compared to the distribution of placebo estimates. They propose that one should not simply compare mean squared prediction errors of treated and placebo units in the post-treatment period, but scale these errors by the respective mean squared prediction errors in the pre-treatment period. In our case, we find that the ratio of mean squared prediction errors of the treated air quality station is larger than the ratios of all 32 control stations.

3.5 The effect of Madrid Central on consumption spending

The results in Section 3.3 indicate that MC achieved its goal of reducing car traffic and pollution levels in the city center of Madrid. However, this may come at the cost of distorting citizens' habits and market outcomes. One of the most salient and controversial dimensions of these distortions is the possible impact that MC may have had on consumption behavior. An increase in the cost of transportation to the MC area can potentially discourage consumption in that area. In this section, we empirically examine whether MC actually affected consumer behavior, and if so, how. Understanding the costs of pollution-reducing policies is as important as evaluating their benefits. Therefore, the results of this section may help policy makers derive conclusions for the introduction of similar policies in the future.

Our theoretical framework yields predictions of the impact of an increase in transportation costs (actual transportation costs or disutility through inconvenience) for consumers living outside the MC area when they make purchases of goods and services from businesses within MC area. In our context, changes in transportation costs induced by MC should not directly affect: (i) purchases of residents from the MC area in businesses within the MC area; (ii) purchases of residents outside the MC area in businesses outside the MC area, as the regulation only restricts traffic inside the MC area; and (iii) purchases of residents from the MC area in businesses outside the MC area. In other words, we are able to clearly define which "trade flows" are directly affected by the policy and which are unaffected. Therefore, the predictions from our theoretical framework and our empirical analysis allow us to identify the impact of the increase in transportation costs for those affected, whilst controlling for demand shocks and supply shocks at different zip codes.

Following this intuition, we aggregate transactions at the week level for each combination of seller-zip code and buyer-zip code dyad available in the data. The resulting data set contains weekly information on how much consumers of each zip code are buying from sellers of each zip code in Madrid.[19]

3.5.1 Theoretical framework and research design

We build our identification strategy using a theoretical framework based on a standard gravity model and the seminal work of Anderson (1979), Eaton and Kortum (2002) and Baier and Bergstrand (2007). Assume a city with N zip codes, and each zip code has buyers and sellers. For simplicity, we consider buyers indexed by their zip code $i = 1, \ldots, N$ and sellers indexed by their zip code $j = 1, \ldots, N$. The sellers in each zip code sell an item differentiated from all items sold in other zip codes. Buyers may choose to buy items from any zip code, and sellers can sell to buyers from any zip code. While this is effectively a static model, we allow for multiple periods indexed according to their week of the year $w = 1, \ldots, W$, and their year $y = 1, \ldots, Y$.

Consider then a representative consumer model with a CES demand function in which the buyer residing in zip code i, in week w of year y, has to decide how much to buy from each of the seller zip codes j (Q_{ijwy}). There is a seasonal (weekly) taste-specific shock θ_{ijw} at the level of the buyer-seller-week. A seller of zip code j cannot price discriminate across different buyers and therefore sets a price P_{jwy} common to all buyers. Moreover, buyers incur iceberg transportation cost τ_{ijwy}. Because we want to study the impact of the introduction of MC on spending flows between zip codes, we allow transportation costs to vary at the buyer-seller-week-year level. In our case, we hypothesize that the introduction of MC will affect the purchases in zip codes inside the MC area from buyers in zip codes outside of MC area. Therefore, the objective function U_{iwy} is the following.

$$U_{iwy} = \left(\int \theta_{ijw} Q_{ijwy}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} - \int \left(\tau_{ijwy} P_{jwy} \right) Q_{ijwy} dj$$

[19]This data structure is comparable to that found in the international trade literature for the estimation of gravity equations (Head and Mayer, 2014; Atalay et al., 2019). Analogously to the trade literature, our data allows us to study how "trade flows" between different geographical areas change when transportation costs change exogenously.

Each consumer maximizes her consumer surplus with respect to Q_{ijwy} taking preferences, prices and other parameters as given.

Let $\widetilde{P}_{iwy} = \left(\int \theta_{ijw}(\tau_{ijwy}P_{jwy})^{1-\sigma}dj\right)^{\frac{1}{1-\sigma}}$ be the price index of buyer i, in week w of year y. Let also $\widetilde{Q}_{iwy} = \left(\int \theta_{ijw}Q_{ijwy}^{\frac{\sigma-1}{\sigma}}dj\right)^{\frac{\sigma}{\sigma-1}}$ be the total amount consumed by buyer i, in week w of year y. Then, the total value of consumption by buyers residing in zip code i, in establishments of sellers in zip code j, in week w of year y will be equal to

$$P_{jwy}Q_{ijwy} = (\widetilde{P}^{\sigma}_{iwy}\widetilde{Q}_{iwy})(P^{1-\sigma}_{jwy})(\theta^{\sigma-1}_{ijw})(\tau^{-\sigma}_{ijwy})$$

Here we can see how an increase in transportation costs τ_{ijwy}, like the one induced by the introduction of MC, will reduce consumption levels.[20] Moreover, this expression can be mapped one-to-one (using logs) to the following equation that we will actually estimate with our data,

$$Y_{ijwy} = \alpha_{iwy} + \gamma_{jwy} + \delta_{ijw} + \beta Treatment_{ijwy} + u_{ijwy} \tag{3.2}$$

where Y_{ijwy} measures (log) expenditures of residents in zip code i in establishments in zip code j during week w of year y. The variable $Treatment_{ijwy}$ is a dummy variable that takes value one if i is a buyer-zip code outside the MC area, j is a seller-zip code inside the MC area, and we are in a week-year in which the MC regulations are in effect. Note this dummy is aimed to capture increases in transportation cost between a zip code pair triggered by the introduction of MC, and that β is the coefficient of interest as it measures the effect of MC on purchases of buyers from outside the MC area in establishments inside the MC area once the policy is in effect. Additionally,

[20] The increase in transportation costs induced by the introduction of MC will have a direct impact on the level of purchases from buyer-zip codes outside of MC area in establishments inside the MC area. In turn, if the reduction of consumption in MC area has spillovers in consumption levels in other zip codes, these should be controlled for by the fixed effects structure. We will not be able to separate this indirect effect of the introduction of MC from aggregate shocks at the buyer-zip code level. However, note that this impact should be economically small if the number of zip codes is large enough. We have 126 zip codes, which should make our case comparable to the usual International Trade framework modeling trade across countries.

α_{iwy} is the buyer-by-week fixed effect, and γ_{jwy} is the seller-by-week fixed effect.[21] The variable δ_{ijw} is the buyer-by-seller fixed effect specific for each week of the year. We allow this dyad-specific fixed effect to vary by the week of the year to account for seasonality patterns (e.g. during Christmas time people living in the outskirts of the city may disproportionately increase their shopping in the city center). Finally, u_{ijwy} is the error term.

As a result, through Specification (3.2) we aim to identify the effect of MC on spending levels from buyers living in zip codes outside the MC area in establishments inside the MC area, both relative to the shopping of these same consumers in other areas of the city and relative to the shopping in downtown Madrid of consumers living within MC area.

The coefficient of interest, β, identifies the partial equilibrium effect of the increased transportation costs due to Madrid Central. Additionally, these cost changes also have a general equilibrium effect as a result of demand substitution. In the case of CES-demand, this is captured by changes in the price index $\widetilde{P}_{iwy}$ and, hence, they affect consumption spending of buyers located in zip codes outside of the MC area in all seller-zip codes, both inside and outside of the MC area (Larch and Yotov, 2016; Piermartini and Yotov, 2016).[22] In Equation (3.2), these changes in the price index $\widetilde{P}_{iwy}$ are captured by the set of buyer-by-week fixed effects α_{iwy}.[23]

[21]The buyer-zip code-week fixed effect α_{iwy} and the seller-zip code-week fixed effect γ_{jwy} would correspond to the importer-period and exporter-period fixed effects in trade models. The parameter α_{iwy} controls for changes over time in the average level of expenditures of people living in zip code i. The parameter γ_{jwy} controls for changes in the attractiveness of shopping in zip-codes inside the MC area.

[22]A similar mapping exists for a logistic specification of demand to our estimates (e.g. Berry, 1994).

[23]Arguably, in addition to the increase in transportation for consumers living outside of the MC area, the traffic ban also changed the attractiveness of the MC area (e.g. because walking in that area is nicer after the introduction of MC). In our model, that would imply that θ_{ijw} increases for all buyer-zip codes i when buying in zip codes j that are in MC area after the regulation came into effect. Because this is a general effect for all buyer-zipcodes, its impact would be fully captured by the seller-by-week fixed effect γ_{jwy}. Our baseline specification will not allow us to separate this potential change in the attractiveness of seller-zip codes in the MC area from other supply shocks taking place simultaneously in those places that are also captured by the seller-by-week fixed effect. To study whether there are changes in the attractiveness of the MC area, we show diff-in-diff results in Table 3.5.4.

3.5.2 Results

We proceed next with our "gravity-like" methodology. Because the outcome variables in this section are measured in logs and the spending flows between two zip codes in a given week can be zero, we add the value one to the dependent variable of interest throughout this section.[24]

We estimate β from Specification (3.2) and show results of the triple difference estimation in Table 3.5.1. In columns 1 and 2, we use total transaction revenue as the dependent variable and find with a statistically insignificant decrease of 3% in spending in the MC area by affected consumers. Columns 3 to 6 examine the impact of MC on brick-and-mortar and online transactions separately. While MC decreases brick-and-mortar spending between 4.7 and 8.9%, it increases online spending between 9.4 and 12.1%. All four columns show statistically significant findings. Therefore, these results suggest that, upon the increase in transaction costs due to the implementation of MC, consumers in zip codes outside the MC area switched consumption spending from brick-and-mortar to online transactions. Note that we show in Appendix Table 3.A.6 that mean transaction values for all, brick-and-mortar, and online transactions did not statistically change due to the introduction of MC. Additionally, Appendix Table 3.A.7 regresses the share of online revenue and the share of online transactions per buyer-seller zip codes dyad and finds results consistent with those in Table 3.5.1, as well as no statistical change in the relative size of brick-and-mortar to online transaction values.

Estimates in Table 3.5.1 correspond to the partial equilibrium impact of Madrid Central, as explained in Section 3.5.1. This means that the changes in consumption of affected consumers shopping inside the MC area are expressed relative to their total consumption. In a general equilibrium context, this total consumption can also adjust because of substitution effects. For instance, the 9% decrease in brick-and-mortar consumption in the MC area by affected consumers is relative to the total consumption of these consumers, which might change in general equilibrium. If substitution effects are only small, our estimates should also be close to the total effects. If they are large, the total effect would be smaller than 9%. Because only five out of 126 zip codes were directly affected by Madrid Central, we anticipate that general equilibrium responses

[24] Up to 14.1% of the dyad-week flows are zero in our sample. This is substantially lower than in usual setups of country trade flows where there are around 50% of zeros (Helpman et al., 2008).

TABLE 3.5.1: Baseline Results

	Total		B&M		Online	
	Rev (1)	Trans (2)	Rev (3)	Trans (4)	Rev (5)	Trans (6)
Treatment	−0.039	−0.029	−0.090	−0.047	0.121	0.094
	(0.038)	(0.024)	(0.040)	(0.025)	(0.061)	(0.044)
Buyer-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968

Notes: Standard errors are clustered at the buyer-seller pair level. The dependent variable is log of revenue and log of number of transactions for all transactions, brick-and-mortar and online transactions at the seller-zip code by buyer-zip code level in a given week. The variable Treatment takes the value one when (1) a seller-zip code is within the MC area and the MC regulations are in place, and (2) the buyer-zip code is outside the MC area, and 0 otherwise.

do not play a large role. This clarification does not affect the result that aggregate spending in the MC area by affected consumers is unchanged, as we find no effect, but it could matter when we qualify the observed decrease in brick-and-mortar spending and the increase in online spending.

An interesting departure from Specification (3.2) is one where we split the dummy of interest, MC_{ijwy}, into two: one for those buyers living in zip codes within the limits of the city of Madrid (and outside the MC area), and one for those zip codes located outside the city of Madrid and still inside the metropolitan area of Madrid. This alternative specification accounts for potential differences in the set of available transportation options to the center of the city of Madrid. Because the policy restricts driving into the city center, consumers now could consider other means of transportation. Those living in zip codes closer to the MC area are more likely to be able to switch at low costs by walking or taking public transport. By contrast, consumers living further away from the MC area might find it more difficult to substitute their car for other means of transportation, as they are not able to walk to the city center and their access to public transport is less convenient. In Appendix Table 3.A.8 we use estimates on travel times from the Google Maps Distance API to show that the relative time loss when switching from car to public transport is indeed larger for zip codes located farther away from the city center. Therefore, one would expect the impact of MC on consumers living

Table 3.5.2: Heterogeneous effects

	Total		B&M		Online	
	Rev (1)	Trans (2)	Rev (3)	Trans (4)	Rev (5)	Trans (6)
Zip codes city	-0.0376	-0.0299	-0.0804	-0.0399	0.049	0.0763
	(0.0394)	(0.0251)	(0.0418)	(0.0261)	(0.0629)	(0.0454)
Zip codes out of city	-0.0391	-0.0285	-0.0959	-0.0524	0.172	0.107
	(0.041)	(0.0254)	(0.0441)	(0.0265)	(0.064)	(0.0451)
Buyer-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes	Yes	Yes	Yes
p-val equal effects	0.956	0.921	0.581	0.4	0.00155	0.0699
Observations	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968

Notes: Standard errors are clustered at the buyer-seller pair level. The dependent variable is log of revenue and log of number of transactions for the full sample, brick-and-mortar and online transactions at the seller-zip code by buyer-zip code level in a given week. The variable Zip codes city takes value one when (1) the seller-zip code is within the MC area and the MC regulations are in place and (2) the buyer-zip code is outside the MC area but within the city of Madrid, and 0 otherwise. The variable Zip codes out of city is defined in the same way but for buyer-zip codes outside the city of Madrid.

farther away from the city to be more severe than for those living closer to the MC area.

The results of this exercise are reported in Table 3.5.2. Columns 1 and 2 show that MC did not have a statistically significant impact on total revenue or the number of transactions, neither from Madrid city zip codes nor from zip codes outside of Madrid. These findings are consistent with those shown in Table 3.5.1.

Next, we examine spatial heterogeneity of the impact of MC on brick-and-mortar revenues and number of transactions. The results in columns 3 and 4 show that brick-and-mortar revenues and transactions decrease across the board. The magnitude of the decrease in revenues and transactions is larger for buyer-zip codes outside the city of Madrid, although we cannot statistically reject the hypothesis that they are the same. Interestingly, we find opposite findings regarding online transactions. Columns 5 and 6 show that online revenues and transactions increased across the board. The magnitude of the increase is larger for revenues and transactions in buyer-zip codes outside the city of Madrid, where the increase in transportation cost to the MC area is likely to be largest. Moreover, we can reject the null hypothesis that the effects on online transac-

TABLE 3.5.3: Gravity

	Total revenue	B&M revenue	Online revenue	Share online revenue
	(1)	(2)	(3)	(4)
Log distance between zip codes	-1.722	-1.767	-0.608	2.080
	(0.0196)	(0.0196)	(0.0157)	(0.0792)
Buyer-week-year FE	Yes	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes	Yes
Observations	3,460,968	3,460,968	3,460,968	3,460,968

Notes: Dependent variable: Log of revenue for all transactions, B&M and online transactions separately and percentage share of online revenue. Standard errors are clustered at the buyer-seller pair level. The dependent variable is the log of revenue for all transactions, brick-and-mortar and online transactions at the seller-zip code by buyer-zip code level in a given week (cols. 1-3), and the percentage share of online revenue for that seller-buyer-week triple (col. 4). The independent variable is the log of distance measured in km between the centroid of the seller-zip code and the centroid of the buyer-zip code. In all columns, we control for seller-week-year specific FE and buyer-week-year specific FE.

tions are equal for buyer-zip codes inside and outside the city of Madrid.[25] Again, this indicates a substitution between brick-and-mortar and online consumption due to the increase in transportation costs for residents living outside the MC area.

We have made two important observations. First, transportation costs matter as MC affects consumer behavior. The increase in transportation decreases consumer spending in brick-and-mortar establishments. Second, when transportation costs increase, there is a switch from brick-and-mortar to online spending. We examine these general mechanisms in Table 3.5.3 and find that, indeed, consumption exhibits gravity. Spending decreases with distance for both brick-and-mortar and online consumption. But relatively, the impact of distance is greater for brick-and-mortar consumption such that the share of online purchases increases with distance. This is important for two distinct purposes. On the one hand, we show that transportation costs within a city matter. On the other hand, we confirm the need to control for seller-buyer zip codes pair fixed effects to clear problems of endogeneity. The existence of gravity implies that those buyers located furthest away from the MC area were buying little in the MC area to begin with. Consequently, the introduction of MC increased transportation costs of those buyers located further away from the MC area to a greater extent. This

[25]Note that these findings also provide an alternative way of testing the impact of MC on online consumption levels of consumers living outside the city of Madrid by using as a control group online consumption levels of consumers inside the city of Madrid and outside the MC area.

means that we must include buyer-seller zip code fixed effects to avoid negative bias in the estimation of the effect of MC on consumption spending.[26]

As mentioned earlier, MC might not only affect transportation costs for a group of consumers but also increase the attractiveness of Madrid's city center for all shoppers. The triple differences strategy accounts for this confounding factor by including seller-specific fixed effects and by comparing the behavior of consumers living inside and outside the MC area. When including such fixed effects, our results show that consumers living outside the MC area do not decrease their consumption in the MC area relative to those living inside. However, this strategy precludes the quantification of the potential increase of the attractiveness of the MC area for all consumers. We can examine this pathway in a simple difference-in-difference specification, comparing sales inside and outside the regulated area. Nevertheless, such a specification cannot control for unobserved supply shocks in different areas and demand shocks for different groups of consumers. Thus, we should be careful in drawing strong conclusions from a difference-in-difference specification of sales.

Aggregating Equation (3.2) to the seller-zip code level, we obtain the following difference-in-differences specification.

$$Y_{jwy} = \alpha_{wy} + \gamma_j + \theta_j t + \beta MC_{jwy} + u_{jwy} \tag{3.3}$$

The outcome Y_{jwy} measures how much sellers in zip code j sell in week w of year y. The variable MC_{jwy} is a dummy that takes value one if seller j is in a zip code inside the MC area and in a week-year in which MC is in effect. The parameter β is the coefficient of interest as it measures the effect of MC on sales of establishments within the MC area once the policy is in effect. The parameter α_{wy} is the week-year fixed effect, γ_j is the seller-zip code fixed effect, and $\theta_j t$ is a seller-zip code specific time trend. Therefore, the identification of the effect occurs conditional on the seller-zip code specific level and its trend. Finally, u_{jwy} is the usual error term.

Columns 1 and 2 of Table 3.5.4 show an insignificant increase of 5% in sales and a 6% increase in transactions, which is significant at the 5% level. Note that this is a rather convenient result for policy makers confronted with opposition by local business in

[26] Appendix Table 3.A.9 replicates Table 3.5.1 without buyer-seller-zip code fixed effects and shows the importance of controlling for the underlying variation across zip code pairs.

Table 3.5.4: Seller-zip code sales

	Total		B&M		Online	
	Rev (1)	Trans (2)	Rev (3)	Trans (4)	Rev (5)	Trans (6)
Madrid Central	0.054	0.060	0.059	0.044	−0.026	0.068
	(0.044)	(0.027)	(0.047)	(0.028)	(0.142)	(0.163)
Week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller trends	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,594	27,594	27,594	27,594	27,594	27,594

Notes: Standard errors are clustered at the seller-zip code level. The dependent variable is the log of revenue for total transactions, brick-and-mortar and online transactions at the seller-zip code level in a given week. The variable Madrid Central takes value 1 when a seller-zip code is within the MC area and the MC regulations are in place, and 0 otherwise.

the MC area: MC reduced traffic congestion and air pollution with no statistically significant impact on economic activity. In columns 3 to 6, we split the revenue and number of transactions at the seller-zip code level into brick-and-mortar and online transactions but find no significant effects. Together with the previous results, this could suggest a potential increase of the potential attractiveness of the MC area.[27]

3.5.3 Robustness checks

A concern with our triple difference specification is the potential existence of different pretrends. Therefore, different trends in the propensity of different buyer-zip codes to buy in each seller-zip code could invalidate the results in Table 3.5.1. To address this concern, in Table 3.5.5, we present results from a falsification test. We define an additional falsification variable which assumes that the introduction of MC took place approximately one month and a half before it actually did (it was introduced in week 49 of 2018 and we assume it was introduced in week 43).[28] We would expect to observe some effect if differential trends play a role. This is not the case. This result also allows us to rule out the existence of potential anticipatory effects that might induce consumers to bring forward consumption in MC area as a result of the imminent increase in transaction costs.

[27] Appendix Table 3.A.10 shows that excluding zip code-specific time trends yields no qualitatively different results except for online sales.

[28] The findings in Table 3.5.5 are robust to using other placebo starting dates such as 4 or 8 weeks before week 49 of 2018. These results are available upon request.

Table 3.5.5: Falsification Test

	B&M (1)	Online (2)	Total (3)
Treatment	-0.0895	0.121	-0.0385
	(0.0409)	(0.0606)	(0.038)
Falsification	0.00509	0.000127	0.0437
	(0.048)	(0.102)	(0.0482)
Buyer-week-year FE	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes
Observations	3,460,968	3,460,968	3,460,968

Notes: Standard errors are clustered at the buyer-seller pair level. The dependent variable is log of revenue and log of number of transactions for the full sample, brick-and-mortar and online transactions at the seller-zip code by buyer-zip code level in a given week. The variable Treatment takes value 1 when (1) a seller-zip code is within the MC area and the MC regulations are in place, and (2) the buyer-zip code is outside the MC area, and 0 otherwise. The variable Falsification is defined in the same way but assumes that MC was introduced six weeks earlier.

Another potential concern arises from the incidence of zeros in trade flows between some zip code pairs. So far, we have adopted the traditional solution of adding one to the dependent variable of interest to avoid dropping observations once we take logs. Table 3.5.6 shows an alternative approach using Poisson Pseudo Maximum Likelihood. This method accommodates zero trade flows with no transformations of the dependent variables since these are in levels (Santos Silva and Tenreyro, 2006). Column 1 shows very similar results for the effects on brick-and-mortar revenues. Column 2 finds a positive impact of MC on online revenues. Column 3 shows that the impact on total revenues is smaller in magnitude and statistically insignificant. This set of results is consistent with the main findings in Table 3.5.1.

A third set of robustness checks is concerned with the fact that differences in transportation costs may have changed differently across zip codes at the same physical distance. Table 3.5.7 presents further evidence consistent with the fact that an increase in transportation costs drives the substitution between brick-and-mortar and online consumption. Using the Google Maps Distance API, we calculate travel times by car and public transport from the centroid of each zip code to the geographic centroid of the MC area. We divide zip codes between those above and below the median of the increase in travel time. We observe that, although the decrease in brick-and-mortar

Table 3.5.6: Poisson Pseudo Maximum Likelihood

	B&M (1)	Online (2)	Total (3)
Treatment	-0.0797 (0.031)	0.0668 (0.0246)	-0.0286 (0.0299)
Buyer-week-year FE	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes
Observations	3,460,968	3,460,968	3,460,968

Notes: Standard errors are clustered at the buyer-seller pair level. The dependent variable is revenue for brick-and-mortar, online and total transactions at the seller-zip code by buyer-zip code level in a given week. The variable Treatment takes value 1 when (1) a seller-zip code is within the MC area and the MC regulations are in place, and (2) the buyer-zip code is outside the MC area, and 0 otherwise. Estimation by Poisson Maximum Likelihood.

sales is not statistically significant for zip codes with high and low travel time increments, those zip codes with a high increase in travel times drive the increase in online purchases.

We also test a number of alternative specifications. Appendix Table 3.A.11 shows results of dividing buyer zip codes into two groups, closer and farther than 6 km from the MC area, as well as results from dropping observations more than 15 km away from the MC area. Appendix Table 3.A.12 includes specifications where buyer-zip codes are weighted by their volume of sales. It also includes other specifications with week-year fixed effects interacted with the distance between each zip code and the center of the MC area to control for changes over time in spending behavior which correlate with geographic consumer location. Our results are qualitatively robust to these alternative specifications.

Table 3.5.7: Heterogeneity by Increase in Travel Time

	B&M (1)	Online (2)	Total (3)
Zip codes low increase	-0.0915	0.0906	-0.0546
	(0.0418)	(0.0635)	(0.0387)
Zip codes high increase	-0.0876	0.148	-0.0239
	(0.0448)	(0.0641)	(0.042)
Buyer-week-year FE	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes
p-val equal effects	0.893	0.151	0.276
Observations	3,460,968	3,460,968	3,460,968

Notes: Standard errors are clustered at the buyer-seller pair level. The dependent variable is level of revenue for brick-and-mortar, online and total transactions at the seller-zip code by buyer-zip code level in a given week. The variable Zip codes low increase takes value one when (1) the seller-zip code is within the MC area and the MC regulations are in place and (2) the buyer-zip code faces a low travel time increase, and 0 otherwise. The variable Zip codes high increase takes value one when (1) the seller-zip code is within the MC area and the MC regulations are in place and (2) the buyer-zip code faces a high travel cost increase, and 0 otherwise.

3.6 Conclusion

This chapter analyzes the benefits and costs of the introduction of constraints to vehicle circulation in the center of Madrid. By restricting access by car, transportation costs increase for those consumers living outside the area affected by the policy, potentially discouraging consumption spending in that particular area. We show that the regulation had the intended effect of reducing traffic congestion in the affected area, and consequently we observe a significant decrease in air pollution. This first set of results clearly states direct benefits from the implementation of MC in the city of Madrid.

However, our data allow for further investigation on the impact of the policy on economic activity. In particular, we use credit card transaction data from a large bank to examine whether consumers affected by the regulation reduced consumption spending in the city center of Madrid as a result of the increase in transportation costs. The granularity of our data grants the identification of purchases of all possible pairs of buyer zip codes and seller zip codes in the city of Madrid. Our findings show that there was not a statistically significant impact on total spending and number of transactions due to the policy. Yet, when we separate brick-and-mortar and online transactions, we find that brick-and-mortar spending and transactions by the directly affected consumers decreased while online spending and transactions by the affected consumers increased. The effect of the policy is larger for those zip codes where buyers face larger transportation constraints. This shows that when consumers face larger transportation costs, they switch from brick-and-mortar to online consumption spending.

Driving bans impose a cost on consumers by making shopping in brick-and-mortar establishments less attractive. While air quality improvements are significant and provide large benefits, brick-and-mortar commerce can be negatively affected. Our results show that, on aggregate, consumers substitute to online purchases, which could compensate the loss in brick-and-mortar spending. However, these substitutions are usually made at different types of sellers so that a driving ban might have unintended distributional effects on smaller businesses.

Thus, our work contributes to the literature in that it provides evidence of the impact of environmental policies on economic activity, more specifically, on spending and number of transactions of consumers in establishments directly affected by the policy. Most importantly, we offer evidence that these effects are not homogeneous and vary along different dimensions. A novel result in our analysis is the potential role played by e-commerce in attenuating the impact of environmental regulation, and its implication for policy makers regarding e-commerce and online transactions. Future research on the impact of environmental policies, regardless of the type of pollution regulated, should aim to provide direct evidence of their cost through diminished economic activity. Similarly, understanding the distributional effects of such policies is a crucial part of the information necessary for the design of future environmental regulations and their respective policy implementations. Furthermore, our results speak about the relevant role that e-commerce may play in smoothing the impact of increases in consumer transportation costs generated by other factors than environmental regulations. For instance, future research should study how consumers resorted to online purchasing during lockdown periods through the Covid-19 pandemic and how e-commerce adoption allowed establishments to weather such critical situation.

Appendix

3.A Additional Tables

TABLE 3.A.1: Descriptive statistics on weather conditions

	Mean (1)	SD (2)	Min (3)	Max (4)	Obs (5)
Temperature [°C]	15.63	7.74	3.39	30.87	616,297
Precipitation [0.1mm]	10.43	17.78	0	131.43	616,297
Cloud cover [okta]	3.45	1.76	0	7.57	616,297
Sunshine [h]	8.2	2.92	1.13	13.41	616,297
Pressure [hPa]	1017.25	6.09	998.21	1035.54	616,297
Humidity [%]	58.29	14.54	26.29	90.29	616,297
Wind speed [0.1 m/s]	19.73	8.46	0.43	75.43	616,297
$0° \leq$ Wind direction $< 45°$	0.19	0.19	0	1	616,297
$45° \leq$ Wind direction $< 90°$	0.17	0.18	0	0.71	616,297
$90° \leq$ Wind direction $< 135°$	0.1	0.14	0	0.86	616,297
$135° \leq$ Wind direction $< 180°$	0.05	0.1	0	0.57	616,297
$180° \leq$ Wind direction $< 225°$	0.1	0.14	0	0.86	616,297
$225° \leq$ Wind direction $< 270°$	0.22	0.21	0	1	616,297
$270° \leq$ Wind direction $< 315°$	0.12	0.15	0	0.86	616,297
$315° \leq$ Wind direction $< 360°$	0.06	0.1	0	0.86	616,297

Notes: The table shows descriptive statistics on weather conditions at each traffic station, where weekly weather is obtained from the closest weather monitor.

TABLE 3.A.2: Effects on traffic levels: Spillovers

	Vehicles per hour (1)	Time occupied [%] (2)	Utilized capacity [%] (3)	Log Vehicles per hour (4)	Log Time occupied (5)	Log Utilized capacity (6)
Madrid Central	-52.63	-1.987	-6.043	-0.153	-0.179	-0.241
	(12.14)	(0.85)	(0.777)	(0.0247)	(0.0496)	(0.0379)
Surroundings	-25.51	-0.318	-2.248	-0.0375	-0.00897	-0.0899
	(3.741)	(0.171)	(0.199)	(0.00623)	(0.0171)	(0.0109)
Station-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	456.3	6.571	20.19	5.648	1.493	2.905
N×T	597,221	596,895	596,328	597,031	592,874	571,518
N	3,948	3,948	3,948	3,948	3,927	3,823

Notes: Standard errors are clustered at the station level. The variables take value 1 when a station is located in the indicated area and the MC regulations are in place, and 0 otherwise. The weather controls are second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators.

TABLE 3.A.3: Effects on Air Pollution Levels: Poisson Regressions

	NO_2 > 40		
	(4)	(5)	(6)
Madrid Central	-0.178	-0.182	-0.158
	(0.0223)	(0.0248)	(0.0571)
Surroundings		-0.0255	-0.00185
		(0.0461)	(0.0683)
City of Madrid			0.0308
			(0.06)
Station-Week FE	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes
Mean dep. var.	0.46	0.46	0.46
N×T	6,381	6,381	6,381
N	33	33	33

Notes: Standard errors are clustered at the station level. The variables take value 1 when a station is located in the indicated area and the MC regulations are in place, and 0 otherwise. The weather controls are second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators. Estimation is done by Poisson regression.

Table 3.A.4: Regression of NO_2 levels with station-specific trends

	Log NO_2			$NO_2 > 40$		
	(1)	(2)	(3)	(4)	(5)	(6)
Madrid Central	-0.158	-0.157	-0.146	-0.135	-0.136	-0.130
	(0.0157)	(0.0173)	(0.0398)	(0.0165)	(0.0168)	(0.0301)
Surroundings		0.0149	0.0254		-0.00735	-0.00172
		(0.0299)	(0.0467)		(0.0618)	(0.066)
City of Madrid			0.0146			0.0078
			(0.044)			(0.0318)
Station-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	3.545	3.545	3.545	0.408	0.408	0.408
N×T	7,187	7,187	7,187	7,187	7,187	7,187
N	33	33	33	33	33	33

Notes: Standard errors are clustered at the station level. The variables take value 1 when a station is located in the indicated area and the MC regulations are in place, and 0 otherwise. The weather controls are second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators.

TABLE 3.A.5: Regression of NO_2 levels including Barcelona

	Log NO_2			$NO_2 > 40$		
	(1)	(2)	(3)	(4)	(5)	(6)
Madrid Central	-0.157	-0.147	-0.167	-0.103	-0.0847	-0.0723
	(0.0154)	(0.023)	(0.0185)	(0.0127)	(0.0185)	(0.0181)
Surroundings		0.0182	-0.00232		0.0339	0.0464
		(0.0315)	(0.0278)		(0.0256)	(0.0264)
City of Madrid			-0.0391			0.0239
			(0.0471)			(0.0398)
Station-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	3.563	3.563	3.563	0.418	0.418	0.418
N×T	8,708	8,708	8,708	8,708	8,708	8,708
N	40	40	40	40	40	40

Notes: Standard errors are clustered at the station level. The variables take value 1 when a station is located in the indicated area and the MC regulations are in place, and 0 otherwise. The weather controls are second order polynomials of temperature, precipitation, cloud cover, humidity, pressure and wind speed, as well as wind direction indicators interacted with station indicators.

TABLE 3.A.6: Effects on Transaction Values

	Total	B&M	Online
	(1)	(3)	(5)
Treatment	-0.00942	-0.0423	0.0265
	(0.0267)	(0.0288)	(0.0461)
Buyer-week-year FE	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes
Observations	3,460,968	3,460,968	3,460,968

Notes: Dependent variable: Log of mean transaction value for all transactions, B&M and online transactions. Standard errors are clustered at the buyer-seller pair level. The dependent variable is log of the mean transaction value (calculated as the ratio between total revenue over number of transactions) for all transactions, brick-and-mortar and online transactions at the seller-zip code by buyer-zip code level in a given week.

Table 3.A.7: Online Shares

	Share online revenue (1)	Share online transactions (2)	Ratio transaction values (3)
Treatment	3.380	1.434	-0.0141
	(0.675)	(0.495)	(0.143)
Buyer-week-year FE	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes
Observations	3,460,968	3,460,968	3,460,968

Notes: Dependent variable: Percentage share of online revenue, online number of transactions and ratios between online and B&M transaction values. Standard errors are clustered at the buyer-seller pair level.

Table 3.A.8: Changes in Travel Time

	Change in travel time when choosing public transport (minutes) (1)
Zip codes out of city	12.40
	(2.46)
Observations	126

Notes: The change in travel time is the difference between using public transport and or a car for traveling from a zip code centroid to the MC area. We regress the difference between travel time to MC by public transportation and car on a dummy if a zip code is out of city.

Table 3.A.9: No buyer-seller pair specific FE

	Total		B&M		Online	
	Rev	Trans	Rev	Trans	Rev	Trans
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.835	-1.909	-2.007	-1.994	-1.360	-0.779
	(0.158)	(0.161)	(0.163)	(0.163)	(0.133)	(0.0932)
Buyer-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer-seller-week FE	No	No	No	No	No	No
Observations	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968

Notes: Dependent variable: Log of revenue and log of number of transactions for all transactions, B&M and online transactions. Standard errors are clustered at the buyer-seller pair level. The dependent variable is log of revenue and log of number of transactions for all transactions, brick-and-mortar and online transactions at the seller-zip code by buyer-zip code level in a given week. The variable Treatment takes value 1 when (1) a seller-zip code is within the MC area and the traffic restriction is in place, and (2) the buyer-zip code is outside the MC area, and 0 otherwise. In all columns we control for buyer-week-year specific FE, and seller-week-year specific FE.

Table 3.A.10: Seller-zip code sales without trends

	Total		B&M		Online	
	Rev	Trans	Rev	Trans	Rev	Trans
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.0987	0.00298	-0.123	0.0302	-0.284	-0.419
	(0.148)	(0.0604)	(0.160)	(0.0606)	(0.118)	(0.0786)
Week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller trends	No	No	No	No	No	No
Observations	27,594	27,594	27,594	27,594	27,594	27,594

Notes: Dependent variable: Log revenue and log transactions for all, B&M and online transactions. Standard errors are clustered at the seller-zip code level. The dependent variable is the log of revenue for brick-and-mortar, online and total transactions at the seller-zip code level in a given week. The variable Treatment takes value 1 when a seller-zip code is within the MC area and the traffic restriction is in place. In all columns we control for week specific FE, and seller zip code FE. In columns 2, 4, and 6 we further include seller-zip code specific trends and trends squared.

Table 3.A.11: Robustness Results I

	All Sample			Zip codes within 15km		
	B&M	Online	Total	B&M	Online	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Zip codes < six km	-0.0719	0.0185	-0.0463	-0.0612	0.04	-0.0312
	(0.0436)	(0.0668)	(0.0407)	(0.0423)	(0.071)	(0.0398)
Zip codes > six km	-0.0948	0.152	-0.0361	-0.0939	0.147	-0.0387
	(0.0422)	(0.062)	(0.0393)	(0.0416)	(0.0698)	(0.0395)
Buyer-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes	Yes	Yes	Yes
p-val equal effects	0.413	0.0015	0.707	0.259	0.0291	0.792
Observations	3,460,968	3,460,968	3,460,968	1,575,050	1,575,050	1,575,050

Notes: Dependent variable: Log of revenue. Standard errors are clustered at the buyer-seller pair level. The variable Zip codes < six km takes value 1 when (1) the seller-zip code is within the MC area and the MC regulations are in place and (2) the buyer-zip code is outside the MC area but within six km of the MC area, and 0 otherwise. The variable Zip codes > six km takes value 1 when (1) the seller-zip code is within the MC area and the MC regulations are in place and (2) the buyer-zip code is further than six kms from the MC area, and 0 otherwise. In all columns, we control for buyer-week-year specific FE, seller-week-year specific FE and buyer by seller by week of the year FE.

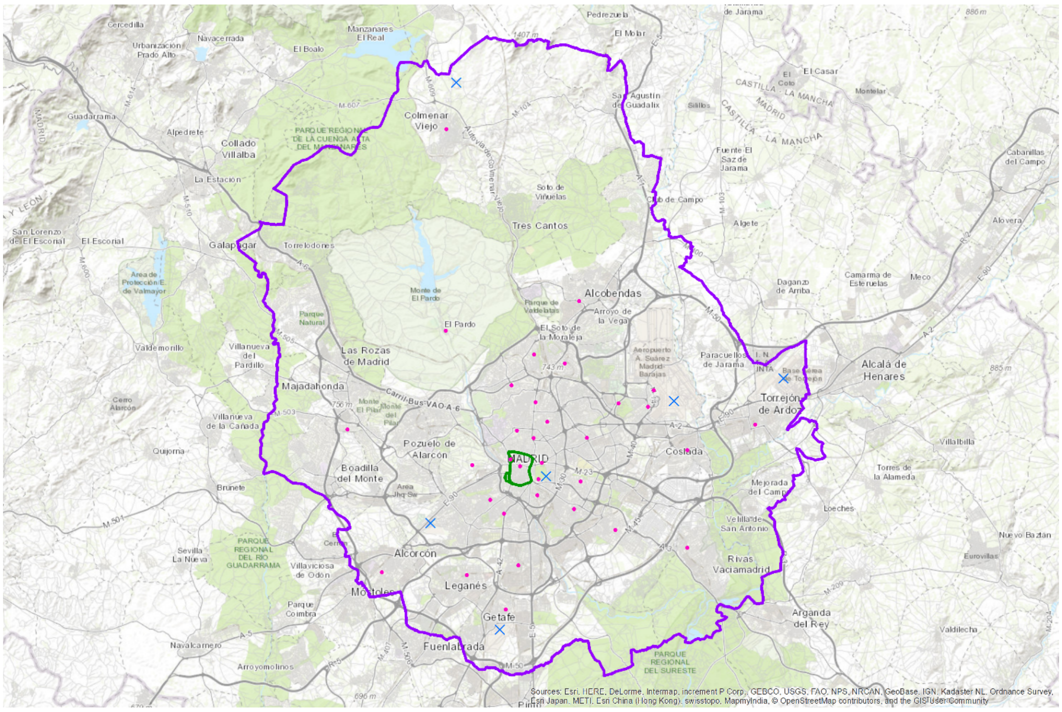
Table 3.A.12: Robustness Results II

	Weights			Extra Controls		
	B&M	Online	Total	B&M	Online	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Zip codes city	-0.0788	0.0413	-0.049	-0.0831	0.105	-0.0442
	(0.0338)	(0.0647)	(0.0322)	(0.0419)	(0.0628)	(0.0395)
Zip codes out of city	-0.138	0.186	-0.0847	-0.101	0.284	-0.0525
	(0.0357)	(0.0694)	(0.0337)	(0.0444)	(0.0639)	(0.0414)
Buyer-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller-week-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer-seller-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Week-year FE × zip code distance	No	No	No	Yes	Yes	Yes
p-val equal effects	0.0124	0.000946	0.115	0.519	0.00000444	0.766
Observations	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968	3,460,968

Notes: Dependent variable: Log of revenue. Standard errors are clustered at the buyer-seller pair level.

3.B Additional Figures

FIGURE 3.B.1: Map of stations



Notes: The map displays the locations of the 33 air quality monitoring stations within the city of Madrid. The pollution monitoring stations appear with pink circles in the map whereas weather stations appear in blue crosses.

Chapter 4

Killing Prescriptions Softly: Low Emission Zones and Child Health from Birth to School

Joint with Hannah Klauber, Nicolas Koch, Nico Pestel, Nolan Ritter, and Alexander Rohlf

4.1 Introduction

Poor air quality is a major public health concern worldwide. Various policy interventions have curbed ambient air pollution considerably over recent decades (Shapiro and Walker, 2018). To further improve air quality, policymakers in high-income countries increasingly focus on clean air programs in urban areas where motor vehicles are the major source of emissions. For instance, more than 200 European cities with already moderate pollution levels have implemented Low Emission Zones (LEZs) banning emission-intensive vehicles from entering city areas to achieve additional air pollution reductions. However, this raises questions about the returns to investments in such ancillary clean air regulations which impose significant costs while the specific health benefits from slight improvements in air quality remain uncertain. In particular, important latent health effects are likely to remain undetected. First, it is difficult to isolate causal effects in the presence of a health stock that changes slowly over time (Almond and Currie, 2011; Currie et al., 2014). Estimates of contemporaneous pollution impacts may substantially underestimate "fully formed" health benefits. Few studies address this issue indirectly by connecting pollution exposure at birth to productivity and earnings in adulthood (Isen et al., 2017). However, quasi-experimental evidence for the longer-term dose-response relationship between air pollution and physical health is missing. Second, in contexts characterized by a combination of high baseline population health, universal access to health care, and comparatively low pollution levels, health benefits are likely to take subtle forms that are hard to detect in the most widely used health measures such as low birth weight, hospitalizations, or mortality.

In this chapter, we examine the persistence of the impact of early-life exposure to air pollution on children's health stock from birth to school enrollment. Using quasi-experimental variation in the roll-out of LEZs in Germany and administrative health insurance data covering one third of the entire population, we test whether children born just before and just after policy-induced reductions in particulate matter concentrations exhibit persistent differences in rarely studied medication usage for up to five years after treatment. We characterize how health benefits accumulate over the course of pre-school childhood and disentangle immediate from longer-term health effects in response to lower pollution levels. By doing so, this work provides novel evidence on the adjustment trajectory of the health stock in the "middle years" between birth and school enrollment that have not been studied before (Almond et al., 2018). Our findings suggest that moderate improvements in air quality due to the adoption of LEZs persistently reduce the number and cost of prescriptions for respiratory diseases.

Our work draws on a data set and a setting which allow us to gain insight into whether pollution reductions below levels that are already low by the standards of high-income countries have lasting health benefits.[1]. Using medical records provided by Germany's largest public health insurer (AOK), we track children from birth to school enrollment and move beyond the usual measures of infant health. Our preferred outcome variables are the cumulative number and the cumulative costs of prescriptions for pharmaceuticals. Prescriptions are a sensitive, real-time, and monetizable

[1]Notably, even prior to LEZ implementation, the annual mean concentration of particles smaller than 10 μm (PM_{10}) of 26.4 μg/m^3 is close to the corresponding WHO guideline level of 20 μg/m^3

health measure that captures subtle effects, e.g. slightly reduced medication requirements manifesting only over a prolonged time frame. Because medication alleviates morbidity, its effects are likely undetectable when using the usual indicators such as hospitalizations. The German context is well suited for the analysis of pharmaceutical prescriptions because Germany's universal public health care system covers all prescription costs without any deductibles for minors. Mandatory insurance coverage implies that we observe health care for the general population rather than for a specific group with insurance, as is the case in pioneering studies using pharmaceutical data for the U.S. (Deschênes et al., 2017; Williams and Phaneuf, 2019; Deryugina et al., 2019) and China (Barwick et al., 2018).

The identification of causal effects rests on the staggered implementation of LEZs across 49 counties in Germany. *De facto*, LEZs constitute a ban of old, emission-intensive diesel vehicles from city areas.[2] Previous research has shown that LEZs are robust predictors of statistically significant reductions in local concentrations of particulate matter smaller than 10 μm (PM_{10}), which are however moderate in magnitude (Wolff, 2014; Gehrsitz, 2017; Pestel and Wozny, 2019; Margaryan, 2021). Thus, we exploit the staggered policy introduction as an instrumental variable for exogenous variation in PM_{10} exposure from motor vehicles. To isolate health effects, we follow Isen et al. (2017) and compare cohorts of children born just before and just after the policy-induced improvements in air quality. That way, cohorts experience different levels of pollution exposure in early life but the same levels throughout subsequent years. Our sample is limited to children from "non-attainment" counties that violated EU air quality standards regarding PM_{10}. While treated counties introduce LEZs, control counties comprise late-adopters and not-to-date-adopters.

Our research design accommodates important methodological advances in difference-in-differences (DID) settings with staggered policy introduction (Goodman-Bacon, 2018). We implement a "stacked DID" estimator (Fadlon and Nielsen, 2019; Deshpande and Li, 2019) that deals with the challenge of heterogeneous treatment effects inherent in the standard two-way fixed effect DID estimator. The stacked design aligns treatment events by event-time not calendar time. Therefore, it allows

[2]Restrictions are generally based on the Euro emission standards for gasoline and diesel vehicles. However, German LEZs have banned very few gasoline vehicles because of the prevalence of catalytic converters that transform pollutants into less toxic gases (Cyrys et al., 2014).

us to exploit a richer fixed effect structure: We remove (i) event-time trends caused by policymakers relying on local pollution or socio-economic trends in the years prior to the policy introduction as criteria to decide whether or when to introduce an LEZ, and (ii) time-invariant unobservables that may influence outcomes and the decision if or when to introduce an LEZ. Using this design, we demonstrate that pre-trends run parallel for all outcomes studied.

We produce three key findings. First, we present evidence that the number of prescriptions for children decreases significantly over early childhood in response to a moderate air quality improvement caused by the adoption of LEZs. A one $\mu g/m^3$ reduction in the concentration of PM_{10} *in utero* and in the first year of life decreases the total number of prescriptions for respiratory diseases over the five pre-school years by about 0.55 or by 3.9% in relative terms. The number of prescriptions decreases by up to 6.8% when analyzing the subgroup of pharmaceuticals closely tied to asthma. A decreasing share of children diagnosed with asthma (i.e. the extensive margin) drives the total reduction of asthma drug prescriptions and, thereby, also reduces prescriptions for respiratory diseases in general. However, we also find that children suffering from any kind of respiratory disease require less medication. This additional effect at the intensive margin suggests positive effects on children's respiratory health beyond asthma. Unconditional quantile regressions (Firpo et al., 2009) reveal suggestive evidence for substantial heterogeneity at the intensive margin. With a point estimate of 2.8 fewer prescriptions, children who suffer most from respiratory diseases may benefit nearly fives times as much from LEZs as the average sufferer.

Second, when analyzing prescriptions for each pre-school year separately, we find that contemporaneous reductions in medication needs account for only about 6–20% of the cumulative savings over the first five years of life. For asthma, the contemporaneous effect remains statistically insignificant. It is only from the second year of life that the initially latent health response materializes in fewer children diagnosed with asthma and, thus, fewer total prescriptions. This finding is consistent with the common notion of a health stock that adjusts slowly to changes in pollution levels which has rarely been documented because of a lack of real-time health measures. The observed delay in adjustment may also explain why prior studies on LEZs could not provide evidence for improved infant health measured by low birth weight (Gehrsitz, 2017; Pestel and Wozny, 2019). We corroborate this conjecture by using hospital data on fetal development disorders as alternative outcome variables for which treatment effects remain statistically insignificant.

Third, the identified health benefits yield economically meaningful cost savings for public health insurers. We find that a reduction of one $\mu g/m^3$ PM_{10} reduces costs by 10.54 € per child in their first five years of life. LEZs reduce PM_{10} levels by 1.37 $\mu g/m^3$ so that cost savings are 14.44 € per child. With average pre-treatment costs of 218.62 € per child, LEZs cause an economically relevant relative cost reduction of 6.6%. With 1,836,434 children protected by LEZs in their *in utero* period and their first year of life, treatment reduces pharmaceutical spending for children born between 2008 and 2017 over their pre-school years by approximately 26.5 million €, or 42.4 million € when accounting for positive spillover effects. Because all children benefit from LEZs from their second year of life onward, cost savings originate from a single life year with slightly improved air quality. While these savings represent only a fraction of the total policy benefits, they already account for about 22% (or 35% when accounting for spillovers) of the up-front costs of owners of vehicles that fail to meet LEZ standards calculated by Rohlf et al. (2020). Finding considerable health benefits, even at overall low pollution levels, suggests that reducing pollution can have large positive effects on children's respiratory health in many settings.

This work makes several contributions. First, our study is an important step forward in credibly estimating the sustained health benefits of lower exposure to air pollution in early life. There is robust evidence for particularly severe impacts of contemporaneous air pollution on infant and fetal health (Chay and Greenstone, 2003a,b; Currie and Neidell, 2005; Currie et al., 2009a; Sanders and Stoecker, 2015; Knittel et al., 2016). Motivated by the "fetal origins hypothesis," a few papers investigate the persistent effects of early-life exposure to air pollution on economic outcomes such as human capital formation (Sanders, 2012; Bharadwaj et al., 2017) and labor market outcomes (Isen et al., 2017) in the long-run. Relatedly, Simeonova et al. (2019) exploit the sequential introduction of Stockholm's congestion charge to demonstrate that health benefits increase with the *duration of exposure* to cleaner air. In comparison, we present evidence for the *persistence of health benefits* from exposure to cleaner air in a fixed period of time, i.e. the period before age one. By doing so, we provide a first quasi-experimental estimate for the longer-term dose-response relationship between PM pollution and *physical* health. We reveal a slowly adjusting health response over the middle years between birth and school that suggests that commonly used contemporaneous infant health measures may be too rough to detect health benefits from mild air quality improvements in the context of already low pollution levels.

Second, we provide the first quasi-experimental study that links moderate improvements in air quality from banning high-emission vehicles to health benefits across the child population. The few seminal papers on the health impact of traffic pollution focus on rare health outcomes (infant mortality, Knittel et al., 2016) and more disadvantaged populations (children living next to highway toll stations, Currie and Walker, 2011). Alexander and Schwandt (2019) exploit the dispersion of emissions-cheating diesel cars in the U.S. to quantify the morbidity costs of diesel pollution for a broader population of children. Given the low single-digit diesel share in U.S. passenger vehicles (only 1.5% of all light duty vehicles in 2014, U.S. DOT, 2015) it is unclear whether their estimates are generalizable. In contrast, we examine the link between child morbidity and diesel pollution in the context of German cities where (i) diesel vehicles are pervasive (45.9% in 2016, KBA, 2017) and a major source of PM, and (ii) initial pollution levels are low. Both characteristics are widespread in Europe and unaccounted for in prior work.[3]

The third feature that sets this study apart is its focus on the use and the costs of pharmaceuticals. Guided by the medical literature (Fanta, 2009) and a seminal paper by Deschênes et al. (2017), we argue that direct morbidity and mortality conditions are a function of pollution and compensatory adaptation in terms of drug therapy. Failing to account for the pharmaceutical expenditures means to underestimate the benefits of clean air policies. With the notable exceptions of Deschênes et al. (2017), Williams and Phaneuf (2019), and Deryugina et al. (2019), the effect of air quality on defensive pharmaceutical expenditures remains largely unevaluated. In our setting in Germany with universal health care access, our finding of significant longer-term reductions in defensive spending is relevant from a public finance perspective. Reduced pharmaceutical expenditures may lower insurance contributions, lower labor costs for employers, and increase net income for households.[4] The quantified policy benefits in terms of persistently lower defensive spending for children's medication are a complement to recent

[3]He et al. (2019a) provide a careful study in the context of exceptionally high pollution levels, showing a decrease in hospitalizations after the opening of a beltway diverting diesel trucks from passing through São Paulo.

[4]Presently, health care contributions are set to 14.6% of gross wages equally shared between employers and employees.

estimates of Pestel and Wozny (2019) and Margaryan (2021), indicating that LEZs also lead to contemporaneous reductions in hospital treatments and ambulatory care claims related to cardiovascular and respiratory diseases for all age groups. While hospitalizations and outpatient treatments of adults may also reflect medical histories linked to pollution exposure and confounding influences in the distant past, such confounders can be ruled out for infants observed from birth onward.

4.2 Low Emission Zones as a research design

In Europe, LEZs are the main instrument for cities to meet EU air quality standards which are among the strictest worldwide. To improve air quality and protect public health, the EU enacted several directives that set legally binding limits for criteria pollutants. Since 2005, for example, the annual mean of PM_{10} pollution must not exceed $40\,\mu g/m^3$. Moreover, daily PM_{10} readings must not exceed $50\,\mu g/m^3$ more than 35 times per year at any measuring station (Directive 2008/50/EC). For $PM_{2.5}$, the EU implemented legally binding limits only in 2015. EU Member States that violate these limits face considerable fines. In Germany, the 16 federal states are responsible for compliance with the EU air quality standards. In case of violations, state governments are obliged to develop city-specific Clean Air Plans (*Luftreinhaltepläne*). The implementation of LEZs is by far the most tangible compliance strategy the Clean Air Plans offer. LEZs explicitly target PM_{10} pollution. To date, 65 counties implemented them in a staggered process where the time of introduction of each individual LEZ depended on several idiosyncratic factors. First, the decision-making process usually involves the respective city administrations and city councils as well as other stakeholders, but state governments ultimately have to approve local Clean Air Plans.[5] They may overrule the decisions of local authorities and force cities to implement LEZs. Because of often conflicting interests between state and local policymakers, the length of the decision-making process regarding the introduction of an LEZ varies. Second, NGOs and private citizens frequently appeal to the courts to advocate against air quality regulations which creates further plausibly exogenous variation in the timing of LEZ

[5] At the federal level, regulations first had to (1) establish vehicle emission categories and (2) designate an official road sign for LEZs. This was done in late 2007.

introductions. Court rulings based on EU air quality legislation have generally sped up the adoption of LEZs.[6]

Figure 4.2.1 shows that almost every year since 2008 there have been waves of new LEZ introductions. Some LEZs cover entire counties while others rather ban emission-intensive vehicles from inner-cities. To secure access to LEZs, a vehicle must display an appropriately colored windscreen sticker based on EU-wide tailpipe emissions categories. The most emission-intensive diesel vehicles up to Euro1 standards (equivalent to 0.14g PM_{10} per km) are banned from LEZs. Petrol-driven vehicles are banned if they do not have a catalytic converter, which is very rare in Germany. Therefore, LEZs are *de facto* bans of old, emission-intensive diesel vehicles.[7] Police and local public order authorities enforce the policy with penalties for its violation of currently 100€.

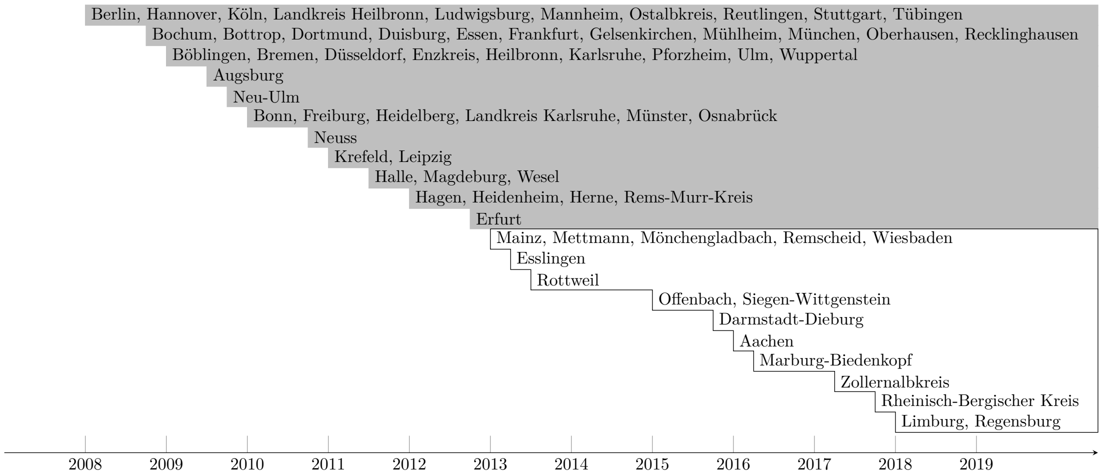
We exploit the temporal and spatial variation in the introduction of LEZs to break well-known sources of endogeneity in the link between health and pollution. For instance, local economic conditions not only affect ambient air pollution (Chay and Greenstone, 2003b) but also infant health (Dehejia and Lleras-Muney, 2004; Lindo, 2011). Therefore, we instrument changes in PM_{10} with the implementation of LEZs. Prior research finds that LEZ implementation decreased local PM_{10} concentrations by 4 – 9% (Wolff, 2014; Malina and Scheffler, 2015; Gehrsitz, 2017; Pestel and Wozny, 2019; Margaryan, 2021). Our approach of using a policy intervention as an instrumental variable is similar to the identification strategy by Chay and Greenstone (2005); Bento et al. (2015); Isen et al. (2017) who use county attainment status under the U.S. Clean Air Act as an instrument for changes in pollutant concentrations.

To address concerns that LEZs are not introduced randomly but in areas where air quality is deteriorating, we follow Wolff (2014) and include in our sample only non-attainment counties, that failed to meet the PM_{10} limits. Of these counties (i) 65 implemented an LEZ and (ii) 63 have not implemented an LEZ to date despite non-

[6]Examples from Wiesbaden and Halle shed light on the different paths leading to the introduction of LEZs. In late 2010, Wiesbaden's city council proposed the introduction of an LEZ which the state government of Hesse rejected. After the appeal of an environmental interest group, Wiesbaden's administrative court ruled that the state government had to approve the LEZ in order to comply with air quality standards as fast as possible (Wiesbadener Tagblatt, 2011). In contrast, the city of Halle appealed plans by the state government of Saxony-Anhalt to implement an LEZ. The city council decided to abandon their case against an LEZ after it was made clear that the EU would otherwise penalize the city almost a million € per day (Mitteldeutsche Zeitung, 2011).

[7]Less emission-intensive vehicles with Euro2 and 3 standard are eligible for red or yellow stickers while cars that meet Euro4 standards are eligible for green stickers.

FIGURE 4.2.1: The Staggered Implementation of LEZs



Notes: The figure depicts introduction dates for all counties with an LEZ. The eleven implementation waves considered in this work are marked in gray.

compliance. Only the 49 counties that implement their LEZ until 2012 count towards the treated because we want to be able to follow children after LEZ implementation for five years. From the sample of non-attainment counties, we construct for each group of treated counties that introduces LEZs in a given year-quarter a different "clean" control group consisting of later-adopting counties and not-to-date-adopters (see Section 4.3.2). Our identifying assumption is that in the absence of LEZ introduction, air pollution and health outcomes in treated counties would have evolved similarly to the control counties.

To test the validity of our identifying assumption, we use a dynamic difference-in-differences design that aligns treatment events by event-time and not calendar-time, which allows for a richer fixed effect structure (see Section 4.3.2). First, we can remove LEZ wave-specific time trends that may result if policymakers use local pollution or socio-economic trends in the years prior to implementation as criteria for introducing LEZs in a given year-quarter. Second, we are able to remove time-invariant differences between treatment and control groups from each wave of LEZ implementation that could be driving outcomes and selection into LEZ adoption and earlier or later adoption. Conditional on our fixed effects, we consistently show that there are no differences in the trends of air pollution and health outcomes across the treatment and control groups before LEZ implementation.

4.3 Methodology

4.3.1 Isolating early-exposure effects

Our goal is to estimate the causal effect of PM exposure in the *in utero* period and the first year of life on later life health outcomes measured before school enrollment. Due to rapid cell proliferation and an intense phase of epigenetic programming, children in the prenatal and immediate postnatal development period are especially vulnerable to the toxicological effects of pollution (Holt, 1998; Šrám et al., 2005; Gluckman et al., 2008; Baccarelli and Bollati, 2009). While the staggered introduction of LEZs provides us with an instrument to address concerns regarding the endogeneity of pollution exposure, the additional empirical challenge is to isolate the long-term effect of PM exposure before age one from any exposure throughout the subsequent lifetime. We want to compare individuals who are exposed to different levels of PM pollution up to age one but who are exposed to the same levels of PM pollution thereafter.

To this end, we resort to a cohort study design as proposed by Isen et al. (2017). Our analysis compares differences in health outcomes of children born just before and just after LEZ implementation relative to the difference in health outcomes between children from the control group born at the same times. We restrict our sample so that within each county all children experience the same pollution levels after age one. Of the cohorts born after LEZ implementation we include all those born at least four quarters post-treatment. LEZs protect these children from conception onward. Of the cohorts born pre-treatment we include the cohort born exactly four quarters prior to LEZ implementation. These children are not protected before age one but they are protected thereafter (see Figure 4.3.1). We exclude children born in the three quarters prior or post to LEZ implementation from the sample because they are partly treated *in utero* or during their first year of life. Note that to ensure equal exposure after age one, we would need to exclude all cohorts born more than four quarters prior to LEZ implementation. To avoid limiting ourselves to a single pre-treatment observation of the treated counties, we keep the two cohorts born five and six quarters prior to LEZ

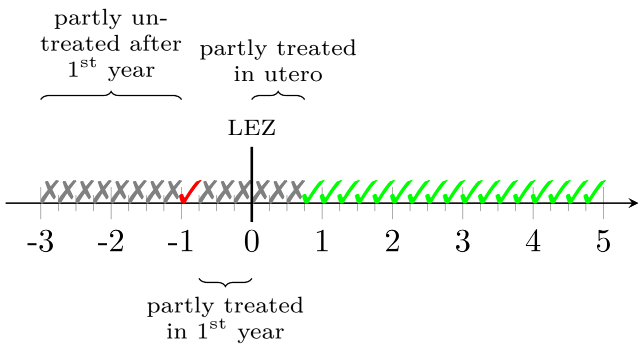


FIGURE 4.3.1: Isolating early-exposure effects

Notes: The figure shows which cohorts are included in a given implementation wave. The numbers on the timeline measure years with respect to the introduction date defining the implementation wave of interest. The green marks indicate the post-treatment cohorts while the red mark indicates the pre-treatment cohort.

implementation in our sample and ensure that this does not influence our results in a robustness check.

4.3.2 Stacked difference-in-differences design

In what follows, we describe our research design that combines the cohort data with the staggered introduction of LEZs as an instrumental variable. To this end, we use a stacked difference-in-differences (stacked DID) design (Deshpande and Li, 2019; Cengiz et al., 2019; Fadlon and Nielsen, 2019) that is well suited to deal with the challenge of heterogeneous treatment effects in settings with staggered policy introductions.

Of the 65 counties that implemented LEZs, 49 adopted them at some point between 2008 and 2012. We estimate treatment effects for these 49 counties to ensure that we can follow children born after the implementation of the ban for at least five years. The 16 other counties that adopt LEZs after 2012 are potential controls.

For the period between 2008 and 2012, there are 11 distinct LEZ *implementation waves*, i.e. year-quarters in which at least one county implemented an LEZ. Subscript j denotes these unique implementation waves. We create an individual data set for each implementation wave.

Counties that introduce LEZs in the year-quarter defining the respective implementation wave count towards the treatment group. Of these counties we include cohorts of children born in the pre- and post-treatment periods defined in Figure 4.3.1.[8]

Counties that do not introduce an LEZ in implementation wave j, either because they do so at a different point in time or never, are eligible for the control group. Because we expect that the LEZ treatment effect on PM concentrations levels off one year after implementation[9] and because we want to have a balanced control group (see discussion in Section 4.5.5), we define an exclusion window to further refine the selection of control counties. Given our choice of an event time window of up to three years prior and five years subsequent to LEZ implementation in the event study specification, we determine the exclusion window to start one year earlier (-4 to +5 years). For the control group, we keep only the counties that do not implement an LEZ within this period. Thus, our control group only holds LEZ-counties that implement the policy measure at least four years before or five years after treatment wave j. For instance, the county Mainz serves as a control unit for Mannheim because it implements its own LEZ more than five years later, in 2013. Likewise, Mannheim serves as a control unit for the county Hagen because it implemented its own LEZ already four years earlier (see Figure 4.2.1). We stack the data sets for the 11 implementation waves j for a pooled regression.

The selected event time window of three years pre-treatment and five years post-treatment is subject to an important trade-off. The longer the time window, the longer we can observe deviations from the parallel trend assumption and effects of LEZs post implementation. The shorter the time window the higher the statistical power. However, the event time window also determines the length of our exclusion window which defines our control group. A shorter event time window, thus, also means that a higher number of LEZ adopters serve as control observations. Our preferred choice of the event time window is based on the dates of the implementations of the LEZs, the availability of data, and the observation of cumulative benefits over five years. To rule out that this choice drives our results, we provide estimates for alternative time-window specifications in the robustness analysis.

[8]The number of included cohorts varies across implementation waves because our sample only comprises children born between 2006 and 2012.

[9]Results are almost identical if we instead exclude all already-treated from the control group to allow for persistent dynamics. See Figure 4.B.1 in the Appendix.

Our stacked DID design addresses concerns about bias arising from the combination of potentially heterogeneous treatment effects and the weighting implied by the two-way fixed effects regression in DID settings with staggered policy introduction (Goodman-Bacon, 2018; Athey and Imbens, 2018; de Chaisemartin and d'Haultfoeuille, 2020). Goodman-Bacon (2018) shows that the two-way fixed effect DID estimator consists of comparisons between all combinations of early treated, late treated, and untreated units. Each comparison is weighted by the size of the subgroup and the variance of treatment. If treatment occurs either early or late, this results in lower treatment variance and, hence, a lower weight. Thus, units that are treated in the middle of the study period have higher weights in the regression than those treated at the beginning or at the end. Because the effectiveness of treatment may vary across implementation waves j, e.g. early adopted LEZs may be more effective at reducing PM emissions than the later adopted as the stock of old diesel vehicles decreases over time, we have to ensure that the weighting of cohorts does not bias our average treatment effect.

Our stacked DID design aligns treatment events by event-time, not calendar time. This results in a setting that is equivalent to one where the treatment events occur all at the same time instead of in a staggered fashion. This prevents the unintended weighting of events driven by the variance of treatment (see also Abraham and Sun, 2020). Regression weights now mainly depend on the size of the subgroup. In addition, the stacked DID design prevents potential bias from using units as controls that have been treated shortly before and might yet be on differential trends.

4.3.3 Estimation equations

First-stage estimation The first-stage regression of the instrumental variable (IV) estimator is a stacked DID regression model specified as

$$P_{ctj} = \alpha(Treated_{cj} \times Post_{tj}) + \sum_{\tau} \delta_j^{\tau} D_{tj}^{\tau} + \lambda_j D_{cj} + W'_{ctj}\rho + X'_{c\bar{t}j}\pi_t + \gamma_c + \eta_{st} + \nu_{ctj} \tag{4.1}$$

where the dependent variable P_{ctj} is the mean PM_{10} exposure in µg/m^3 during the first quarter of life of a cohort born in county c and year-quarter t for treatment wave j. The binary variable $Treated_{cj}$ is equal to 1 if county c introduces an LEZ in implementation wave j. $Post_{tj}$ is a binary variable equal to 1 if quarter t is after the implementation quarter of wave j. For every implementation wave, the indicators D^{τ}_{tj} are equal to 1 if quarter t is τ quarters before or after the quarter of LEZ implementation j, where $\tau \in \{-6, -5, -4, 4, 5, 6, \ldots, 19\}$ (see Section 4.3.1).[10] Every wave j also has its own indicator D_{cj} which is equal to 1 if county c introduces an LEZ specifically in that implementation wave. The vector W'_{ctj} comprises weather controls in the county of birth c. Weather controls are exogenous and, thus, included for every year of life. The county fixed effects γ_c control for time-invariant, unobserved determinants of pollution exposure for children born in county c. The fixed effects η_{st} account for time-varying determinants of pollution exposure that are common to all children born in state s in year-quarter t; we refer to them as birth state–birth quarter fixed effects.[11] The coefficient $\widehat{\alpha}$ provides a difference-in-differences estimate of the impact of LEZ implementation on quarterly PM_{10} levels at the county level.

In some specifications, we also include pre-treatment controls for socio-economic characteristics (e.g. population density, employment, income, and transfers) interacted with year-quarter fixed effects denoted by $X'_{c\bar{t}j}$. We generate these controls by interacting year-quarter dummies with terciles of the variables measured in $\bar{t} = 2007$, the year prior to the first LEZ implementations (Barrot and Sauvagnat, 2016).[12] Observing that the estimated treatment effect changes significantly after allowing for these trends that vary with the levels of the exogenous covariates would suggest that the results are driven by differential trends in pollution across socio-economic characteristics (see Jaeger et al. (2020) and Hoynes et al. (2016) for similar approaches).

Our stacked DID design enables us to remove important unobservables that may simultaneously drive treatment selection and outcomes. First, the indicators D^{τ}_{tj} in Equation (4.1) remove LEZ wave-specific unobservables that appear in event-time rather than calendar-time. This eliminates trends that emerge, for instance, if policymak-

[10] For instance, $\tau = 4$ refers to the fourth quarter after LEZ implementation. Similarly, $\tau = -4$ refers to the fourth quarter prior to LEZ implementation.

[11] Note that the counties of Berlin, Hamburg, and Bremen are states in their own right. Including state–quarter fixed effects absorbs the variation explained by the treatment in these counties. Therefore, we allocate these three counties to neighboring states, i.e. we assign Berlin to Brandenburg, Hamburg to Schleswig-Holstein, and Bremen to Lower Saxony. We ensure that this allocation does not determine the results by estimating alternative specifications.

[12] Appendix 4.D.1 holds a detailed description of the control variables.

ers use local pollution or socio-economic trends in the years prior to implementation as decision criteria whether and when to introduce LEZs. Accommodating calendar time effects will not eliminate such pre-trends. One could restrict the effect of the wave-specific event-time binary variables D^{τ}_{tj} to be equal across treatment waves, that means replacing δ^{τ}_{j} by δ^{τ}. However, we would then only remove the variation in event-time pooled over all implementation waves. Second, by including D_{cj} and allowing its effect to vary by implementation wave, we remove time-invariant differences between treatment and control groups for each LEZ implementation wave j and between different implementation waves.[13] This accounts for time-invariant unobservables that may drive outcomes and selection into LEZ adoption and earlier or later adoption. In case we only included one wave-unspecific binary variable $Treated_{cj}$, we would only control for these differences between the pooled treatment and the pooled control group.

Second-stage estimation In the second stage, predicted values for PM_{10} from Equation (4.1) serve as an explanatory variable in

$$H_{ctj} = \beta\widehat{P_{ctj}} + \sum_{\tau} \sigma^{\tau} D^{\tau}_{tj} + \psi_j D_{cj} + W'_{ctj}\kappa + X'_{c\bar{t}j}\mu_t + \gamma_c + \eta_{st} + \epsilon_{ctj} \tag{4.2}$$

where the coefficient of interest β describes the marginal effect of a one $\mu g/m^3$ LEZ-driven increase in PM_{10} exposure on the average health outcome H_{ctj} of a cohort born in county c and year-quarter t. The health outcome is accumulated over the first five years of a child's life and is averaged over the children in cohort ct. We obtain the health outcome through an auxiliary regression that exploits the available information at the level of the individual child (i.e. gender and area of residence within a county at the five-digit zip code), which controls for some of the observed heterogeneity in individual health when aggregating the data.[14] All regressions are weighted by the number of children in a cohort (see Appendix 4.C). We cluster standard errors at the county level, the level at which the treatment is assigned.

[13] The latter is only true because we include D^{τ}_{tj} as well.

[14] Aggregating individual level data *via* auxiliary regressions and conducting regressions on the obtained aggregates is a common approach in the literature (e.g. Baker and Fortin, 2001; Shapiro, 2006; Albouy, 2009a,b; Angrist and Lavy, 2009; Currie et al., 2015; Isen et al., 2017; Notowidigdo, 2020) and asymptotically equivalent to using the individual level data itself (e.g. Donald and Lang, 2007). We provide a detailed description of the procedure in the Appendix 4.C.

The unbiasedness of $\widehat{\beta}$ depends on two crucial assumptions. First, LEZs need to be a strong instrument for changes in particulate matter pollution. We subsequently present evidence for a strong first-stage, showing that PM_{10} levels decline significantly and persistently in response to LEZ implementation. Second, for consistency, it must be the case that LEZ introduction affects health outcomes only via its impact on air pollution and not, for instance, through changes in the population's composition or behavior. Given that our results are largely insensitive to the choice of controls, and conditional on the full set of fixed effects, we believe that the exclusion restriction holds. While we cannot conclusively show its validity, we conduct various indirect tests. First, we comprehensively assess whether LEZs shift the composition in the underlying population in LEZ counties but find no evidence thereof. Second, we re-estimate our IV model for medical conditions unrelated to air pollution. These placebo tests provide further evidence in favor of the validity of our IV design. Third, we explore the unlikely possibility that the compliance costs of LEZs reduce industrial activity and, thereby, emissions of other health-damaging local pollutants, most notably sulfur dioxide (SO_2). We find no evidence for any statistically significant effects on SO_2. Finally, the exclusion restriction may be violated if LEZs affect pollutants other than PM_{10} that also have a direct impact on respiratory health (see discussion in Section 4.5.5). This is a potential threat to all analyses in which a single instrument could affect multiple pollutants. In fact, we show that LEZs reduce both PM and NO_2. However, we also show that the policy exclusively affects these pollutants and argue that our identification strategy suffices to identify the health effects of pollutants from motor vehicles rather than of PM_{10} in particular. Similarly, other studies attribute estimated IV-effects to air pollution more generally rather than to a particular pollutant (Chay and Greenstone, 2003a; Currie and Neidell, 2005; Arceo et al., 2016; Knittel et al., 2016; Deryugina et al., 2019; Sager, 2019; Colmer et al., 2020).

Event-study specification We estimate event-study specifications of our stacked DID model to assess the parallel trends assumption. To this end, we expand the pre-treatment window of our sample to include all observations up to three years prior to LEZ implementation. Furthermore, we include observations from the three quarters before and the three quarters after LEZ implementation, so that $\tau \in \{-12, \ldots, 19\}$. To gain precision and prevent noise potentially linked to estimating 32 quarter coefficients, we group the event study coefficients at the year level throughout this chapter. By doing so, the first stage becomes

$$P_{ctj} = \sum_{\upsilon} \theta_{\upsilon}(Treated_{cj} \times D^{\upsilon}_{tj}) + \sum_{\tau} \delta^{\tau} D^{\tau}_{tj} + \lambda_j D_{cj} + W'_{ctj}\rho + X'_{c\bar{t}j}\pi_t + \gamma_c + \eta_{st} + \omega_{ctj} \quad (4.3)$$
$$\text{with} \quad \upsilon = \lfloor \frac{\tau}{4} \rfloor$$

The parameter of interest is θ_{υ}. It captures the marginal effect of LEZs on the mean PM_{10} concentration in year υ prior or post to treatment. We set $\theta_0 = 0$ so that the year prior to LEZ implementation is the reference category. The figures presented in the following sections plot the θ_{υ} estimates in event time. The main difference between the standard two-way fixed effect event study and our dynamic estimator in Equation (3) is that we eliminate time-invariant unobservables both within and between LEZ implementation waves by including D_{cj} as well as wave-specific event-time trends that do not appear in calendar time by including D^{τ}_{tj}.

We also estimate an event study specification that explores the treatment effect dynamics for the long-run health outcomes. Replacing the dependent variable in Equation (4.3) with H_{ctj} results in a reduced-form event study model. This specification allows us to examine how LEZs affect long-run health depending on a child's age at the time of exposure. Because the year prior to LEZ implementation is the reference category, we essentially test for differential effects of exposure relative to exposure at age one and older.

4.4 Data

Air pollution and LEZ data The German Environment Agency (UBA) provides data on air pollution for the years 2006 through 2013 for 128 counties that violated EU-wide limits for PM_{10}. Specifically, monitoring stations record daily concentrations of regulated air pollutants particulate matter (PM_{10}), nitrogen dioxide (NO_2), ozone (O_3), and sulphur dioxide (SO_2). For the sample period, there is no consistent data on fine particulate matter ($PM_{2.5}$) concentrations available from UBA because there were no legally binding thresholds for $PM_{2.5}$ before 2015. Stations located at the roadside primarily measure peak pollution exposure from traffic, while the remaining stations measure the permanent exposition to pollution in residential areas (UBA, 2019). We combine measurements from both types of stations and interpolate the point measures into county space using Inverse Distance Weighting (Currie and Neidell, 2005; Karlsson and Ziebarth, 2018). Based on the daily records, we construct a weighted average of quarterly levels of PM_{10}, NO_2, O_3, and SO_2. Following recent studies (i.e. Chay and Greenstone, 2003a; Isen et al., 2017), we use weights proportional to the number of monitor observations within a quarter and limit our data to stations with at least 60 measurements per quarter (see Appendix 4.D.2 for a detailed description). UBA also provides data on the exact dates at which LEZs are implemented.

Health care data The health data comes from Germany's largest public health insurer AOK and is at the level of the individual child and the individual quarter. It provides us with information on pharmaceutical prescriptions and in-hospital doctoral care.[15] We obtained access to the anonymized data through the "AOK Research Institute" (WIdO). The data holds information on about a third of all children born in Germany, approximately 200,000 annually. The sample includes the full medical records of all children born between January 1, 2006, and December 31, 2012 until the end of 2017, exploiting the full range of data available at WIdO when conducting this research. This allows for observing all children in the sample from birth until age five. Because our analysis focuses on early childhood pollution effects, we exclude children from our sample who move out of their birth county within the first year of life. Likewise, we exclude children who are not continuously insured with AOK until the age of five. Overall, our sample holds observations on about 1.1 million children across all counties in Germany. The sample of counties underlying our analysis that either implement or consider implementing an LEZ comprises 550,000 children.

The focal point of our analysis are respiratory health outcomes. It is well established in the medical literature (e.g. Li et al., 2003) that PM pollution causes inflammations in the respiratory tract that may irreversibly reduce lung growth and function. In consequence, affected children are more vulnerable to suffer from respiratory problems in general and are at a higher risk of developing severe chronic diseases such as asthma later in life. Because respiratory diseases are primarily treated with pharmaceuticals,

[15]Note, that we only observe stationary hospital treatments that accommodate overnight cases.

we restrict the main analysis to prescriptions. Pharmacies electronically provide health insurers with data on prescriptions (Swart et al., 2005). These data hold information on costs and pharmaceutical substances classified according to the ATC-Code system.[16] In Germany's universal public health care system, all prescription costs are covered by the insurance without any co-payments for children. This applies to over-the-counter drugs as well, as public insurances are legally obligated to bear the expenses for children up to the age of 12 (§34 Abs. 1 Satz 5 SGB V).

Prescription data are not linked to ICD-10 codes[17] that reveal the diagnoses for which pharmaceuticals are prescribed. Therefore, it is necessary to identify the relevant pharmaceutical substances for the therapy of respiratory diseases. To this end, we follow two different approaches. First, we use a publication akin to the Red Book ("Gelbe Liste") to link pharmaceuticals and diagnoses for more than 120,000 drugs. This provides us with a broad group of pharmaceuticals used to treat respiratory diseases. Second, we consider the 20 most often prescribed substances for asthma and chronic obstructive pulmonary disease (ATC R03) in a given year. This returns a small, strict subset of the pharmaceuticals identified in the first approach. Appendix 4.D.3 describes the procedures in detail. Costs of prescriptions are in real values normalized to the fourth quarter of 2017 and account for market price changes, such as expiring patents (see Appendix 4.D.3).

We complement the analysis of respiratory health outcomes with placebo tests. To this end, we consider three outcomes that are independent of air pollution but correlated with socio-economic status: hospital treatments of injuries of the head (S00-S09), the arm (S40-S49), or several body parts (T00-T07). Additionally, we use hospital data on pregnancy duration and fetal growth (ICD-10 codes P05, P07, and P08) in our analysis of common infant health measures.[18] On the insured themselves, the medical

[16]The Anatomical Therapeutic Chemical (ATC) classification system categorizes drugs based on their active ingredients according to the organ or the system on which they act as well as their therapeutic, pharmacological, and chemical properties. It is compiled by the World Health Organization (WHO) and adapted to the German market on an annual basis (Swart et al., 2005).

[17]The ICD-10-Code is an international system for the statistical classification of diseases and related health problems provided by the WHO. Germany uses the extended version ICD-10-GM. Outpatient and inpatient physicians are legally required (§§295 and 301 SGB V) to classify diagnoses accordingly.

[18]In accordance with common practice, we consider the discharge diagnosis as the main reason for hospitalization (Swart et al., 2005).

TABLE 4.4.1: Summary Statistics of Pollution and Health Outcomes

	(1) mean	(2) sd	(3) min	(4) max	(5) N
Air pollution					
PM_{10} (μg/m^3)	23.7	6.2	9.0	57.3	3,584
Prescriptions for respiratory diseases					
Number of prescriptions over five years per child	12.7	2.8	4.4	23.8	3,584
Prescription expenditures over five years per child (€)	195.8	54.1	63.6	635.3	3,584
Share of sufferers per cohort (%)	76.1	7.5	38.3	94.6	3,584
Prescriptions for asthma					
Number of prescriptions over five years per child	2.0	0.8	0.1	6.1	3,584
Prescription expenditures over five years per child (€)	62.2	33.9	1.6	489.8	3,584
Share of sufferers per cohort (%)	18.3	6.1	1.3	50.6	3,584
Number of children per cohort	159.3	171.6	10	1,593	3,584

Notes: The table reports summary statistics of PM_{10} pollution (measured in μg/m^3) and of cumulative prescriptions over the five pre-school years linked to a broad group of respiratory diseases and asthma specifically. The variables are defined for our study period from 2006 to 2012 and our sample of 128 German counties that violated EU-wide limits for PM_{10}. Cumulative prescriptions are calculated based on data until 2017. Health measures are in terms of the number or the costs of prescriptions per child. Costs of prescriptions are in real values normalized to the fourth quarter of 2017. The share of sufferers reflects the share of children in the cohort that require at least one prescription for a respiratory disease or asthma, respectively.

records additionally offer information about the birth dates, gender, and the precise location of residence within a county at the five-digit zip code.

Table 4.4.1 provides summary statistics of the prescription data for respiratory diseases and pollution levels at the county and quarter level. Overall, the cross-sectional dimension of our data covers 128 counties, while the longitudinal dimension covers 28 quarters from 2006 to 2012. A detailed description is provided below the table.

Additional data on weather and county characteristics Weather is a strong correlate of pollution and health (Karlsson and Ziebarth, 2018). Therefore, we obtain data from the German Weather Service (DWD) on a battery of weather phenomena at the level of the individual weather station as well as in an interpolated grid format. We combine weather, pollution, and health data at the county level. In the regressions, we include linear and quadratic terms of precipitation as well as 12 temperature bins[19]

[19]Temperature bins count the number of days with temperatures above 0, 5, 10, 15, 20, 25, 29, 30, 31, 32, 33 and 34 degrees Celsius.

in addition to mean temperature, sunshine duration, relative humidity, pressure, and wind speed. We include additional control variables from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) such as socio-economic and demographic characteristics as well as information on the age of mothers giving birth. We use pre-treatment measures of these variables in 2007 and categorize them into terciles. For a detailed description see Appendix 4.D.1.

4.5 Results

4.5.1 Ambient air pollution

First, we present evidence that the relationship between LEZ implementation and particulate matter levels is strong. Table 4.5.1 shows estimates of the effect of LEZ introduction on average quarterly PM_{10} concentrations following Equation (4.1) for three different sets of control variables that increase in stringency from left to right. All regressions include birth county and birth state–birth quarter fixed effects. Standard errors clustered at the county level are in parentheses. The reported mean outcomes represent weighted averages for the dependent variable in the pre-treatment period over all treated counties with weights equal to the number of children per county.

Our most stringent specification shows that the presence of an LEZ reduces mean quarterly PM_{10} concentrations by about 1.37 $\mu g/m^3$. This estimate is statistically significant at the 0.1% level and robust across control sets. The stability of our treatment effect estimates after including heterogeneous trends that vary with the pre-treatment levels of the socio-economic controls (see Section 4.3.3) suggests that the LEZ effect is not driven by differential trends in pollution across socio-economic characteristics. Compared to the mean pollution exposure of 26.44 $\mu g/m^3$, an LEZ decreases particulate matter by about 5.2%, which is in line with the LEZ literature. An F-statistic of 20.14 provides evidence for a fairly strong first stage relationship. To accommodate remaining concerns about potential bias from weak instruments (Andrews et al., 2019; Lee et al., 2020), we subsequently also report robust Anderson-Rubin (AR) confidence intervals for our IV-estimates.

TABLE 4.5.1: The Effect of LEZ Implementation on PM_{10} Concentrations

	First Stage Estimation PM_{10} Pollution in μg/m^3	
	(1)	*(2)*
LEZ treatment	-1.30	-1.37
	(0.34)	(0.30)
Mean outcome	26.44	26.44
First stage F-statistic	14.25	20.14
Weather controls	x	x
Socio-economic controls		x

Notes: This table reports coefficients from two variants of the first stage regression in Equation (4.1). The dependent variable is the quarterly mean PM_{10} concentration in a given county and year in μg/m^3. Both columns include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 9, 609.

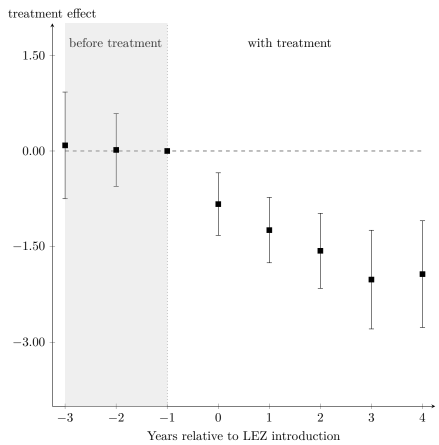
Figure 4.5.1 plots the event-study results for the effect of LEZs on PM_{10}.[20] The post-treatment patterns suggest that LEZs cause a level shift to persistently lower PM_{10} concentrations with a strong immediate effect. Moreover, coefficients prior to treatment are close to zero and statistically insignificant, which is in line with common trends in LEZ and non-LEZ counties in the years preceding the policy interventions. Recall that our reported event study coefficients exploit a rich fixed effect structure to remove (i) event-time trends caused by policymakers relying on local pollution or socio-economic trends as criteria to decide whether and when to introduce an LEZ and (ii) time-invariant unobservables that may drive outcomes and selection into LEZ adoption and earlier or later adoption.

4.5.2 Medication of respiratory diseases

In the following, we show the persistent effects of lower PM_{10} pollution on the medication of respiratory diseases. Our analysis comprises two groups of prescriptions to detect health effects of varying severity. To comprehensively capture respiratory health effects, we consider a broad group of pharmaceuticals. To capture effects re-

[20] Table 4.A.1 in the Appendix provides the full regression results.

FIGURE 4.5.1: Event-study Estimates of the Effect of LEZ Implementation on PM_{10} Concentrations



Notes: The figure presents event-study coefficients from Equation (4.3) that show how LEZs affect the quarterly mean PM_{10} concentration in µg/m^3 in the years before and after LEZ implementation. The gray shaded area indicates the pre-treatment period. The coefficient in the year prior to implementation is normalized to zero. The regression includes county fixed effects, state–quarter fixed effects, LEZ wave–event time fixed effects, LEZ wave–treated fixed effects as well as weather and socio-economic controls. It is weighted by the county–quarter cell size. Standard errors are clustered at the county level. Confidence intervals refer to the 5% level of significance.

lated to asthma specifically, we consider a subset of pharmaceuticals that is closely linked to the therapy of this chronic disease. For both groups of prescriptions, we provide reduced form estimates that indicate the health effect of LEZ implementation (upper panel) and IV estimates representing the health effect of a one µg/m^3 increase in PM_{10} levels (lower panel) in Table 4.5.2. The dependent variable is either the number of prescriptions (left side) or their costs in € (right side) that accumulate on average over the first five years of a child's life.

All regression estimates show that LEZs benefit child health. The magnitude of our results is robust and statistically significant across control sets. For the broad group of respiratory diseases the most stringent IV estimate in column (2), for instance, shows that an LEZ-caused decrease in pollution of one µg/m^3 *in utero* and in the first year of life reduces the number of medical prescriptions by about 0.55 per child, on average. With a standard error of 0.17, the estimate is statistically significant at the 1% level. Relating the treatment effect to the pre-treatment average of 14.14 prescriptions per child reveals a relative reduction of 3.9%. Likewise, the coefficient in column (4) shows that the costs of prescriptions decrease by 10.54 € (4.8%) over the first five years of a child's

TABLE 4.5.2: The Effect of Early-Life PM_{10} Exposure on Medication of Respiratory Diseases throughout Early Childhood

	A. Number of prescriptions		B. Costs of prescriptions (€)	
	(1)	*(2)*	*(3)*	*(4)*
	Reduced Form Estimation			
Respiratory diseases	-0.65	-0.75	-10.98	-14.41
	(0.17)	(0.18)	(3.58)	(4.32)
Mean outcome	14.14	14.14	218.62	218.62
Asthma	-0.21	-0.23	-6.34	-8.01
	(0.07)	(0.08)	(2.57)	(3.22)
Mean outcome	2.5	2.5	73.27	73.27
	IV Estimation			
Respiratory diseases	0.50	0.55	8.46	10.54
	(0.16)	(0.17)	(3.36)	(3.97)
Mean outcome	14.14	14.14	218.62	218.62
Asthma	0.16	0.17	4.88	5.86
	(0.06)	(0.07)	(2.19)	(2.74)
Mean outcome	2.5	2.5	73.27	73.27
First stage F-statistic	14.25	20.14	14.25	20.14
Weather controls	x	x	x	x
Socio-economic controls		x		x

Notes: This table reports reduced form estimates that indicate the health effect of LEZ implementation (upper panel) and IV estimates from Equation (4.2) representing the health effect of a one $\mu g/m^3$ increase in PM_{10} levels during the *in utero* period and the first life year (lower panel). The dependent variable is either the number of prescriptions per child (left side) or their costs in € per child (right side) that accumulate over the first five years of a child's life on average. It refers to either prescriptions for respiratory diseases in general or asthma specifically. The dependent variable is composition-adjusted for the birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level.

life on average. The relative reductions are even higher in magnitude when considering the subset of prescriptions for asthma. While the number of prescriptions decreases by about 6.8%, the costs decrease by about 8.0% on average. Moreover, comparing the estimated coefficients for respiratory diseases in general and asthma specifically, we find that about 56% (5.86/10.54) of the cost savings accrue due to changes in chronic asthma diseases, while about 31% (0.17/0.55) of the reduction in the number of prescriptions is attributable to asthma. AR confidence intervals (CIs) for the IV-estimates reported in Table 4.A.2 of the Appendix corroborate that all average effects are positive and significantly different from zero.

Based on the reported results we approximate the total cost savings from LEZ protection during the *in utero* period and the first year of life. The most recent birth statistics tell us that 1,836,434 children are protected by LEZs *in utero* and in the 12 months after birth. Multiplying this number with 14.41 € in cost savings per child, the first reduced form coefficient in column (4), we find that treatment reduces long-run pharmaceutical costs by approximately 26.5 million € in children born until 2017.[21] Similarly, the calculation linked to asthma medication highlights cost savings of about 14.7 million €.

These specific savings represent an important component for a cost-benefit analysis of the policy but only a fraction of the total health benefits that need to be included. In particular, our study considers only new-born children rather than the whole population, only physical rather than behavioral effects, and only prescriptions rather than doctor visits, hospitalizations, and fatalities. Moreover, the estimated savings derive exclusively from early-exposure effects measured until school enrollment. Benefits that persist over half a decade, however, are unlikely to cease suddenly with enrollment. Nonetheless, the estimated 26.5 million € already account for about 22% of the up-front costs of owners of vehicles that fail to meet LEZ standards calculated by Rohlf et al. (2020). In combination with other empirically identified LEZ-benefits, namely reduced hospitalizations (Pestel and Wozny, 2019), ambulatory care claims (Margaryan, 2021), and prescriptions (Rohlf et al., 2020) in the general population, it is realistic to conclude that LEZ-costs can be recovered within a few years. Our IV estimates may also serve to approximate the counterfactual situation of implementing the WHO's PM_{10} guideline of 20 µg/m^3 already in 2008. Holding constant health benefits at 10.54 € per child per µg/m^3 and assuming all children born into LEZ-counties until 2017 had benefited, the 6.4 µg/m^3 PM_{10} reduction from the annual mean of 26.4 µg/m^3 could potentially have reduced longer-run pharmaceutical costs by about 6.4×10.54 € $\times$ $2,301,305$ children ≈ 155 million €.

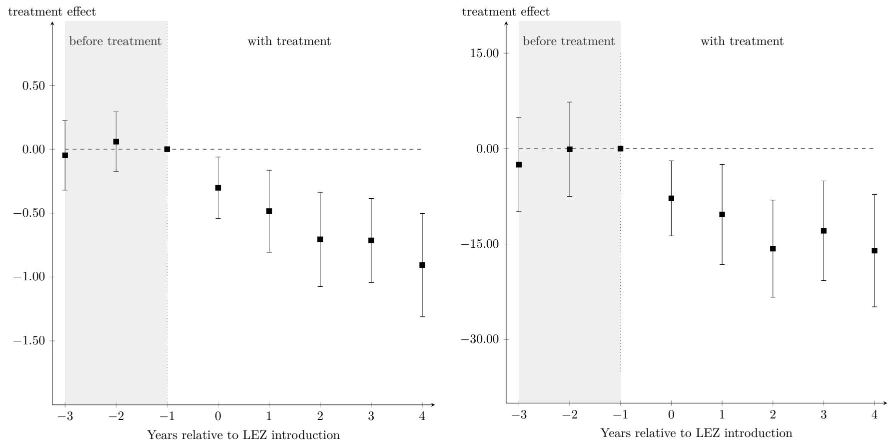
Figure 4.5.2 plots the event-study results for the number and the costs of prescriptions for respiratory diseases.[22] They allow us to examine how LEZs affect long-run health depending on a child's age at the time of exposure. Because the year prior to

[21]This back-of-the envelope calculation does not account for potential differences between those insured with AOK and the general population and is based on the assumption that children born after 2012 benefit equally from the policy.

[22]Table 4.A.1 in the Appendix provides the full regression results.

FIGURE 4.5.2: Event-study Estimates of LEZ Effects on Medication for Respiratory Diseases

(A) Number of prescriptions (B) Costs of prescriptions



Notes: The figure presents event-study coefficients based on the specification in Equation (4.3) that show how LEZs affect the medication of respiratory diseases depending on the time between birth and LEZ implementation in years. The dependent variable is the number of prescriptions per child (Panel A) or their costs in € per child (Panel B) that accumulate over the first five years of a child's life on average. The gray shaded area indicates the pre-treatment period. The coefficient in the year prior to implementation is normalized to zero. The regression includes county fixed effects, state–quarter fixed effects, LEZ wave–event time fixed effects, LEZ wave–treated fixed effects as well as weather and socio-economic controls. It is weighted by the county–quarter cell size. Standard errors are clustered at the county level. Confidence intervals refer to the 5% level of significance.

LEZ implementation is the reference category, we essentially test for differential effects of exposure relative to exposure at age one and older. Therefore, the significant post-treatment decrease in the number and the costs of prescriptions suggests additional benefits of exposure to cleaner air between conception and age one relative to exposure at age one and later. Also note that children born two (three) years prior to implementation are treated no later than from age two (three) onward. Therefore, significant pre-treatment coefficients could indicate differential benefits from exposure beginning at age two or three relative to exposure at age one. However, they could also indicate that the common trends assumption is violated. The fact that pre-treatment coefficients are close to zero, suggests that neither of the two scenarios apply. The event-study plots for asthma prescriptions are similar to the ones for respiratory diseases in general (see Figure 4.B.2 in the Appendix).

Disaggregation by year of life Using outcome measures that aggregate medication over the first five years of life neither reveal whether health effects are persistent nor whether their intensity is constant over time. To gain further insights in how early pollution exposure propagates through early life, we analyze the pharmaceuticals prescribed in each of the five years separately in Panel A of Table 4.5.3. Because numbers and costs of prescriptions have shown to exhibit very similar behavior with respect to LEZ implementation, we focus on the number of prescriptions from here on.

With regards to respiratory diseases in general, the reduced form and the IV estimations consistently reveal that early pollution exposure persistently and statistically significantly affects medication. For instance, the IV estimates indicate that a one μg/m^3 decrease in PM_{10} exposure *in utero* and in the first year of life decreases the average number of prescriptions per child in each of the five pre-school years. However, the coefficient in year five is statistically significant only at the 10% level. Contemporaneous reductions in medication needs in the first year of life account for only about 20% of the cumulative savings on prescriptions over the first five years of life (0.11/0.55). With regards to asthma related prescriptions, this share reduces to 6% (0.01/0.17). Moreover, the contemporaneous effect on asthma medication remains statistically insignificant and effects occur only from the second year on. This suggests that it requires time for improvements in chronic diseases to materialize. Note that the sum of the coefficients estimated for each individual pre-school year is identical to the cumulative effect over all five years presented previously and presented again in column (1) of Table 4.5.3.

Disentangling the extensive from the intensive margin We assess how the extensive and the intensive margin drive the overall treatment effect in Panel B and C of Table 4.5.3. The dependent variable in Panel B is the share of children in the cohort that require at least one prescription for either a respiratory disease or asthma specifically (i.e. extensive margin). The dependent variable in Panel C is the average number of prescriptions a child diagnosed with any respiratory disease or asthma specifically requires (i.e. intensive margin).

For respiratory diseases in general, our IV estimate shows that a one μg/m^3 decrease in PM_{10} pollution *in utero* and in the first year of life reduces the share of sufferers by 1 percentage point (column 7). This finding is statistically significant at the

5% level. In addition, the lower pollution level decreases the number of prescriptions in children diagnosed with any kind of respiratory disease on average by about 0.42 over the five pre-school years (column 13). For asthma specifically, we find a 1 percentage point reduction in the share of children newly diagnosed with asthma (column 7). The prescription requirements of children suffering from asthma, however, are not affected in a statistically significant way (column 13), which may reflect that existing chronic diseases need constant treatment. Because the medication of respiratory diseases comprises that of asthma and because we find clear treatment effects at the intensive margin for respiratory diseases in general, our findings suggest that lower PM_{10} levels have positive effects on children's respiratory health beyond asthma. While we cannot positively identify these effects, we are able to make tentative inferences. For example, Beatty and Shimshack (2014) find that air pollution affects upper respiratory infections such as sinusitis and lower respiratory infections such as acute bronchitis or acute bronchiolitis.

Note that the estimated coefficients for the extensive and intensive margin confirm the overall effect presented in column (1) of Table 4.5.3. For instance, for asthma the overall effect given by the IV-coefficient in column (1) is approximated by the sum of the product of the extensive margin coefficient in column (7) and the pre-treatment mean at the intensive margin in column (13) and the product of the intensive margin coefficient in column (13) and the pre-treatment mean at the extensive margin in column (7) ($0.01\times11.32+0.24\times0.22 \approx 0.17$).[23] Moreover, analyzing the health effects at the extensive and intensive margin for each year separately, we mostly observe initially latent effects that become prominent only after the first year of life. This finding once more suggests that the health stock adjusts slowly. Overall, our results are robust to controlling the false discovery rate following Benjamini and Hochberg (1995) for the 76 hypotheses we test (Table 4.A.11 in the Appendix).

Because the degree of suffering may vary substantially among children afflicted by respiratory diseases, we deepen our analysis of heterogeneous treatment effects at the intensive margin by applying unconditional quantile regressions (Firpo et al., 2009) at the level of the individual child. For computational tractability, the estimation is based

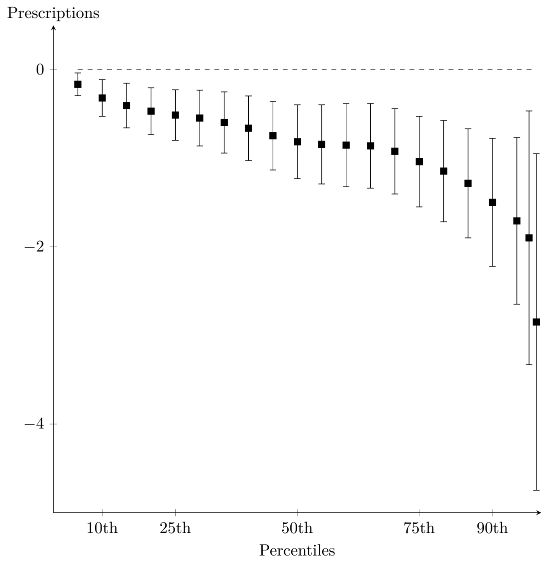
[23]The formal relation is given as $\frac{d(P/N)}{d(x)} = \frac{d(P/S)}{d(x)} \times \frac{S}{N} + \frac{d(S/N)}{d(x)} \times \frac{P}{S}$, where P is the number of prescriptions per child over the five pre-school years, N is the total number of children in the cohort and S is the share of children in the cohort suffering from the disease.

TABLE 4.5.3: The Effect of Early-Life Exposure to PM_{10} on Medication of Respiratory Diseases by Year of Life and at the Extensive and the Intensive Margin

	A. Total Number of prescriptions per child						B. Extensive Margin Share of sufferers						C. Intensive Margin Number of prescriptions per sufferer					
	(1) year 1-5	*(2)* year 1	*(3)* year 2	*(4)* year 3	*(5)* year 4	*(6)* year 5	*(7)* year 1-5	*(8)* year 1	*(9)* year 2	*(10)* year 3	*(11)* year 4	*(12)* year 5	*(13)* year 1-5	*(14)* year 1	*(15)* year 2	*(16)* year 3	*(17)* year 4	*(18)* year 5
								Reduced Form Estimation										
Resp. diseases	-0.75	-0.15	-0.23	-0.20	-0.10	-0.08	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	-0.58	-0.08	-0.18	-0.18	-0.07	-0.07
	(0.18)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.16)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
Mean outcome	14.14	2.17	3.34	2.94	3.08	2.60	0.79	0.71	0.84	0.81	0.82	0.77	17.69	3.03	3.97	3.63	3.73	3.33
Asthma	-0.23	-0.02	-0.08	-0.06	-0.04	-0.02	-0.01	0.00	-0.02	-0.02	-0.01	-0.01	-0.33	-0.07	-0.14	-0.06	-0.04	-0.02
	(0.08)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.28)	(0.06)	(0.08)	(0.07)	(0.08)	(0.09)
Mean outcome	2.50	0.50	0.67	0.50	0.46	0.38	0.22	0.24	0.28	0.22	0.20	0.16	11.32	2.05	2.38	2.26	2.31	2.33
								IV Estimation										
Resp. diseases	0.55	0.11	0.17	0.14	0.07	0.06	0.01	0.02	0.02	0.01	0.01	0.00	0.42	0.06	0.13	0.13	0.05	0.05
	(0.17)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.15)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Mean outcome	14.14	2.17	3.34	2.94	3.08	2.60	0.79	0.71	0.84	0.81	0.82	0.77	17.69	3.03	3.97	3.63	3.73	3.33
Asthma	0.17	0.01	0.06	0.04	0.03	0.02	0.01	0.00	0.02	0.02	0.01	0.01	0.24	0.05	0.10	0.04	0.03	0.02
	(0.07)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.21)	(0.05)	(0.06)	(0.05)	(0.06)	(0.07)
Mean outcome	2.50	0.50	0.67	0.50	0.46	0.38	0.22	0.24	0.28	0.22	0.20	0.16	11.32	2.05	2.38	2.26	2.31	2.33
Weather controls	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
SE controls	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Notes: This table reports reduced form estimates (upper panel) and IV estimates (lower panel) for disaggregated health effect in each of the five pre-school years and at the extensive and the intensive margin. The dependent variable is either the number of prescriptions per child (Panel A), the share of children in the cohort that requires at least one prescription (Panel B), or the number of prescriptions per child diagnosed with a disease (Panel C). It either refers to prescriptions for respiratory diseases in general or asthma specifically and it is either aggregated over the first five years of a child's life or given for each year separately. The dependent variable is composition-adjusted for the birth county–birth quarter cell in panel A. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects as well as weather and socio-economic controls. The regressions are weighted by birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 9,609.

FIGURE 4.5.3: Unconditional Quantile Treatment Effects of LEZs on Medication for Respiratory Diseases



Notes: The figures present coefficients from unconditional quantile regressions (Firpo et al., 2009) at the level of the individual child (Appendix 4.E for further information). The dependent variable is the number of prescriptions for respiratory diseases that accumulate over the first five years of a child's life. The severity of suffering increases from left to right. The bars indicate the 95% confidence interval.

on a reduced-form standard DID.[24] Our estimates in Figure 4.5.3 provide suggestive evidence that children who suffer worst from respiratory diseases may benefit the most from LEZs. For example, with 2.8 fewer prescriptions, point estimates for children in the 99th percentile indicate that they may benefit nearly fives times as much from LEZs as the average sufferer with 0.58 fewer prescriptions (Table 4.5.3).[25]

Comparison to the literature To provide context for the magnitude of our findings, we compare our estimates to related epidemiological and economic research that focuses on asthma.

In a meta-study, Khreis et al. (2017) summarize the available epidemiological research on the impact of early life exposure to air pollution on the prevalence of asthma in children. Overall, the research suggests odds ratios of 1.025 for associations between

[24]For further information on the unconditional quantile estimator see Appendix 4.E.

[25]Table 4.A.3 in the Appendix features all coefficients and standard errors.

PM_{10} and asthma at any age. Taking the odds ratio as an approximation of relative risk, we can compare the magnitude of our estimates for the share of sufferers to these results. Our estimate for asthma in column (7) in Table 4.5.3 implies a risk ratio of 1.045 at the mean, which is outside the meta-study's 95% confidence interval for studies that consider children from age three to young adults up to age 21. However, odds ratios from analyses limited to the ages three to six come very close to our estimates (Clark et al. (2010): 1.068, Deng et al. (2016): 1.048, Liu et al. (2016): 1.029).

Economic studies focus on contemporaneous improvements in child health. Using the case of the Stockholm congestion charge, Simeonova et al. (2019) show that persistently lower PM_{10} exposure reduces asthma-related hospital admissions of children below six years of age with an implied elasticity of 3.7. The elasticities we estimate for asthma drug prescriptions (1.8), expenditures (2.1), and the share of sufferers (1.2) are smaller.[26] This difference could be attributed to the fact that Simeonova et al. (2019) examine contemporaneous benefits of persistently improved air quality over a longer time period, while we study longer run health benefits from exposure to cleaner air in a single year.

Other economic studies consider short-run variations in air pollution exposure, but mainly focus on $PM_{2.5}$.[27] Alexander and Schwandt (2019) study the impact of emissions cheating by car manufacturers on $PM_{2.5}$ and child health outcomes. Their estimates imply that a one $\mu g/m^3$ increase in $PM_{2.5}$ increases asthma related hospital admissions of children aged four and younger by 0.42 per 1,000. Evaluated at the reported means, the elasticity is 3.01. Barwick et al. (2018) study changes in health-related consumption in China for $PM_{2.5}$ using data on bank card transactions. They estimate that a 10 $\mu g/m^3$ decrease in $PM_{2.5}$ reduces health spending in children's hospitals by 1.13%, implying an elasticity of 0.06.

[26] The calculations are based on the IV estimates for asthma in columns (2) and (4) of Table 4.5.2 and in column (7) of Table 4.5.3, respectively. These point estimates are then multiplied with the mean PM_{10} exposure and divided by the mean of the outcome to obtain elasticities.

[27] Beatty and Shimshack (2014) is a notable exemption. Based on data from young children in England, they relate respiratory treatments for children to monthly PM_{10} exposure. The estimated coefficient on PM_{10} is, however, statistically insignificant but would imply an elasticity of only 0.1.

TABLE 4.5.4: Severe Effects of Exposure to PM_{10} During the Perinatal Period on Fetal Development

	IV Estimation Hospital treatments per 1,000 children		
	developmental disorder and malnutrition *(1)*	short gestation period and low birth weight *(2)*	long gestation period and high birth weight *(3)*
Mean PM_{10} ($\mu g/m^3$)	-0.03	1.50	0.12
	(0.61)	(2.49)	(0.39)
Mean outcome	2.93	50.22	1.14
Weather controls	x	x	x
Socio-economic controls	x	x	x

Notes: This table reports coefficient estimates from three variants of the IV regression in Equation (4.2). The dependent variable is the number of stationary hospital treatments of three different disorders linked to abnormal fetal development per 1,000 children. It is the composition-adjusted average for a birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The first stage F-statistic is 24.39. The sample size is 9,609.

4.5.3 Common infant health measures

Health effects may be subtle if changes in pollution exposure are small and health effects may be latent if the health stock only adjusts slowly. Our results indicate that this could be the case in the context of LEZ implementation. To assess this hypothesis further, we estimate the effect of a one $\mu g/m^3$ decrease in PM_{10} exposure during the prenatal period on fetal development. The outcome variables are postpartum stationary hospitalization due to abnormal birth weight, unusual pregnancy duration, and fetal malnutrition.[28] The estimated treatment effects in Table 4.5.4 remain statistically insignificant. Given our findings for respiratory health, this indicates that infant health measures revolving around hospitalizations immediately postpartum may be too coarse or too focused on the short-term to detect health effects. Moreover, it might explain why prior studies on LEZs could not provide evidence for improved infant health measured in the form of low birth weight (Gehrsitz, 2017; Pestel and Wozny, 2019).

[28] As before, the sample comprises three cohorts born prior to and all cohorts born up to five years subsequent to LEZ implementation. However, we exclude the first three cohorts born after implementation which are only partially protected by LEZs *in utero*.

4.5.4 Effect mechanisms

Different mechanisms may explain the effects of LEZs on pollution and health. First, overall traffic could have decreased because of mode switching to public transport. Second, individuals could have substituted their banned vehicles with "greener" vehicles. To assess these two potential channels we draw on additional annual data at the county level provided by the Federal Highway Research Agency (BAST) and the German Federal Motor Transport Authority (KBA). The BAST data provides information on the number of passing vehicles on all freeways (*Autobahnen*) and federal roads (*Bundesstraßen*) recorded by traffic monitors. The KBA data provides information on the number of registered private and commercial passenger vehicles by fuel type (diesel and gasoline) and emission class.

Table 4.5.5 shows that LEZ implementation has no statistically significant effect on the traffic volume on an average day (column 1).[29] However, it reveals a shift in the composition of the vehicle fleet. As expected, we observe a significant 20% decrease ($e^{0.18} - 1$) in the number of old diesel vehicles classified as Euro1 and lower that are banned by the LEZs (column 2). Even vehicles with emission classes Euro2 through Euro4 decrease significantly (column 3). The banned vehicles seem to be replaced by used gasoline vehicles with the emission standards Euro2 through Euro4 (column 6). We find no significant changes in the number of the newest diesel and gasoline cars (column 4 and 6).

In combination, these findings provide tentative evidence on the emission source of the policy-induced PM_{10}-reduction. Because overall traffic is not affected, particles from wear and tear of brakes, tires, and road surfaces are unlikely to have changed upon LEZ implementation. The "de-dieselization" of the vehicle fleet, however, suggests that after implementation the same amount of traffic is caused by fewer dirty diesel vehicles and more cleaner gasoline vehicles.[30] Thus, diesel exhaust may be the main driver of the observed PM_{10}-reductions. Toxicological and epidemiological studies suggest that PM from diesel exhaust is particularly harmful because it mainly con-

[29] Given that LEZs usually cover city centers, it is possible that traffic from within the LEZ shifts to other, unregulated areas of the county. While the data from BAST do not allow a closer examination, Wolff (2014) and Gehrsitz (2017) provide evidence against this hypothesis.

[30] Diesel vehicles with Euro1 may emit up to 140 µg tailpipe PM_{10} per km. Gasoline vehicles with Euro2-4 emit about 1-2 µg PM_{10} per km according to HBEFA.

TABLE 4.5.5: The Effect of LEZ Implementation on Traffic Volume and the Vehicle Fleet

	Traffic volume	Diesel vehicles			Gasoline vehicles		
	vehicles/24 hours	Euro1 and worse	Euro2-4	Euro5 and better	Euro1	Euro2-4	Euro5 and better
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
LEZ treatment	-0.01	-0.18	-0.03	-0.01	0.04	0.02	-0.01
	(0.02)	(0.02)	(0.01)	(0.05)	(0.06)	(0.01)	(0.15)
Sample size	5,584	6,514	6,514	5,418	6,514	6,514	5,418
Weather controls	x	x	x	x	x	x	x
Socio-economic controls	x	x	x	x	x	x	x

Notes: The dependent variable is the number of motor vehicles counted over the 24 hours of an average day in column (1), the number of vehicles with a diesel engine in the emission class Euro1 or worse in column (2), in emission class Euro2 through Euro4 in column (3), and in emission class Euro5 or better in column (4), and the number of vehicles with a gasoline engine in emission class Euro1 in column (5), in emission class Euro2 through Euro4 in column (6), and in emission class Euro5 or better in column (7). The sample in column (4) and (7) is limited to the period after 2008 because the emission class Euro5 was only introduced in 2009. All outcome variables are transformed with the inverse hyperbolic sine function. Accordingly, the percentage change in the outcome variable is given by $(e^{\beta} - 1) \cdot 100$. All regressions include county, state–year, LEZ wave–event time, LEZ wave–treated fixed effects, as well as weather and socio-economic controls. The weather variables comprise only precipitation, temperature, and sunshine duration because they are available for the entire period in which we observe the BAST and KBA data. The regressions are weighted by the average cohort size. Standard errors in parentheses are clustered at the county level.

sists of small particles that can penetrate far into the human body (Krzyżanowski et al., 2005; HEI, 2010).

4.5.5 Robustness checks

Alternative control groups Changes in the composition of the control group do not alter our results. Our baseline estimates rely on a control group that comprises counties that implement LEZs at some point and counties that violate EU PM_{10} standards but have no LEZ to date. Neither of the event-study plots in Figure 4.5.1 or 4.5.2 point to differential trends in air pollution and health outcomes across the treatment and control groups before policy introduction. This alleviates concerns that LEZ adoptions are determined by any omitted local conditions or shocks that simultaneously affect air pollution and health outcomes. Nonetheless, we additionally restrict our sample to counties that actually introduce LEZs to ensure that our results are not determined by deliberate selection of counties into treatment. Recall that our full sample comprises 128 counties of which 65 actually implement LEZs; 49 between 2008 and 2012. Figure 4.B.3 in the Appendix shows event-study estimates based on the sample of the 65 ever-adopter counties. Although the restriction to ever-adopters nearly slashes our sample in half, the event-study plots for the first stage and the reduced form continue showing well-behaved patterns that are similar to those in Figure 4.5.1 and 4.5.2.

In our preferred DID setting, we observe cohorts born up to five years after treatment and exclude all cohorts from counties that implement LEZs in this time period from the control group in each of the stacked data sets. By choosing a five-year time window after treatment, we are able to identify longer-run effects of LEZ introduction. However, the length of the time window has implications for the composition of the control group. The shorter the window, the more comparable are the control and the treatment group. This is because more control units that introduce an LEZ with close proximity in time are eligible to the control group. In Table 4.A.4 in the Appendix we show that treatment effects remain robust when shortening the time window after treatment sequentially from five to two years.

Spillover effects Some of the counties in our control group directly neighbor counties implementing an LEZ. These counties may be subject to positive or negative spillovers. This would bias our estimated effects on pollution and health. In particular, we are concerned that drivers may change their routes to circumnavigate the LEZ such that traffic would merely be displaced. In this case, LEZs would have worsened air pollution in neighboring counties and our estimates would overestimate the policy‘s effectiveness. To address this concern, Table 4.5.6 replicates the estimations from Tables 4.5.1 and 4.5.2. The only difference is that we include a binary variable that takes on a value of 1 if a neighboring county implements an LEZ.[31]

We find negative but statistically insignificant policy effects in neighboring counties with regard to either PM_{10} pollution, prescriptions for respiratory diseases, or asthma. In line with these findings, Wolff (2014) and Gehrsitz (2017) show that treatment effects on pollution measuring stations outside of the LEZs are negative but insignificant. However, when controlling for neighboring counties, the effect of LEZs on air pollution within LEZ counties is -2.18 (Table 4.5.6) while our main estimate is -1.37 (Table 4.5.1). The reduced form effects of LEZs on the number and costs of prescriptions are also higher when controlling for neighboring counties compared to

[31]In alternative tests, we exclude all neighboring counties from each treatment wave in the sample and account for the number of neighboring LEZs when estimating spillover effects. In either case, the results remain almost identical.

our preferred estimates in Table 4.5.2, the only exception being the coefficient for the number of prescriptions for asthma. However, the IV effects remain almost unchanged given that the difference in the estimated effects on pollution and health are of similar proportion.

While we cannot recover statistically significant effects on neighboring counties, the higher magnitude of the reduced form treatment effects when controlling for neighboring counties indicates that positive spillover effects are likely present. This conjecture is corroborated when testing whether the policy-induced "de-dieselization" of the vehicle fleet identified in Section 4.5.4 expands across county borders. In line with Wolff (2014), we find that neighboring counties also exhibit a statistically significant reduction (-0.04, $t = -2.06$) in the number of banned diesel cars classified as Euro1 and lower. This effect can be plausibly linked to the fact that LEZs cover city centers where many workplaces and points of interest are located. Individuals living in the neighboring counties have an incentive to ensure that their vehicles allow access.

Overall, the findings alleviate concerns about overestimating and instead mark our main analysis as conservative. In fact, when accounting for positive spillover effects the estimated effectiveness of the policy could be considerably greater. While in Section 4.5.2 we estimate that treatment reduces long-run pharmaceutical costs by approximately 26.5 million € in children born until 2017, the estimates in Table 4.5.6 suggest cost savings of 42.4 million € ($1,836,434$ children $\times$ 23.10 €).

Accounting for changes in population characteristics A potential threat to identification in our IV design is that improved air quality might change the composition of the cohorts in LEZ counties, leading to changes in the unobservable characteristics of the children born there. This occurs if the implementation of LEZs changes how individuals move in and out of the county which would invalidate the exclusion restriction. For example, if LEZs either attract good health risks or induce bad health risks to locate to other counties, improvements in child health may result from the changed socio-economic structure of counties rather than from reduced pollution. While there exists no evidence in the literature that LEZs affect the socio-economic structure of counties, we nevertheless report a range of additional estimates that strengthen our confidence that the exclusion restriction holds.

Table 4.5.6: Air Pollution and Health Effects on Neighboring Counties

	PM_{10} Pollution	Number of prescriptions		Costs of prescriptions	
	$\mu g/m^3$	Respiratory diseases	Asthma	Respiratory diseases	Asthma
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
			A. Reduced Form Estimation		
LEZ effect on LEZ-counties	-2.18	-1.09	-0.24	-23.10	-14.34
	(0.53)	(0.28)	(0.12)	(6.38)	(5.39)
LEZ effect on neighbor-counties	-0.41	-0.16	-0.03	-1.56	-0.54
	(0.25)	(0.13)	(0.07)	(3.56)	(2.83)
			B. IV Estimation		
PM_{10} Pollution	-	0.49	0.11	10.34	6.38
		(0.15)	(0.06)	(3.47)	(2.73)
Weather controls	x	x	x	x	x
Socio-economic controls	x	x	x	x	x

Notes: This table replicates the first stage regression in Table 4.5.1 and the reduced form and IV regressions in Table 4.5.2. In addition, the treatment effect on counties adjacent to those that implement an LEZ is estimated. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects, as well as weather and socio-economic controls. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 7,665.

First, we analyze whether LEZ implementation significantly affects how individuals migrate in and out of counties (Appendix Table 4.A.5) and we conduct placebo tests (Appendix Table 4.A.6). To estimate treatment effects on mobility patterns we consider net migration in the overall county population, net migration among families, and the fraction of AOK-insured children moving out of their birth county. As placebo health outcomes we use hospital treatments of injuries of the head, the arm, and of other body parts. All outcomes are typical health issues in children and strongly correlated with socio-economic status (e.g. Faelker et al., 2000; Birken and MacArthur, 2004; Yates et al., 2006), while the air pollution literature does not indicate a relationship. Table 4.A.5 and Table 4.A.6 in the Appendix do not reveal any statistically significant policy effects on migration patterns and placebo outcomes. However, because the coefficients are estimated imprecisely, we turn to another indirect test on changes in the socio-economic composition of cohorts.

If unobservable population characteristics adjust gradually over time, comparing treated and untreated cohorts will suffer less bias the closer they are in the time dimension. Therefore, we reuse our previous robustness analysis in which we sequentially reduced the five-year time window after treatment to two years. It is unlikely that population characteristics change markedly within just two years. In the absence of

TABLE 4.5.7: The Effect of LEZ Implementation on Different Pollutants

	(1) $PM_{2.5}$	(2) NO_2	(3) O_3	(4) SO_2
LEZ treatment	-0.30	-3.27	-0.04	-0.12
	(0.12)	(1.07)	(0.56)	(0.25)
Mean outcome	15.51	40.81	39.27	4.3
First stage F-statistic	5.79	9.44	0.01	0.24
Weather controls	x	x	x	x
Socio-economic controls	x	x	x	x

Notes: This table reports coefficient estimates for the effect of LEZs on four different air pollutants. The dependent variable is either the mean concentration of $PM_{2.5}$, NO_2, O_3, or SO_2 in μg/m^3. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects as well as weather and socio-economic controls. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 3,466 in column (1) and 9,609 in columns (2) through (4).

gradual changes in unobserved characteristics, the estimates for the limited samples should be close to those in the main analysis. Table 4.A.4 in the Appendix shows that treatment effects remain indeed robust as the post-treatment time window decreases.

Effect on other air pollutants We use PM_{10} as a measure for PM exposure. By definition, PM_{10} includes particles below 10 μm, including the finer $PM_{2.5}$ particles. We rely on PM_{10} because policymakers in Europe are highly focused on this pollutant, and LEZs explicitly target PM_{10}. Moreover, the EU only set legally binding limits for $PM_{2.5}$ in 2015. To evaluate whether LEZs also decrease $PM_{2.5}$, in a first robustness check, we resort to satellite-based $PM_{2.5}$ estimates from van Donkelaar et al. (2019).[32] This data is available on a fine resolution grid (0.01 degrees) but only at the annual level.[33] Thus, we lose quarterly observations and the corresponding fixed effects. Table 4.5.7 shows that the introduction of LEZs reduces mean $PM_{2.5}$ concentrations by about 2%. The relatively modest magnitude compared to that for PM_{10} may reflect attenuation bias from non-classical measurement error.

[32]Data from the German air monitoring network for $PM_{2.5}$ is very limited. We have about 70% fewer observations for $PM_{2.5}$ than for PM_{10}.

[33]van Donkelaar et al. (2019) merge satellite measurements of aerosol optical depth with a particulate transport model, and combine them with data from air monitoring stations to obtain estimates of $PM_{2.5}$ for Europe.

Our second robustness check is motivated by the fact that diesel vehicles emit significant quantities of nitrogen oxides. In fact, road traffic emissions of nitrogen dioxide (NO_2), which serve as an indicator for different nitrogen oxides, are caused primarily by diesel vehicles.[34] Therefore, we assess whether LEZ implementation also leads to notable changes in ambient NO_2 concentrations using our data from the German air monitoring network. Table 4.5.7 shows that LEZs significantly reduce NO_2 by about 3.27 µg/m^3 (8.0%) on average. This finding is consistent with the fact that LEZs are *de facto* bans of old diesel vehicles.

Because LEZs reduce both PM and NO_2, we cannot conclusively infer that PM_{10} determines our observed health effects alone. Therefore, we caution against interpreting our IV results as a causal estimate of the health effects of PM_{10} as a stand-alone pollutant. Instead, we argue that our IV results represent health effects linked to air pollutants from diesel vehicles. Similarly, other papers generalize their results to air pollution effects (c.p. Chay and Greenstone, 2003a; Currie and Neidell, 2005; Arceo et al., 2016; Knittel et al., 2016; Deryugina et al., 2019; Colmer et al., 2020). We subsequently show that the policy does not affect other pollutants.

First, we examine whether LEZs have unintended effects on ozone (O_3) concentrations. O_3 is negatively correlated with other local air pollutants, in particular with NO_2 which is one of its precursors. Therefore, we may be concerned that the implementation of LEZs increases O_3 concentrations. However, Table 4.5.7 does not provide evidence for an unintended increase in O_3.

Second, environmental regulation can have adverse impacts on firm output and productivity. Therefore, we may be concerned that LEZs decrease industrial activity and, thereby, reduce emissions of industrial pollutants, most notably SO_2. To rule out that health effects are subject to this channel, we also estimate effects for SO_2 concentrations. Table 4.5.7 does not reveal any statistically significant effects. Because transport only accounts for about 2% of total SO_2 emissions, this robustness check also serves as a placebo test, suggesting that our first-stage results are not driven by confounding factors.

[34]In Germany, about 72.5% of NO_2 emissions from on-road traffic are from diesel vehicles (UBA, 2017).

Accounting for treatment differences after the first year of life In our main analysis, we compare children who experience different pollution exposure levels *in utero* and over their first year of life but the same exposure levels afterwards. A strict implementation of this comparison requires that we restrict ourselves to cohorts born exactly four quarters prior to LEZ implementation. However, to avoid limiting ourselves to a single pre-treatment observation of the treated, we additionally include the two cohorts born five and six quarters prior to treatment in our main analysis. The drawback of this approach is that it neglects potentially different exposure levels in the second year of the children's lives. However, our event-study estimates in Figure 4.5.2 do not indicate additional benefits from exposure at age one relative to exposure at age two and three. As a further robustness check, we limit pre-treatment observations to cohorts born four quarters before implementation. The results in Table 4.A.7 in the Appendix show that our findings are robust with respect to this sample restriction. The reduced form and the IV point estimates are very similar in magnitude to the ones reported in Table 4.5.2.

Accounting for increases in policy stringency Our treatment estimate identifies the effect of LEZ introduction. Upon implementation, LEZs ban the most emission-intensive diesel vehicles with tailpipe emission category Euro1 or lower (no sticker). However, LEZs become more stringent over time so that they eventually also ban vehicles with Euro2 (red sticker) and Euro3 (yellow sticker) standards. This gradual adoption of more stringent restrictions raises the concern that children born after LEZ implementation benefit from cleaner air for longer throughout their pre-school years than children born just before LEZ implementation. This would imply that children in our pre- and post-treatment comparison differ not only in exposure during their *in utero* period and their first year of life. Reassuringly, our event-study results for the number and the costs of prescriptions in Figure 4.5.2 provide no evidence that differences in exposure after age one have any additional health benefits. If there were differential benefits from exposure beginning at age two or three relative to exposure at age one, we would expect positive pre-treatment coefficients. Instead, we observe statistically insignificant coefficient estimates close to zero. As a further robustness check we enrich our event study specification for prescriptions in Equation 4.3 by eight additional binary variables: $\sum_{y=2}^{5} f_{cy}^{Euro2} + \sum_{y=2}^{5} f_{cy}^{Euro3}$, where f_{cy}^{Euro2} (f_{cy}^{Euro3}) is equal to one if Euro2 (Euro3) vehicles are banned in year of life y of cohort c. These dummy variables should absorb potential additional health benefits in year of life two through five linked to the two more stringent LEZ regimes banning Euro2 and Euro3 vehicles, respectively. If there were additional benefits from Euro2 and 3 vehicle bans, we would expect post-treatment coefficients to decrease in event time.[35] However, Figure 4.B.4 in the Appendix exhibits very similar patterns to our event study in Figure 4.5.2 that lacks the additional dummies capturing changes in stringency. This suggests there are no additional benefits from the increasing stringency of LEZs.

Two-way fixed effect DID estimation We also estimate the two-way fixed effect equivalent of our stacked difference-in-differences estimator. The coefficient estimates in Table 4.A.8 show similar, robust effects of PM_{10} on child health. However, they tend to be lower in magnitude. For instance, the IV coefficient for respiratory diseases is 0.55 in the stacked DID estimation in column (2) of Table 4.5.2 while it is only 0.39 in the two-way fixed effect DID estimation in Table 4.A.8. We expect that part of this attenuation stems from the weighted aggregation of heterogeneous treatment effects revealed in Goodman-Bacon (2018). It may also indicate violations in common trends that result from including already-treated units in the control groups for the newly treated, although they are on differential trends from prior treatment. Also note that it is impossible in the two-way fixed effect setup to include fixed effects that absorb implementation wave-specific unobservables in event-time and time-invariant differences between treatment and control groups within and across implementation waves.

Functional form We also test the robustness of our results with respect to the functional form. The outcome variables in our baseline specifications are in levels. Using per capita prescriptions as outcome, we implicitly assume that prescriptions per child would have evolved with the same absolute changes in the absence of treatment. However, if prescriptions per child changed at the same rate in the absence of any LEZ intervention instead, the parallel trends assumption would be violated. Although our event-study plots do not reveal pre-trends that differ in a statistically significant manner, we re-estimate our main results with logged outcome variables in Table 4.A.9 in

[35] We do not expect shifts in post-treatment patterns because the timing of Euro2 and 3 vehicle bans varies across counties relative to the implementation date.

the Appendix. We find that the estimated relative effects are nearly identical to the ones derived from Table 4.5.2.

4.6 Conclusion

This chapter provides a quasi-experimental study that links moderate improvements in air quality in a single year from banning emission-intensive vehicles to substantial health benefits across children's pre-school years. The context of our study are urban counties in Germany, where motor vehicles are a major source of air pollution. Yet, average pre-treatment pollution levels are low. These characteristics are widespread in Europe, so that our results are most likely generalizable. Exploiting unique public health insurance data at the patient level on one million children, we examine whether individuals born just before and just after reductions in PM concentrations caused by the adoption of Low Emission Zones exhibit persistent differences in rarely studied medication use up to five years after treatment. We focus on children's pharmaceutical prescriptions as a sensitive, real-time health measure that overcomes the challenge of capturing health effects that may be both subtle if changes in pollution exposure are moderate and initially latent if the health stock adjusts slowly.

We present strong evidence that the cumulative number and the cumulative costs of pharmaceutical prescriptions over early childhood decrease significantly after LEZ implementations improve air quality. For instance, the number of prescriptions for asthma decreases by 6.8% and their costs decrease by about 8.0% on average for every one $\mu g/m^3$ reduction in PM_{10} concentration. Our findings provide strong support for the notion of health as a stock that changes relatively slowly over time. It is only from the second year of life that the initially latent health response materializes in fewer children diagnosed with asthma and, thus, fewer total prescriptions. Contemporaneous reductions in prescriptions of asthma medication account for less than 6% of the cumulative savings over the first five years of life. This highlights that estimates of contemporaneous pollution impacts may substantially underestimate 'fully formed' health benefits. We identify economically meaningful cost savings for public health insurers. With 1,836,434 children protected by Low Emission Zones *in utero* and during their first year of life, treatment reduces costs for prescriptions in children born between 2008 and 2017 for respiratory diseases by about 26.5 million € over their pre-school years, or 42.4 million € when accounting for positive spillover effects. Because we compare children who differ only in their pollution exposure during their *in utero* period and their first year of life, these cost savings originate from a very short period with slightly improved air quality.

Our results inform contentious policy debates. Across Europe, vintage- and fuel-specific driving bans are a widespread and ever more stringent policy intervention. Major metropolitan areas, including Paris, Madrid, and Rome, are even committed to full diesel bans by 2025. Yet, opponents of driving restrictions prominently question whether a narrow focus on diesel bans is a rational choice to effectively improve air quality and public health. Our study seeks to provide first answers by quantifying an important fraction of the reduction in the public health burden accomplished by LEZs - Germany's flagship policy at the local level to comply with air quality standards set by the EU clean air directives. Finding meaningful health improvements, even at low pre-treatment pollution levels, suggests that vintage-specific driving bans that target particularly old and emission intensive diesels can have large and long-lasting positive effects on children's respiratory health in many settings.

Our study cannot assess whether additional restrictions for newer vehicles would yield further health improvements. This is an important policy question for future research. Another research question is how treatment effects progress through life. In our study, health effects are latent in the first year before materializing. It is far from obvious whether health effects persist permanently and how they impede cognitive or non-cognitive skill formation. To address the progression of effects, we would need to follow the children in our study and rerun our estimations at later points in time, for example at the end of elementary school, at the end of high school, and some years into their professional lives.

Appendix

4.A Additional Tables

TABLE 4.A.1: Event-study Estimates – The Effect of LEZ Implementation on PM_{10} Concentrations and Health Outcomes by Year

	PM_{10} Pollution (µg/m^3) *(1)*	Number of prescriptions *(2)*	Costs of prescriptions *(3)*
LEZ treatment ($\theta = -2$))	0.026	-0.079	-3.094
	(0.411)	(0.137)	(3.659)
LEZ treatment ($\theta = -1$))	0.007	0.034	-0.455
	(0.284)	(0.116)	(3.694)
LEZ treatment ($\theta = 1$))	-0.896	-0.304	-8.006
	(0.254)	(0.122)	(2.969)
LEZ treatment ($\theta = 2$))	-1.269	-0.480	-10.13
	(0.249)	(0.166)	(3.969)
LEZ treatment ($\theta = 3$))	-1.613	-0.674	-14.897
	(0.289)	(0.189)	(3.852)
LEZ treatment ($\theta = 4$))	-2.040	-0.675	-12.34
	(0.387)	(0.17)	(3.967)
LEZ treatment ($\theta = 5$))	-1.929	-0.788	-14.192
	(0.388)	(0.200)	(4.243)
Weather controls	x	x	x
Socio-economic controls	x	x	x

Notes: This table reports estimated event-study coefficients underlying Figure 4.5.1 and Figure 4.5.2. The dependent variables are the average PM_{10} level in µg/m^3, the number or the costs in € of prescriptions that accumulate over the first five years of a child's life on average, respectively. The dependent variables are composition-adjusted for the birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects as well as weather and socio-economic controls. The regressions are weighted by the birth county–birth quarter cell size. They are based on an expanded pre-treatment window including all observations up to three years prior to LEZ implementation. Furthermore, observations from the 3 quarters before and the 3 quarters after LEZ implementation are included. The resulting sample size is 19,290. Standard errors in parentheses are clustered at the county level.

Table 4.A.2: Anderson Rubin Confidence Sets

	A. Number of prescriptions		B. Costs of prescriptions (€)	
	(1)	*(2)*	*(3)*	*(4)*
		Respiratory diseases		
CS_{AR}	[0.25 - 1.03]	[0.28 - 1.04]	[3.01 - 19.31]	[4.07 - 22.07]
F_{AR}	15.11	16.72	9.41	11.10
p-value	0.0001	0.0000	0.0022	0.0009
		Asthma		
CS_{AR}	[0.06 - 0.33]	[0.06 - 0.34]	[1.05 - 11.44]	[1.20 - 13.48]
F_{AR}	8.83	9.04	6.07	6.18
p-value	0.0030	0.0026	0.0138	0.0129
Weather controls	x	x	x	x
Socio-economic controls		x		x

Notes: This table reports weak-instrument-robust inference for the IV-estimates in Table 4.5.2. The AR-confidence sets (CS_{AR}) provide robust confidence intervals with a coverage probability of 95%. The F-distributed AR-statistic (F_{AR}) and its p-value test the null hypothesis that the coefficient of the endogenous variable PM_{10} in the structural equation is equal to zero.

Table 4.A.3: Unconditional Quantile Regression Estimates of the Effect of Early-Life Exposure to LEZs on Respiratory Prescriptions throughout Childhood

	Q-5	Q-10	Q-15	Q-20	Q-25	Q-30	Q-35
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
LEZ treatment	−0.166	−0.310	−0.406	−0.470	−0.514	−0.547	−0.597
	(0.065)	(0.106)	(0.129)	(0.135)	(0.146)	(0.161)	(0.176)
	Q-40	Q-45	Q-50	Q-55	Q-60	Q-65	Q-70
	(8)	*(9)*	*(10)*	*(11)*	*(12)*	*(13)*	*(14)*
LEZ treatment	−0.662	−0.746	−0.814	−0.844	−0.853	−0.861	−0.922
	(0.186)	(0.197)	(0.213)	(0.228)	(0.239)	(0.244)	(0.246)
	Q-75	Q-80	Q-85	Q-90	Q-95	Q-97.5	Q-99
	(15)	*(16)*	*(17)*	*(18)*	*(19)*	*(20)*	*(21)*
LEZ treatment	−1.039	−1.146	−1.284	−1.498	−1.707	−1.899	−2.848
	(0.261)	(0.292)	(0.314)	(0.369)	(0.480)	(0.730)	(0.969)

Notes: This table reports regression coefficients from unconditional quantile regressions. The dependent variable is the number of prescriptions for respiratory diseases that accumulate over the first five years of a child's life. All regressions include birth county and birth state–birth quarter fixed effects. Standard errors in parentheses are clustered at the county level and bootstrapped using 1,000 draws. The sample size is 556, 898.

TABLE 4.A.4: The Effect of Early-Life PM_{10} Exposure – Different Post-Treatment Time Windows

	$\Delta = 2$ *(1)*	$\Delta = 3$ *(2)*	$\Delta = 4$ *(3)*
	A. First Stage Estimation		
Mean PM_{10}	-1.14	-1.53	-1.55
	(0.3)	(0.3)	(0.3)
Mean outcome	26.44	26.44	26.44
	B. Reduced Form Estimation		
Respiratory diseases	-0.44	-0.50	-0.62
	(0.16)	(0.17)	(0.17)
Mean outcome	14.14	14.14	14.14
Asthma	-0.16	-0.21	-0.22
	(0.07)	(0.07)	(0.07)
Mean outcome	2.5	2.5	2.5
	C. IV Estimation		
Respiratory diseases	0.39	0.32	0.40
	(0.16)	(0.13)	(0.12)
Mean outcome	14.14	14.14	14.14
Asthma	0.14	0.13	0.14
	(0.07)	(0.05)	(0.05)
Mean outcome	2.5	2.5	2.5
First stage F-statistic	14.50	25.92	27.19
Sample size	6,922	8,727	9,575
Weather controls	x	x	x
Socio-economic controls	x	x	x

Notes: This table reports coefficient estimates for shorter post-treatment windows. The time window increases sequentially from two years (column 1) to four years (column 3). Panel A presents coefficients from first stage, Panel B the coefficients from reduced form and Panel C coefficients from IV estimations. The dependent variable in Panel A is the PM_{10} concentration; in panel B and C it is the number of prescriptions for respiratory diseases in general or asthma specifically that accumulate over the first five years of a child's life on average. The dependent variable in Panel B and C is composition-adjusted for the birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects as well as weather and socio-economic controls. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level.

Table 4.A.5: The Effect of LEZ Implementation on Migration Patterns

	Total net migration *(1)*	Net migration among families *(2)*	Movers among AOK children *(3)*
LEZ treatment	0.66	-0.07	-0.56
	(0.47)	(0.55)	(0.33)
Mean outcome	2.93	-1.86	9.44
Weather controls	x	x	x
Socio-economic controls	x	x	x

Notes: This table reports coefficient estimates for the effect of LEZ implementation on county migration. The dependent variable in column (1) is net migration per 1,000 inhabitants of the total county population. The dependent variable in column (2) is net migration per 1,000 inhabitants of the total county population younger than 18 years old and 30 to 50 years old. The dependent variable in column (3) is the fraction of AOK-insured children moving out of the birth county after their first and before their sixth year of life in %. All columns include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects as well as weather and socio-economic controls. Observations of LEZ counties in the three quarters prior and subsequent to implementation are included. The regressions are weighted by the birth county–birth quarter cell size. Standard errors are clustered at the county level and are in parentheses. The sample size is 12,865.

Table 4.A.6: The Effect of Early-Life PM_{10} Exposure on Placebo Health Outcomes (per 1,000 children)

	Arm injuries (1)	Head injuries (2)	Several injuries (3)
		A. Reduced Form	
LEZ treatment	0.58	-0.97	-0.09
	(1.04)	(4.64)	(0.11)
Mean outcome	5.7	68.81	0.12
		B. IV Estimation	
PM_{10} mean	-0.43	0.71	0.06
	(0.75)	(3.39)	(0.08)
Mean outcome	5.7	68.81	0.12
Weather controls	x	x	x
Socio-economic controls	x	x	x

Notes: This table reports reduced form estimates (Panel A) and IV estimates (Panel B) for three different placebo health outcomes. The dependent variable is either the number of stationary hospital treatments of arm injuries, head injuries or injuries involving several body parts, that accumulate over the first five years of a child's life on average and per 1,000 children. The dependent variable is composition-adjusted for the birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects as well as weather and socio-economic controls. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is $9,609$.

TABLE 4.A.7: The Effect of Early-Life PM_{10} Exposure when limiting pre-treatment observations to cohorts born four quarters before implementation

	A. Number of prescriptions		B. Costs of prescriptions (€)	
	(1)	*(2)*	*(3)*	*(4)*
		Reduced Form Estimation		
Respiratory diseases	-0.63	-0.78	-10.85	-16.02
	(0.21)	(0.25)	(4.70)	(6.70)
Mean outcome	13.94	13.94	213.11	213.11
Asthma	-0.30	-0.31	-9.29	-13.12
	(0.08)	(0.1)	(3.53)	(4.87)
Mean outcome	2.46	2.46	71.96	71.96
		IV Estimation		
Respiratory diseases	0.49	0.49	8.48	10.2
	(0.22)	(0.21)	(4.51)	(5.13)
Mean outcome	13.94	13.94	213.11	213.11
Asthma	0.23	0.20	7.26	8.35
	(0.09)	(0.08)	(3.59)	(3.9)
Mean outcome	2.46	2.46	71.96	71.96
First stage F-statistic	9.05	12.75	9.05	12.75
Weather controls	x	x	x	x
Socio-economic controls		x		x

Notes: This table reports reduced form estimates (Panel A) and IV estimates (Panel B) for health effects when we use only cohorts born four quarters prior to LEZ implementation as pre-treatment observations of the treated. The dependent variable is the number of prescriptions that accumulate over the first five years of a child's life on average. It refers to either prescriptions for respiratory diseases in general or asthma specifically. The dependent variable is composition-adjusted for the birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 7, 893.

TABLE 4.A.8: The Effect of Early-Life PM_{10} Exposure on Medication of Respiratory Diseases throughout Early Childhood - Two-Way Fixed Effect Estimation

	A. Number of prescriptions		B. Costs of prescriptions (€)	
	(1)	*(2)*	*(3)*	*(4)*
		Reduced Form Estimation		
Respiratory diseases	-0.50	-0.52	-9.13	-11.83
	(0.15)	(0.18)	(2.97)	(3.80)
Mean outcome	14.14	14.14	218.62	218.62
Asthma	-0.18	-0.23	-6.04	-8.43
	(0.06)	(0.07)	(2.14)	(2.67)
Mean outcome	2.5	2.5	73.27	73.27
		IV Estimation		
Respiratory diseases	0.41	0.39	7.53	8.84
	(0.16)	(0.15)	(3.03)	(3.32)
Mean outcome	14.14	14.14	218.62	218.62
Asthma	0.15	0.17	4.99	6.30
	(0.06)	(0.06)	(2.09)	(2.35)
Mean outcome	2.5	2.5	73.27	73.27
First stage F-statistic	13.26	17.23	13.26	17.23
Weather controls	x	x	x	x
Socio-economic controls		x		x

Notes: This table replicates our main results in Table 4.5.2 using two-way fixed effect estimation. The dependent variable is either the number (panel A) or the costs in € (panel B) of prescriptions that accumulate over the first five years of a child's life on average. It refers to either prescriptions for respiratory diseases in general or asthma specifically. In each panel, coefficients from reduced form and IV estimations are presented. The dependent variable is composition-adjusted for the birth county–birth quarter cell. All regressions include birth county and birth state–birth quarter fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 2, 904.

TABLE 4.A.9: The Effect of Early-Life PM_{10} Exposure on Medication of Respiratory Diseases throughout Early Childhood - Logged outcomes

	A. Number of prescriptions		B. Costs of prescriptions (€)	
	(1)	*(2)*	*(3)*	*(4)*
	Reduced Form Estimation			
Respiratory diseases	-0.05	-0.06	-0.06	-0.07
	(0.02)	(0.02)	(0.02)	(0.02)
Mean outcome	14.14	14.14	218.62	218.62
Asthma	-0.09	-0.10	-0.11	-0.12
	(0.03)	(0.04)	(0.05)	(0.05)
Mean outcome	2.5	2.5	73.27	73.27
	IV Estimation			
Respiratory diseases	0.04	0.04	0.05	0.05
	(0.01)	(0.01)	(0.02)	(0.02)
Mean outcome	14.14	14.14	218.62	218.62
Asthma	0.07	0.07	0.08	0.09
	(0.02)	(0.03)	(0.04)	(0.04)
Mean outcome	2.5	2.5	73.27	73.27
First stage F-statistic	14.25	20.14	14.25	20.14
Weather controls	x	x	x	x
Socio-economic controls		x		x

Notes: This table replicates our main results in Table 4.5.2 using logged outcome variables. The dependent variable is either the number (panel A) or the costs in € (panel B) of prescriptions that accumulate over the first five years of a child's life on average. It refers to either prescriptions for respiratory diseases in general or asthma specifically. In each panel, coefficients from reduced form and IV estimations are presented. The dependent variable is composition-adjusted for the birth county–birth quarter cell. All regressions include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 9, 609.

TABLE 4.A.10: The Effect of LEZ Implementation on PM_{10} Concentrations - no IDW interpolation

	First Stage Estimation PM_{10} Pollution in μg/m^3	
	(1)	*(2)*
LEZ treatment	-1.30	-1.50
	(0.41)	(0.40)
Mean outcome	27.26	27.26
First stage F-statistic	10.27	14.12
Weather controls	x	x
Socio-economic controls		x

Notes: This table replicates Table 4.5.1 when we do not interpolate the pollution data but include only counties with own measuring stations in the sample. The dependent variable is the quarterly mean PM_{10} concentration in a given county and year in μg/m^3. All columns include birth county, birth state–birth quarter, LEZ wave–event time, and LEZ wave–treated fixed effects. Weather and socio-economic controls are added sequentially moving from left to right. The regressions are weighted by the birth county–birth quarter cell size. Standard errors in parentheses are clustered at the county level. The sample size is 8, 286.

TABLE 4.A.11: Multiple Hypotheses Testing

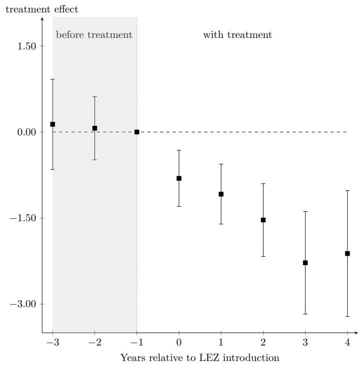
			Respiratory diseases				Asthma			
			Reduced form		IV		Reduced form		IV	
Outcome	Table	Specification	p	$p(BH)$	p	$p(BH)$	p	$p(BH)$	p	$p(BH)$
Prescriptions	3	2	0.000	0.002	0.002	0.011	0.003	0.015	0.012	0.035
Costs	3	4	0.001	0.011	0.009	0.027	0.014	0.039	0.034	0.071
Prescriptions year 1	4	2	0.002	0.011	0.003	0.015	0.423	0.465	0.429	0.465
Prescriptions year 2	4	3	0.000	0.002	0.002	0.011	0.001	0.011	0.010	0.030
Prescriptions year 3	4	4	0.000	0.002	0.002	0.011	0.003	0.015	0.008	0.026
Prescriptions year 4	4	5	0.018	0.044	0.029	0.062	0.036	0.071	0.053	0.092
Prescriptions year 5	4	6	0.055	0.093	0.092	0.139	0.225	0.289	0.247	0.302
Share of sufferers	4	7	0.007	0.026	0.008	0.026	0.008	0.026	0.018	0.044
Share of sufferers (year 1)	4	8	0.022	0.053	0.016	0.043	0.945	0.945	0.945	0.945
Share of sufferers (year 2)	4	9	0.041	0.077	0.036	0.071	0.023	0.053	0.044	0.079
Share of sufferers (year 3)	4	10	0.043	0.077	0.039	0.075	0.001	0.010	0.003	0.015
Share of sufferers (year 4)	4	11	0.025	0.056	0.035	0.071	0.064	0.106	0.078	0.123
Share of sufferers (year 5)	4	12	0.244	0.302	0.276	0.327	0.068	0.110	0.100	0.149
Prescriptions per sufferer	4	13	0.000	0.006	0.007	0.026	0.243	0.302	0.259	0.312
Prescriptions per sufferer (year 1)	4	14	0.123	0.167	0.156	0.204	0.290	0.339	0.309	0.356
Prescriptions per sufferer (year 2)	4	15	0.000	0.002	0.005	0.019	0.091	0.139	0.104	0.150
Prescriptions per sufferer (year 3)	4	16	0.000	0.003	0.004	0.015	0.420	0.465	0.427	0.465
Prescriptions per sufferer (year 4)	4	17	0.102	0.149	0.119	0.164	0.617	0.656	0.621	0.656
Prescriptions per sufferer (year 5)	4	18	0.114	0.161	0.148	0.197	0.801	0.823	0.801	0.823

Notes: This table reports p values for all 76 hypotheses regarding respiratory diseases and asthma tested in Tables 4.5.2 and 4.5.3. Columns labeled p indicate unadjusted p-values while columns labeled $p(BH)$ indicate p-values adjusted for multiple hypotheses testing following Benjamini and Hochberg (1995).

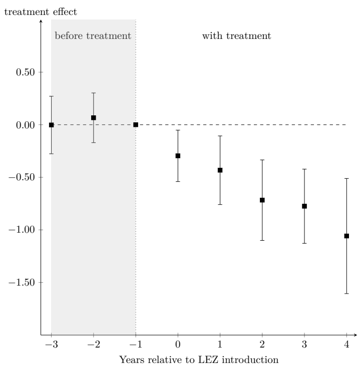
4.B Additional Figures

FIGURE 4.B.1: Event-study Estimates of LEZ Effects - Excluding All Already Treated Counties from the Control Group

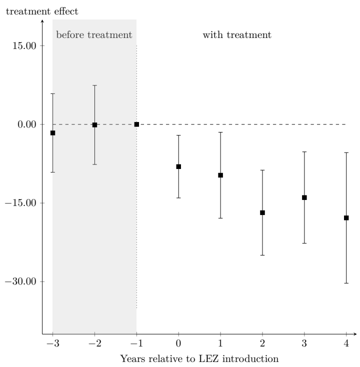
(A) PM_{10}



(B) Number of prescriptions



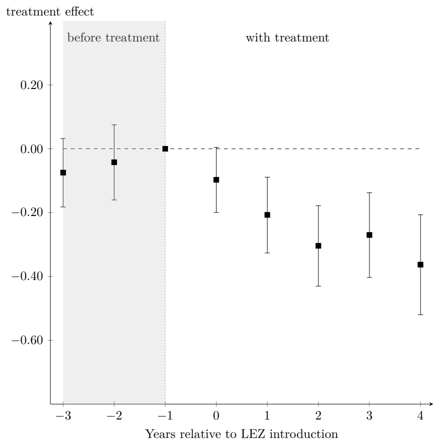
(C) Costs of prescriptions



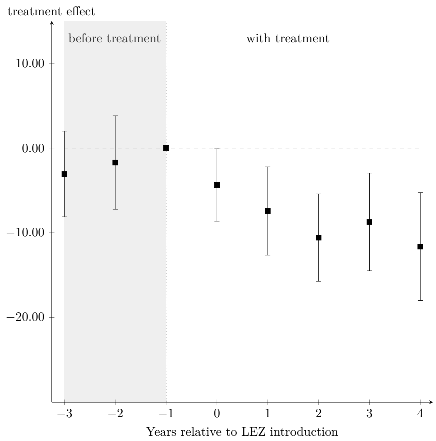
Notes: The figure presents event-study coefficients that show how LEZs affect PM_{10} concentration and the medication of respiratory diseases for a sample that excludes all already treated counties from the control group. The dependent variable is the mean quarterly PM_{10} level (Panel a), the number (Panel b) or the costs (Panel c) of prescriptions that accumulate over the first five years of a child's life on average. The gray shaded area indicates the pre-treatment period. The coefficient in the year prior to implementation is normalized to zero. The regression includes county fixed effects, state–quarter fixed effects, LEZ wave–event time fixed effects, LEZ wave–treated fixed effects as well as weather and socio-economic controls. It is weighted by the county–quarter cell size. Standard errors are clustered at the county level. Confidence intervals refer to the 5% level of significance.

FIGURE 4.B.2: Event-study Estimates of LEZ Effects on Medication for Asthma

(A) Number of prescriptions



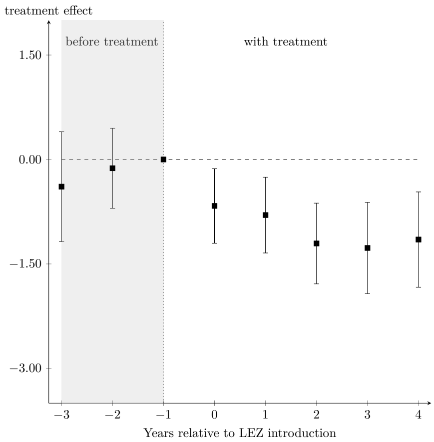
(B) Costs of prescriptions



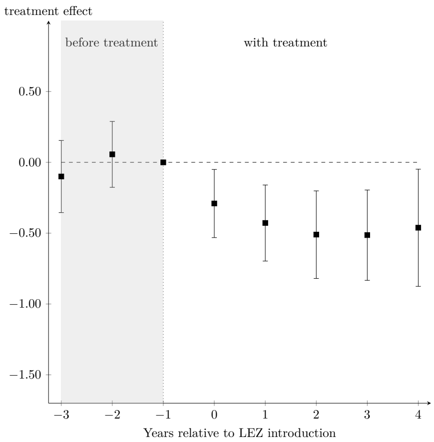
Notes: The figure presents event-study coefficients that show how LEZs affect the medication of asthma depending on the time distance between birth and LEZ implementation in years. The dependent variable is the number (Panel a) or the costs (Panel b) of prescriptions that accumulate over the first five years of a child's life on average. The gray shaded area indicates the pre-treatment period. The coefficient in the year prior to implementation is normalized to zero. The regression includes county fixed effects, state–quarter fixed effects, LEZ wave–event time fixed effects, LEZ wave–treated fixed effects as well as weather and socio-economic controls. It is weighted by the county–quarter cell size. Standard errors are clustered at the county level. Confidence intervals refer to the 5% level of significance.

FIGURE 4.B.3: Event-study Estimates Excluding the Never Treated

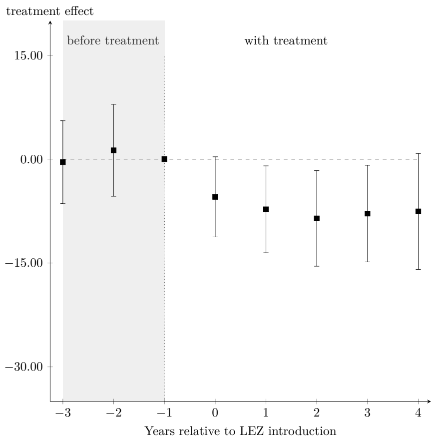
(A) PM_{10}



(B) Number of prescriptions



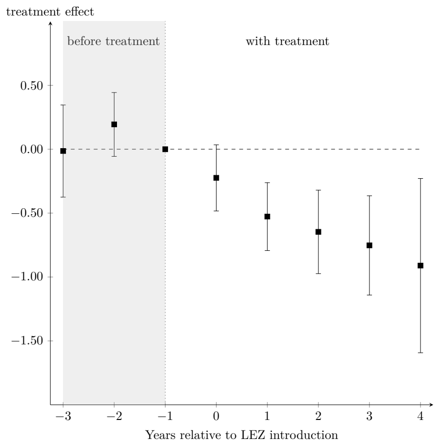
(C) Costs of prescriptions



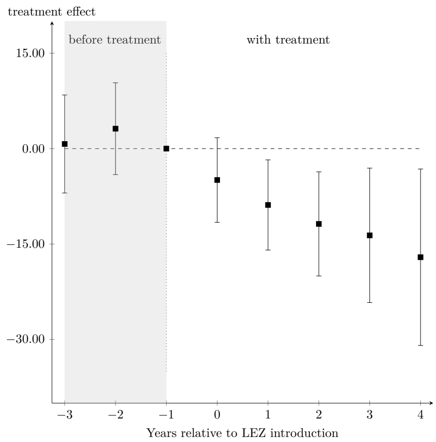
Notes: The figure presents event-study coefficients that show how LEZs affect PM_{10} pollution and the medication of respiratory diseases for a sample that excludes all never treated counties from the control group. The dependent variable is the average PM_{10} level in $\mu g/m^3$ (Panel a) or the number of prescriptions that accumulate over the first five years of a child's life on average (Panel b). We reduce the time window that defines our control group to 4 years post-treatment and 3 years pre-treatment to avoid that the control groups for the later treated become very small. The gray shaded area indicates the pre-treatment period. The coefficient in the year prior to implementation is normalized to zero. The regression includes county fixed effects, state–quarter fixed effects, LEZ wave–event time fixed effects, LEZ wave–treated fixed effects as well as weather and socio-economic controls. It is weighted by the county–quarter cell size. Standard errors are clustered at the county level. Confidence intervals refer to the 5% level of significance.

Figure 4.B.4: Event-study Estimates of LEZ Effects - Accounting for Increases in Policy Stringency

(a) Number of prescriptions



(b) Costs of prescriptions



Notes: The figure presents event-study coefficients that show how LEZs affect the medication of respiratory diseases based on Equation (4.3) enriched by eight additional dummies $\sum_{y=2}^{5} f_{cy}^{Euro2} + \sum_{y=2}^{5} f_{cy}^{Euro3}$, which should absorb potential additional health benefits in life years two to five linked to the two more stringent LEZ regimes banning Euro 2 and Euro 3 vehicles, respectively. The dependent variable is either the number (Panel a) or the costs of prescriptions (Panel b) that accumulate over the first five years of a child's life on average. The gray shaded area indicates the pre-treatment period. The coefficient in the year prior to implementation is normalized to zero. The regression includes county fixed effects, state–quarter fixed effects, LEZ wave–event time fixed effects, LEZ wave–treated fixed effects as well as weather and socio-economic controls. It is weighted by the county–quarter cell size. Standard errors are clustered at the county level. Confidence intervals refer to the 5% level of significance.

4.C Composition-adjusted health outcomes

For the estimation of Equation (4.2), health outcomes observed at the level of the individual child (H_{ict}) are aggregated to the cohort level (H_{ct}). We define a cohort by its birth county c as well as its birth year and birth quarter t. However, additional information at the level of the individual such as the individual's sex or the precise location of residence within a county at the five-digit zip code is available. To exploit this information, we conduct auxiliary regressions that are commonly used in the literature (e.g. Currie et al., 2015).

In a first step, we regress the children's health outcomes on individual-level covariates as well as birth county–birth quarter fixed effects:

$$H_{ict} = W'_{ict}\zeta + \phi_{ct} + \xi_{ict} \tag{4.4}$$

where the dependent variable H_{ict} is the accumulated health outcome over the first five years of life for individual i born in county c and year and quarter t. W'_{ict} is a vector of individual-level covariates that includes gender and location of residence within a county at the five-digit zip code. Additionally, Equation (4.4) controls for a full set of birth county–birth quarter indicators ϕ_{ct}. Their coefficient estimates $\widehat{\phi_{ct}}$ are orthogonal to the covariates at the individual level. In other words, they give the average health outcomes for a birth county–birth quarter cohort after controlling for gender and residence. In line with Isen et al. (2017), we refer to the predicted cohort means obtained by this approach as composition-adjusted. We use the composition-adjusted outcomes as the dependent variable in Equation (4.2).

The use of composition-adjusted group means is asymptotically equivalent to using the individual level data (e.g. Donald and Lang, 2007) if the sampling variance of the composition-adjusted group estimates is taken into account. In accordance with other studies (e.g. Angrist and Lavy, 2009; Albouy, 2009b; Currie et al., 2015; Isen et al., 2017), we estimate all regressions by weighted least squares using the number of individuals in each birth county–birth quarter cell as weights. This is assumed to be a reasonable approximation of weighting by inverse sampling variance. Compared to running regressions on the individual level data, the estimation of models collapsed to the level of variation ensures that tests are of correct size given serial correlation in the within-group errors (Isen et al., 2017). Additionally, it requires substantially less computational power.

4.D Data

4.D.1 Control variables

Table 4.D.1 gives an overview of the county-specific control variables used in the estimations. We observe cohorts over a five-year period and we include weather controls for all of these years. The 2007 values of socio-economic demographic controls are categorized in terciles and interacted with year-quarter dummies in the regressions.

TABLE 4.D.1: Control Variables

Variable	Definition	Source
Weather Controls		
Precipitation	Sum of the precipitation height in mm	DWD
Sunshine	Total sunshine duration in hours	DWD
Temperature	Mean, minimum and maximum temperature, 12 separate terms that count the number of days with temperatures above 0, 5, 10, 15, 20, 25, 29, 30, 31, 32, 33 and 34 degree Celsius	DWD
Wind	Average windspeed 10 m above ground in m/s	DWD
Relative humidity	Relative humidity at 2 m above ground in %	DWD
Pressure	Mean vapor pressure in hpa	DWD
Socio-economic Demographic Controls		
Average age	Average age of the county population	BBSR
Population density	Residents per km^2	BBSR
Migration in	People moving out of county per 1,000 inhabitants	BBSR
Migration out	People moving into county per 1,000 inhabitants	BBSR
Moving AOK children	Share of AOK-insured children moving out of county	WIdO
Women share	Female to male population ratio	BBSR
Share of foreigners	Percentage of people without German citizenship	BBSR
Women share in foreigners	Share of female foreigners among foreigners	BBSR
Employment	Employees subject to social insurance contributions per 100 inhabitants of working age	BBSR
Gross Value Added (GVA)	Total gross value added in 1,000 € per employed person	BBSR
GVA share in primary sector	Share of gross value added in the primary sector in %	BBSR
GVA share in tertiary sector	Share of gross value added in the tertiary sector in %	BBSR
Household income	Average household income in € per inhabitant	BBSR
Housing transfers	Number of households receiving housing benefits, per 1,000 households	BBSR
Education	Share of students graduating with higher education entrance qualification	BBSR
Marriages	Marriages per 1,000 inhabitants 18 years and older	BBSR
Share of young mothers	Births of mothers in the age group 15 to under 20 years per 1,000 women in the age group	BBSR
Share of older mothers	Births of mothers 40 years and older per 1,000 women aged 40 to under 45	BBSR

4.D.2 Aggregating the pollution data

We aggregate the pollution data by averaging the daily PM_{10} readings of all measuring stations in a county and quarter. We weight each observation by the number of station readings in that period (c.p. Chay and Greenstone, 2003a; Isen et al., 2017).

In the few counties in our sample without measuring stations, we interpolate pollution exposure from surrounding stations using Inverse Distance Weighting (IDW). Following Karlsson and Ziebarth (2018), we consider all stations within a 60 km (37.5 miles) radius of the county's centroid. We then calculate the weighted average using both the number of station measurements and the inverse distance of the monitors to the centroid as weights.

To avoid fluctuations in pollution levels linked to stations not being active regularly, we generate the quarterly averages including only stations with at least 60 measurements. Moreover, to avoid bias from interpolating pollution levels from treated to nearby untreated counties, we only use stations outside of LEZ counties for the interpolation. We show that our results are robust when we do not interpolate the pollution data and include only counties with own measuring stations in the sample in Table 4.A.10.

4.D.3 Prescription data

The identification process of pharmaceutical substances that are relevant in the therapy of respiratory diseases and asthma specifically is as follows:

Pharmaceuticals for respiratory diseases We use a publication akin to the Red Book called "Gelbe Liste" by the ISO 9001:2015 certified Medizinische Medien Informations GmbH which serves as a source of information for medical and pharmaceutical professionals. For more than 120,000 drugs, it links ATC-code classified pharmaceutical substances to ICD-10-code classified clinical diagnoses. By linking ATC to ICD codes, we can determine for which diseases different pharmaceutical substances are commonly prescribed. From the registered information we draw 6,479 unique links, of which we select only those related to respiratory diseases (150 substances). While this approach is comprehensive, it suffers from the drawback that it may also cover substances generically administered for a broad variety of diseases.

Pharmaceuticals for asthma Additionally, we define a smaller list of pharmaceuticals that are closely tied to asthma. To this end, we consult annually updated lists of the substances prescribed most often for asthma in a given year, that is substances in the ATC category R03. The lists are prepared by IGES institute for the years 2006 to 2017.[36] In our analysis, we consider only prescriptions of the 20 most often prescribed substances in the year the prescription is issued. Note, that the top 20 substances cover almost the entire market of substances prescribed for asthma and COPD, however, they may not include substances prescribed in rare cases. The pharmaceuticals identified according to this procedure represent a strict subset of those compiled in the prior approach).

The prescription costs are adjusted to allow for intertemporal comparisons as if the average cost per prescription had not changed. In other words, we take both inflation but also ATC-specific market price changes, such as expiring patents, into account. To this end, we calculate ATC-specific price indices normalized to the fourth quarter of 2017 using available prescription data for all children in Germany. Based on the generated price indices we adjust the prescription costs observed in our sample to real values, before aggregating them to the cohort level.

4.E Unconditional quantile regression

We estimate an unconditional quantile regression (Firpo et al., 2009) to flexibly estimate LEZ treatment effects across the unconditional distribution of our health outcomes. The approach is based on the use of the re-centered influence function (RIF) defined in Equation (4.5).

$$RIF(H_i, q_\theta) = q_\theta + IF(H_i, q_\theta) = q_\theta + \frac{\theta - \mathbb{I}(H_i \leq q_\theta)}{f_H(q_\theta)} \tag{4.5}$$

It is the sum of the influence function (IF) and the θth quantile of the unconditional distribution of the health variable H denoted as q_θ. The IF indicates the marginal influence of an observation H_i on the quantile q_θ. It is determined by f_H, the empirical

[36]More information on the underlying data and aggregation methodologies are provided on the IGES website and in the latest published report Häussler and Höer (2016).

density function evaluated at q_θ, and by the indicator $1(h \leq q_\theta)$ which is equal to 1 if H_i is below or equal to q_θ. Thus, an observation's influence is negative if its health status lies below and positive if it lies above the health status at the θth quantile.

The expected value of the RIF equals the quantile of the unconditional distribution.[37] By the law of iterated expectations and integration over the conditional mean, the unconditional quantile q_θ can be expressed as

$$q_\theta = E[RIF(H_i, q_\theta)] = E[E[RIF(H_i, q_\theta)|X_i]] = \int E[RIF(H_i, q_\theta)|X_i]dF_X$$

where X is the vector of covariates and F_X is the marginal distribution function of X. To obtain the marginal treatment effects on the unconditional quantile q_θ, we take the sample quantile $\hat{q}_\theta$ and retrieve the density $\hat{f}_H$ using a Gaussian kernel method.[38] To obtain $\widehat{RIF}$, we substitute both into Equation (4.5). Secondly, we apply RIF-OLS regression to obtain the coefficients representing the marginal *ceteris paribus* effect of an infinitesimal shift in the distribution of the covariates X on the unconditional θth quantile of H:

$$\hat{\beta}_\theta = \left(\sum_{i=1}^{N} X_i' X_i\right)^{-1} \sum_{i=1}^{N} X_i' \widehat{RIF}(H_i, \hat{q}_\theta)$$

The identifying assumption is that in the absence of treatment, the change in the health outcome at each quantile would have been the same in the treatment and the control group. Because endogenous regressors cannot be addressed by the conventional unconditional quantile regression framework, and because estimation times are prohibitively long when using a stacked design, we limit our quantile regression analysis to reduced form estimations using a standard DID design knowing that some caveats may apply (cf. Section 4.5.5)

[37] $E[RIF(H_i, q_\theta)] = E[q_\theta] + \frac{\theta - E[\mathbb{I}(H_i \leq q_\theta)]}{f_H(q_\theta)} = q_\theta + \frac{\theta - \theta}{f_H(q_\theta)} = q_\theta$

[38] $\hat{f}_H(\hat{q}_\theta) = \frac{1}{N \cdot b_H} \cdot \sum_{i=1}^{N} K_H\left(\frac{H_i - \hat{q}_\theta}{b_H}\right)$, where K_H is the kernel function and b_H is a positive scalar bandwidth.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505. [Cited on pages 130 and 131.]

Abendroth, A.-K., S. Melzer, A. Kalev, and D. Tomaskovic-Devey (2017): "Women at Work: Women's Access to Power and the Gender Earnings Gap," *ILR Review*, 70, 190–222. [Cited on page 10.]

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Tech. Rep. 2002-06, Center for Economic Studies, U.S. Census Bureau. [Cited on pages 19 and 20.]

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333. [Cited on pages 9 and 19.]

Abraham, S. and L. Sun (2020): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*. [Cited on page 169.]

Alba, A. (2009): "La incapacidad temporal para el trabajo: análisis económico de su incidencia y su duración," Technical report, Universidad Carlos III de Madrid. [Cited on page 64.]

Albouy, D. (2009a): "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, 117, 635–667. [Cited on page 171.]

——— (2009b): "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values," *NBER Working Paper*, 1–32. [Cited on pages 171 and 215.]

Alexander, D. and H. Schwandt (2019): "The Impact of Car Pollution on Infant and Child Health: Evidence from Emissions Cheating," *Working Paper No. WP-2019-4*. [Cited on pages 162 and 187.]

ALMOND, D. AND J. CURRIE (2011): "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 25, 153–72. [Cited on page 158.]

ALMOND, D., J. CURRIE, AND V. DUQUE (2018): "Childhood Circumstances and Adult Outcomes: Act II," *Journal of Economic Literature*, 56, 1360–1446. [Cited on page 158.]

ALTONJI, J. G. AND R. M. BLANK (1999): "Chapter 48 Race and Gender in the Labor Market," in *Handbook of Labor Economics*, Elsevier, vol. 3, 3143–3259. [Cited on pages 8 and 9.]

ANDERSON, J. (1979): "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69, 106–16. [Cited on page 133.]

ANDREWS, I., J. H. STOCK, AND L. SUN (2019): "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–753. [Cited on page 177.]

ANGRIST, J. AND V. LAVY (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99, 1384–1414. [Cited on pages 171 and 215.]

ARAGÓN, F., 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. [Cited on pages 51, 53, and 78.]

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, 257–280. [Cited on pages 53, 63, 172, and 195.]

ATALAY, E., A. HORTACSU, M. J. LI, AND C. SYVERSON (2019): "How Wide Is the Firm Border?" *The Quarterly Journal of Economics*, 134, 1845–1882. [Cited on page 133.]

ATHEY, S. AND G. W. IMBENS (2018): "Design-based Analysis in Difference-in-differences Settings with Staggered Adoption," *NBER Working Paper*. [Cited on page 169.]

AUFFHAMMER, M., S. M. HSIANG, W. SCHLENKER, AND A. SOBEL (2013): "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change," *Review of Environmental Economics and Policy*, 7, 181–198. [Cited on page 120.]

AZMAT, G. AND R. FERRER (2017): "Gender Gaps in Performance: Evidence from Young Lawyers," *Journal of Political Economy*, 125, 1306–1355. [Cited on pages 5, 10, 17, 18, 29, and 43.]

Babcock, L., S. Laschever, M. Gelfand, and D. Small (2003): "Nice Girls Don't Ask," *Harvard Business Review.* [Cited on pages 5 and 17.]

Baccarelli, A. and V. Bollati (2009): "Epigenetics and Environmental Chemicals," *Current Opinion in Pediatrics*, 21, 243–251. [Cited on page 166.]

Bagues, M. F. and B. Esteve-Volart (2010): "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *The Review of Economic Studies*, 77, 1301–1328. [Cited on pages 6 and 41.]

Baier, S. L. and J. H. Bergstrand (2007): "Do free trade agreements actually increase members' international trade?" *Journal of International Economics*, 71, 72–95. [Cited on page 133.]

Baker, M. and N. M. Fortin (2001): "Occupational Gender Composition and Wages in Canada, 1987–1988," *Canadian Journal of Economics*, 34, 345–376. [Cited on page 171.]

Barahona, N., F. A. Gallego, and J.-P. Montero (2020): "Vintage-Specific Driving Restrictions," *The Review of Economic Studies*, 87, 1646–1682. [Cited on page 115.]

Barnet-Verzat, C. and F.-C. Wolff (2008): "Gender Wage Gap and the Glass Ceiling Effect: A Firm-level Investigation," *International Journal of Manpower*, 29, 486–502. [Cited on page 9.]

Barrot, J.-N. and J. Sauvagnat (2016): "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks," *The Quarterly Journal of Economics*, 131, 1543–1592. [Cited on page 170.]

Barwick, P. J., S. Li, D. Rao, and N. B. Zahur (2018): "The Morbidity Cost of Air Pollution: Evidence from Consumer Spending in China," *NBER Working Paper.* [Cited on pages 159 and 187.]

Beatty, T. K. and J. P. Shimshack (2014): "Air Pollution and Children's Respiratory Health: A Cohort Analysis," *Journal of Environmental Economics and Management*, 67, 39–57. [Cited on pages 184 and 187.]

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300. [Cited on pages 184 and 210.]

Bento, A., M. Freedman, and C. Lang (2015): "Who Benefits From Environmental Regulation? Evidence From the Clean Air Act Amendments," *Review of Economics and Statistics*, 97, 610–622. [Cited on page 164.]

Bentolila, S., J. J. Dolado, and J. Jimeno (2019): "Dual Labor Markets Revisited," Technical report, Centro de Estudios Monetarios y Financieros. [Cited on pages 54 and 81.]

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262. [Cited on page 135.]

Bertrand, M. (2011): "Chapter 17 - New Perspectives on Gender," in *Handbook of Labor Economics*, ed. by D. Card and O. Ashenfelter, Elsevier, vol. 4, 1543–1590. [Cited on page 8.]

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2019): "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway," *The Review of Economic Studies*, 86, 191–239. [Cited on pages 10 and 44.]

Bertrand, M., C. Goldin, and L. F. Katz (2010): "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, 2, 228–255. [Cited on page 18.]

Bharadwaj, P., M. Gibson, J. G. Zivin, and C. Neilson (2017): "Gray Matters: Fetal Pollution Exposure and Human Capital Formation," *Journal of the Association of Environmental and Resource Economists*, 4, 505–542. [Cited on pages 111 and 161.]

Biasi, B. and H. Sarsons (2020): "Flexible Wages, Bargaining, and the Gender Gap," Working Paper 27894, National Bureau of Economic Research. [Cited on page 10.]

Birken, C. S. and C. MacArthur (2004): "Socioeconomic Status and Injury Risk in Children," *Paediatrics & Child Health*, 9, 323–325. [Cited on page 193.]

Bishop, K. C., J. D. Ketcham, and N. V. Kuminoff (2018): "Hazed and Confused: The Effect of Air Pollution on Dementia," NBER Working Papers 24970. [Cited on page 74.]

Blackman, A., F. Alpízar, F. Carlsson, and M. R. Planter (2018): "A Contingent Valuation Approach to Estimating Regulatory Costs: Mexico's Day without Driving Program," *Journal of the Association of Environmental and Resource Economists*, 5, 607–641. [Cited on page 114.]

Blackman, A., P. Qin, and J. Yang (2020): "How costly are driving restrictions? Contingent valuation evidence from Beijing," *Journal of Environmental Economics and Management*, 104. [Cited on page 114.]

Blau, F. D. and L. M. Kahn (2017): "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 55, 789–865. [Cited on pages 5, 8, 9, and 18.]

Blinder, A. S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources*, 8, 436–455. [Cited on pages 7 and 18.]

BLS (2019): "Report 1083," *U.S. Bureau of Labor Statistics.* [Cited on page 5.]

Blum, B. S. and A. Goldfarb (2006): "Does the internet defy the law of gravity?" *Journal of International Economics*, 70, 384–405. [Cited on page 124.]

Bohren, J. A., A. Imas, and M. Rosenberg (2019): "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 109, 3395–3436. [Cited on pages 8, 10, 40, and 41.]

Boletin Oficial Ayuntamiento de Madrid (2018): "Acuerdo de 29 de octubre de 2018 de la Junta de Gobierno de la Ciudad de Madrid por el que se desarrolla el régimen de gestión y funcionamiento de la Zona de Bajas Emisiones "Madrid Central"," *BOAM*, 8268/2199. [Cited on page 116.]

Bonhomme, S. and L. Hospido (2017): "The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data," *Economic Journal*, 127, 1244–1278. [Cited on page 78.]

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019): "Beliefs about Gender," *American Economic Review*, 109, 739–773. [Cited on page 14.]

Braithwaite, I., S. Zhang, J. B. Kirkbride, D. P. J. Osborn, and J. F. Hayes (2019): "Air Pollution (Particulate Matter) Exposure and Associations with Depression, Anxiety, Bipolar, Psychosis and Suicide Risk: A Systematic Review and Meta-Analysis," *Environmental Health Perspectives*, 127, 126002. [Cited on page 74.]

Cabrales, A., J. J. Dolado, and R. Mora (2014): "Dual Labour Markets and (Lack of) On-the-Job Training: PIAAC Evidence from Spain and Other EU Countries," IZA Discussion Papers 8649, Institute for the Study of Labor (IZA). [Cited on pages 54 and 80.]

Card, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957–979. [Cited on pages 82 and 108.]

Card, D., A. R. Cardoso, and P. Kline (2016): "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women," *The Quarterly Journal of Economics*, 131, 633–686. [Cited on pages 9 and 18.]

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2020): "Are Referees and Editors in Economics Gender Neutral?" *The Quarterly Journal of Economics*, 135, 269–327. [Cited on pages 9 and 34.]

Card, D., J. Heining, and P. Kline (2013): “Workplace Heterogeneity and the Rise of West German Wage Inequality,” *The Quarterly Journal of Economics*, 128, 967–1015. [Cited on pages 24 and 38.]

Carrillo, P. E., A. Lopez-Luzuriaga, and A. S. Malik (2018): “Pollution or crime: The effect of driving restrictions on criminal activity,” *Journal of Public Economics*, 164, 50–69. [Cited on page 111.]

Carrillo, P. E., A. S. Malik, and Y. Yoo (2016): “Driving restrictions that work? Quito’s Pico y Placa Program,” *Canadian Journal of Economics/Revue canadienne d’économique*, 49, 1536–1568. [Cited on page 115.]

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019): “The Effect of Minimum Wages on Low-wage Jobs,” *The Quarterly Journal of Economics*, 134, 1405–1454. [Cited on page 167.]

Chang, E. H., K. L. Milkman, L. J. Zarrow, K. Brabaw, D. M. Gromet, R. Rebele, C. Massey, A. L. Duckworth, and A. Grant (2019a): “Does Diversity Training Work the Way It’s Supposed To?” *Harvard Business Review*. [Cited on pages 5 and 44.]

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, 141–169. [Cited on pages 53 and 111.]

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, 151–72. [Cited on page 53.]

Chay, K. Y. and M. Greenstone (2003a): “Air Quality, Infant Mortality, and the Clean Air Act of 1970,” *NBER Working Paper*, 1–43. [Cited on pages 161, 172, 174, 195, and 218.]

——— (2003b): “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession,” *The Quarterly Journal of Economics*, 118, 1121–1167. [Cited on pages 53, 111, 161, and 164.]

——— (2005): “Does Air Quality Matter? Evidence From the Housing Market,” *Journal of Political Economy*, 113, 376–424. [Cited on page 164.]

Chen, Y., G. Z. Jin, N. Kumar, and G. Shi (2013): “The promise of Beijing: Evaluating the impact of the 2008 Olympic Games on air quality,” *Journal of Environmental Economics and Management*, 66, 424–443. [Cited on page 115.]

Chen, Y. and A. Whalley (2012): “Green Infrastructure: The Effects of Urban Rail Transit on Air Quality,” *American Economic Journal: Economic Policy*, 4, 58–97. [Cited on page 115.]

Clark, N. A., P. A. Demers, C. J. Karr, M. Koehoorn, C. Lencar, L. Tamburic, and M. Brauer (2010): “Effect of Early Life Exposure to Air Pollution on Development of Childhood Asthma,” *Environmental Health Perspectives*, 118, 284–290. [Cited on page 187.]

Colmer, J., D. Lin, S. Liu, and J. Shimshack (2020): “Why are Pollution Damages Lower in Developed Countries? Insights from High Income, High-Pollution Hong Kong,” *CEP Discussion Paper*. [Cited on pages 172 and 195.]

Cullen, Z. B. and R. Perez-Truglia (2019): “The Old Boys’ Club: Schmoozing and the Gender Gap,” Working Paper 26530, National Bureau of Economic Research. [Cited on pages 10 and 42.]

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015): “Do Housing Prices Reflect Environmental Health Risks? Evidence from More than 1600 Toxic Plant Openings and Closings,” *American Economic Review*, 105, 678–709. [Cited on pages 171 and 215.]

Currie, J., E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin (2009a): “Does pollution increase school absences?” *Review of Economics and Statistics*, 91, 682–694. [Cited on pages 51, 53, and 161.]

Currie, J. and M. Neidell (2005): “Air Pollution and Infant Health: What Can We Learn from California’s Recent Experience?” *The Quarterly Journal of Economics*, 120, 1003–1030. [Cited on pages 53, 112, 161, 172, 174, and 195.]

Currie, J., M. Neidell, and J. F. Schmieder (2009b): “Air pollution and infant health: Lessons from New Jersey,” *Journal of Health Economics*, 28, 688 – 703. [Cited on page 53.]

Currie, J. and R. Walker (2011): “Traffic Congestion and Infant Health: Evidence from E-ZPass,” *American Economic Journal: Applied Economics*, 3, 65–90. [Cited on pages 115 and 162.]

——— (2019): “What Do Economists Have to Say about the Clean Air Act 50 Years after the Establishment of the Environmental Protection Agency?” *Journal of Economic Perspectives*, 33, 3–26. [Cited on page 115.]

Currie, J., J. G. Zivin, J. Mullins, and M. Neidell (2014): “What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution?” *Annual Review of Resource Economics*, 6, 217–247. [Cited on pages 53, 83, and 158.]

Cyrys, J., A. Peters, J. Soentgen, and H.-E. Wichmann (2014): "Low Emission Zones Reduce PM10 Mass Concentrations and Diesel Soot in German Cities," *Journal of the Air & Waste Management Association*, 64, 481–487. [Cited on page 159.]

Davis, L. (2008): "The Effect of Driving Restrictions on Air Quality in Mexico City," *Journal of Political Economy*, 116, 38–81. [Cited on page 115.]

de Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-way Fixed Rffects Rstimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–96. [Cited on page 169.]

Dehejia, R. and A. Lleras-Muney (2004): "Booms, Busts, and Babies' Health," *The Quarterly Journal of Economics*, 119, 1091–1130. [Cited on page 164.]

Deng, Q., C. Lu, C. Ou, L. Chen, and H. Yuan (2016): "Preconceptional, Prenatal and Postnatal Exposure to Outdoor and Indoor Environmental Factors on Allergic Diseases/Symptoms in Preschool Children," *Chemosphere*, 152, 459–467. [Cited on page 187.]

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019): "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, 109, 4178–4219. [Cited on pages 51, 53, 63, 83, 112, 159, 162, 172, and 195.]

Deschênes, O., M. Greenstone, and J. S. Shapiro (2017): "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program," *American Economic Review*, 107, 2958–89. [Cited on pages 159 and 162.]

Deshpande, M. and Y. Li (2019): "Who is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 11, 213–48. [Cited on pages 159 and 167.]

Díaz-García, C., A. González-Moreno, and F. J. Sáez-Martínez (2013): "Gender Diversity within R&D Teams: Its Impact on Radicalness of Innovation," *Innovation*, 15, 149–160. [Cited on page 39.]

Dohmen, T. J., H. Lehmann, and A. Zaiceva (2008): "The Gender Earnings Gap inside a Russian Firm: First Evidence from Personnel Data - 1997 to 2002," *Zeitschrift für ArbeitsmarktForschung - Journal for Labour Market Research*, 41, 157–179. [Cited on page 9.]

Dolado, J. J., C. García-Serrano, and J. F. Jimeno (2002): "Drawing Lessons from the Boom of Temporary Jobs in Spain," *Economic Journal*, 112, F270–F295. [Cited on pages 54 and 80.]

Dolado, J. J., M. Jansen, and J. F. Jimeno (2005): "Dual Employment Protection Legislation: A Framework for Analysis," CEPR Discussion Papers 5033, C.E.P.R. Discussion Papers. [Cited on page 80.]

Donald, S. G. and K. Lang (2007): "Inference with Difference-in-Differences and Other Panel Data," *The Review of Economics and Statistics*, 89, 221–233. [Cited on pages 171 and 215.]

Eaton, J. and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779. [Cited on page 133.]

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, 36–65. [Cited on pages 51, 53, and 111.]

Egan, M. L., G. Matvos, and A. Seru (2017): "When Harry Fired Sally: The Double Standard in Punishing Misconduct," Working Paper 23242, National Bureau of Economic Research. [Cited on page 10.]

Escudero, M., X. Querol, A. Ávila, and E. Cuevas (2007): "Origin of the exceedances of the European daily PM limit value in regional background areas of Spain," *Atmospheric Environment*, 41, 730 – 744. [Cited on page 62.]

Eskeland, G. and T. Feyzioglu (1997): "Rationing Can Backfire: The "Day without a Car" in Mexico City," *World Bank Economic Review*, 11, 383–408. [Cited on page 115.]

Fadlon, I. and T. H. Nielsen (2019): "Family Health Behaviors," *American Economic Review*, 109, 3162–3191. [Cited on pages 159 and 167.]

Faelker, T., W. Pickett, and R. J. Brison (2000): "Socioeconomic Differences in Childhood Injury: a Population Based Epidemiologic Study in Ontario, Canada," *Injury Prevention*, 6, 203–208. [Cited on page 193.]

Fan, S.-J., J. Heinrich, M. S. Bloom, T.-Y. Zhao, T.-X. Shi, W.-R. Feng, Y. Sun, J.-C. Shen, Z.-C. Yang, B.-Y. Yang, and G.-H. Dong (2020): "Ambient air pollution and depression: A systematic review with meta-analysis up to 2019," *Science of The Total Environment*, 701, 134721. [Cited on page 74.]

Fanta, C. (2009): "Drug Therapy: Asthma," *New England Journal of Medicine*, 360, 1002–1014. [Cited on page 162.]

Fernandez-Val, I. (2009): "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," *Journal of Econometrics*, 150, 71–85. [Cited on page 74.]

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973. [Cited on pages 160, 184, 186, and 219.]

Frederiksen, A., L. B. Kahn, and F. Lange (2019): "Supervisors and Performance Management Systems," *Journal of Political Economy*, 128, 2123–2187. [Cited on pages 6 and 17.]

Galdon-Sanchez, J., R. Gil, and G. Uriz-Uharte (2020): "The Value of Informaton in Competitive Markets: The Impact of IT and Big Data on Small and Medium Enterprises," *Working Paper*. [Cited on page 121.]

Gallego, F., J.-P. Montero, and C. Salas (2013): "The effect of transport policies on car use: Evidence from Latin American cities," *Journal of Public Economics*, 107, 47–62. [Cited on page 115.]

Gehrsitz, M. (2017): "The Effect of Low Emission Zones on Air Pollution and Infant Health," *Journal of Environmental Economics and Management*, 83, 121–144. [Cited on pages 115, 159, 160, 164, 188, 189, and 191.]

Gluckman, P. D., M. A. Hanson, C. Cooper, and K. L. Thornburg (2008): "Effect of in Utero and Early-Life Conditions on Adult Health and Disease," *New England Journal of Medicine*, 359, 61–73. [Cited on page 166.]

Goldin, C. (2014): "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104, 1091–1119. [Cited on pages 5 and 9.]

Goldin, C. and L. F. Katz (2015): "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation," *Journal of Labor Economics*, 34, 705–746. [Cited on pages 17, 27, and 43.]

Goodman-Bacon, A. (2018): "Difference-in-Differences with Variation in Treatment Timing," *NBER Working Paper*, 1–46. [Cited on pages 159, 169, and 197.]

Graff Zivin, J. and M. Neidell (2012): "The Impact of Pollution on Worker Productivity," *American Economic Review*, 102, 3652–3673. [Cited on pages 51 and 53.]

——— (2013): "Environment, Health, and Human Capital," *Journal of Economic Literature*, 51, 689–730. [Cited on pages 53, 59, and 115.]

Greenstone, M. and R. Hanna (2014): "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India," *American Economic Review*, 104, 3038–3072. [Cited on page 112.]

Greenstone, M. and B. K. Jack (2015): "Envirodevonomics: A Research Agenda for an Emerging Field," *Journal of Economic Literature*, 53, 5–42. [Cited on page 112.]

Halldén, K., J. Säve-Söderbergh, and Å. Rosén (2018): “Gender of the Immediate Manager and Women’s Wages: The Importance of Managerial Position,” *Social Science Research*, 72, 115–133. [Cited on page 10.]

Hammitt, J. K. and Y. Zhou (2006): “The Economic Value of Air-Pollution-Related Health Risks in China: A Contingent Valuation Study,” *Environmental and Resource Economics*, 33, 399–423. [Cited on page 112.]

Hanna, R. and P. Oliva (2015): “The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City,” *Journal of Public Economics*, 122, 68–79. [Cited on pages 51, 53, 59, and 63.]

Hansen, A. C. and H. K. Selte (2000): “Air Pollution and Sick-leaves,” *Environmental and Resource Economics*, 16, 31–50. [Cited on pages 53 and 79.]

Hausman, J. A., B. D. Ostro, and D. A. Wise (1984): “Air Pollution and Lost Work,” NBER Working Papers 1263. [Cited on pages 51, 53, and 79.]

Häussler, B. and A. Höer (2016): *Arzneimittel-Atlas 2018: Der Arzneimittelverbrauch in der GKV*, MWV Medizinisch Wissenschaftliche Verlagsgesellschaft. [Cited on page 219.]

He, J., N. Gouveia, and A. Salvo (2019a): “External Effects of Diesel Trucks Circulating Inside the São Paulo Megacity,” *Journal of the European Economic Association*, 17, 947–989. [Cited on pages 112 and 162.]

He, J., H. Liu, and A. Salvo (2019b): “Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China,” *American Economic Journal: Applied Economics*, 11, 173–201. [Cited on pages 53, 63, and 71.]

Head, K. and T. Mayer (2014): “Gravity Equations: Workhorse,Toolkit, and Cookbook,” Handbook of International Economics, Elsevier. [Cited on page 133.]

HEI (2010): *Traffic-related Air Pollution: a Critical Review of the Literature on Emissions, Exposure, and Health Effects*, Health Effects Institute. Panel on the Health Effects of Traffic-Related Air Pollution. [Cited on page 190.]

Helpman, E., M. Melitz, and Y. Rubinstein (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” *The Quarterly Journal of Economics*, 123, 441–487. [Cited on page 136.]

Henrekson, M. and M. Persson (2004): “The effects on sick leave of changes in the sickness insurance system,” *Journal of Labor Economics*, 22, 87–113. [Cited on page 58.]

HERRING, C. (2009): "Does Diversity Pay?: Race, Gender, and the Business Case for Diversity," *American Sociological Review*, 74, 208–224. [Cited on page 39.]

HOFFMAN, M. AND S. TADELIS (forthcoming): "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," *Journal of Political Economy*. [Cited on page 6.]

HOLT, P. G. (1998): "Programming for Responsiveness to Environmental Antigens That Trigger Allergic Respiratory Disease in Adulthood Is Initiated during the Perinatal Period," *Environmental Health Perspectives*, 106, 795–800. [Cited on page 166.]

HOSPIDO, L. AND C. SANZ (2019): "Gender Gaps in the Evaluation of Research: Evidence from Submissions to Economics Conferences," Working Paper 1918, Banco de Espana. [Cited on pages 9, 34, and 41.]

HOYNES, H., D. W. SCHANZENBACH, AND D. ALMOND (2016): "Long-run Impacts of Childhood Access to the Safety Net," *American Economic Review*, 106, 903–34. [Cited on page 170.]

ICHINO, A. AND E. MORETTI (2009): "Biological Gender Differences, Absenteeism, and the Earnings Gap," *American Economic Journal: Applied Economics*, 1, 183–218. [Cited on page 9.]

ISEN, A., M. ROSSIN-SLATER, AND W. R. WALKER (2017): "Every Breath You Take - Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970," *Journal of Political Economy*, 125, 848–902. [Cited on pages 158, 159, 161, 164, 166, 171, 174, 215, 216, and 218.]

JAEGER, D. A., T. J. JOYCE, AND R. KAESTNER (2020): "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?" *Journal of Business & Economic Statistics*, 38, 317–326. [Cited on page 170.]

JOHANSSON, P. AND M. PALME (1996): "Do economic incentives affect work absence? Empirical evidence using Swedish micro data," *Journal of Public Economics*, 59, 195–218. [Cited on page 58.]

——— (2005): "Moral hazard and sickness insurance," *Journal of Public Economics*, 89, 1879–1890. [Cited on page 58.]

JUHN, C. AND K. MCCUE (2017): "Specialization Then and Now: Marriage, Children, and the Gender Earnings Gap across Cohorts," *Journal of Economic Perspectives*, 31, 183–204. [Cited on page 18.]

Kahn, S. (1992): "Economic Implications of Public-Sector Comparable Worth: The Case of San Jose, California," *Industrial Relations: A Journal of Economy and Society*, 31, 270–291. [Cited on page 9.]

Karlsson, M. and N. R. Ziebarth (2018): "Population Health Effects and Health-Related Costs of Extreme Temperatures: Comprehensive Evidence from Germany," *Journal of Environmental Economics and Management*, 91, 93–117. [Cited on pages 174, 176, and 218.]

KBA (2017): "Pressemitteilung Nr. 01/2017 - Fahrzeugzulassungen im Dezember 2016 - Jahresbilanz," https://www.kba.de/DE/Presse/Pressemitteilungen/2016_2019/2017/Fahrzeugzulassungen/pm01_2017_n_12_16_pm_komplett.html, accessed: 2020-09-01. [Cited on page 162.]

Khreis, H., C. Kelly, J. Tate, R. Parslow, K. Lucas, and M. Nieuwenhuijsen (2017): "Exposure to Traffic-Related Air Pollution and Risk of Development of Childhood Asthma: A Systematic Review and Meta-Analysis," *Environment International*, 100, 1–31. [Cited on page 186.]

Kitagawa, E. M. (1955): "Components of a Difference Between Two Rates," *Journal of the American Statistical Association*, 50, 1168–1194. [Cited on pages 7 and 18.]

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019a): "Child Penalties across Countries: Evidence and Explanations," *AEA Papers and Proceedings*, 109, 122–126. [Cited on pages 9 and 27.]

Kleven, H., C. Landais, and J. E. Søgaard (2019b): "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11, 181–209. [Cited on pages 9 and 27.]

Knittel, C. R., D. L. Miller, and N. J. Sanders (2016): "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health," *Review of Economics and Statistics*, 98, 350–366. [Cited on pages 53, 161, 162, 172, and 195.]

Krzyżanowski, M., B. Kuna-Dibbert, and J. Schneider (2005): *Health Effects of Transport-Related Air Pollution*, WHO Regional Office Europe. [Cited on page 190.]

Kunze, A. and A. R. Miller (2017): "Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies," *The Review of Economics and Statistics*, 99, 769–775. [Cited on page 10.]

Kurtulus, F. A. and D. Tomaskovic-Devey (2011): "Do Female Top Managers Help Women to Advance? A Panel Study Using EEO-1 Records:," *The Annals of the American Academy of Political and Social Science*. [Cited on page 10.]

LARCH, M. AND Y. V. YOTOV (2016): "General Equilibrium Trade Policy Analysis with Structural Gravity," . [Cited on page 135.]

LAZEAR, E. P., K. L. SHAW, AND C. T. STANTON (2015): "The Value of Bosses," *Journal of Labor Economics*, 33, 823–861. [Cited on pages 6 and 17.]

LEAPE, J. (2006): "The London Congestion Charge," *Journal of Economic Perspectives*, 20, 157–176. [Cited on page 115.]

LEE, D. L., J. MCCRARY, M. J. MOREIRA, AND J. PORTER (2020): "Valid t-ratio Inference for IV," *arXiv preprint arXiv:2010.05058*. [Cited on page 177.]

LEVINE, S. S., E. P. APFELBAUM, M. BERNARD, V. L. BARTELT, E. J. ZAJAC, AND D. STARK (2014): "Ethnic Diversity Deflates Price Bubbles," *Proceedings of the National Academy of Sciences*, 111, 18524–18529. [Cited on page 39.]

LI, N., C. SIOUTAS, A. CHO, D. SCHMITZ, C. MISRA, J. SEMPF, M. WANG, T. OBERLEY, J. FROINES, AND A. NEL (2003): "Ultrafine Particulate Pollutants Induce Oxidative Stress and Mitochondrial Damage." *Environmental Health Perspectives*, 111, 455–460. [Cited on page 174.]

LI, S. (2016): "Better Lucky Than Rich? Welfare Analysis of Automobile License Allocations in Beijing and Shanghai," SSRN Scholarly Paper ID 2349865, Social Science Research Network, Rochester, NY. [Cited on page 115.]

LI, S., Y. LIU, A.-O. PUREVJAV, AND L. YANG (2019): "Does subway expansion improve air quality?" *Journal of Environmental Economics and Management*, 96, 213–235. [Cited on page 115.]

LICHTER, A., N. PESTEL, AND E. SOMMER (2017): "Productivity effects of air pollution: Evidence from professional soccer," *Labour Economics*, 48, 54 – 66. [Cited on page 53.]

LIN, C.-Y. C., W. ZHANG, AND V. I. UMANSKAYA (2011): "The Effects of Driving Restrictions on Air Quality: São Paulo, Bogotá, Beijing, and Tianjin," . [Cited on page 115.]

LINDO, J. M. (2011): "Parental Job Loss and Infant Health," *Journal of Health Economics*, 30, 869–879. [Cited on page 164.]

LIU, W., C. HUANG, Y. HU, Q. FU, Z. ZOU, C. SUN, L. SHEN, X. WANG, J. CAI, J. PAN, Y. HUANG, J. CHANG, Y. SUN, AND J. SUNDELL (2016): "Associations of Gestational and Early Life Exposures to Ambient Air Pollution with Childhood Respiratory Diseases in Shanghai, China: A Retrospective Cohort Study," *Environment International*, 92-93, 284–293. [Cited on page 187.]

Maida, A. and A. Weber (2019): "Female Leadership and Gender Gap within Firms: Evidence from an Italian Board Reform," Discussion Paper 13476, CEPR. [Cited on pages 10 and 44.]

Malina, C. and F. Scheffler (2015): "The Impact of Low Emission Zones on Particulate Matter Concentration and Public Health," *Transportation Research Part A: Policy and Practice*, 77, 372–385. [Cited on page 164.]

Malkiel, B. G. and J. A. Malkiel (1973): "Male-Female Pay Differentials in Professional Employment," *The American Economic Review*, 63, 693–705. [Cited on page 9.]

Malo, M. A., B. Cueto, C. García Serrano, and J. I. Pérez Infante (2012): "La medición del absentismo: Estimaciones desde la perspectiva de las empresas y de las vidas laborales," Technical report, Ministerio de Empleo y Seguridad Social. [Cited on page 64.]

Margaryan, S. (2021): "Low Emission Zones and Population Health," *Journal of Health Economics*, 76, 102402. [Cited on pages 159, 163, 164, and 181.]

Mengel, F., J. Sauermann, and U. Zölitz (2019): "Gender Bias in Teaching Evaluations," *Journal of the European Economic Association*, 17, 535–566. [Cited on pages 10, 34, and 41.]

Mitteldeutsche Zeitung (2011): "Umweltzone startet auch in Halle zum 1. September," 2020-07-09. [Cited on page 164.]

Montt, G. (2018): "Too polluted to work? The gendered correlates of air pollution on hours worked," *IZA Journal of Labor Economics*, 7. [Cited on page 53.]

Moretti, E. and M. Neidell (2011): "Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles," *Journal of Human Resources*, 46, 154–175. [Cited on page 53.]

Nathan, M. and N. Lee (2013): "Cultural Diversity, Innovation, and Entrepreneurship: Firm-Level Evidence from London," *Economic Geography*, 89, 367–394. [Cited on page 39.]

Neidell, M. (2009): "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations," *The Journal of Human Resources*, 44, 450–478. [Cited on page 53.]

——— (2017): "Air pollution and worker productivity," *IZA World of Labor*, 363. [Cited on page 90.]

NEIDELL, M. J. (2004): "Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma," *Journal of Health Economics*, 23, 1209–1236. [Cited on page 112.]

NIEDERLE, M. AND L. VESTERLUND (2007): "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics*, 122, 1067–1101. [Cited on pages 17, 30, and 43.]

NOTOWIDIGDO, M. J. (2020): "The Incidence of Local Labor Demand Shocks," *Journal of Labor Economics*, 38, 687–725. [Cited on page 171.]

OAXACA, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693. [Cited on pages 7 and 18.]

OECD (2019): "National Accounts and Environment Statistics," https://stats.oecd.org/, accessed on January 28, 2019. [Cited on pages 52 and 88.]

OLIVA, P. (2015): "Environmental Regulations and Corruption: Automobile Emissions in Mexico City," *Journal of Political Economy*, 123, 686–724. [Cited on page 115.]

OSAKWE, R. (2010): "An Analysis of the Driving Restriction Implemented in San José, Costa Rica," *Environment for Development Policy Brief.* [Cited on page 115.]

OSTRO, B. D. (1983): "The effects of air pollution on work loss and morbidity," *Journal of Environmental Economics and Management*, 10, 371 – 382. [Cited on pages 51, 53, and 79.]

PARRY, I. W. H., M. WALLS, AND W. HARRINGTON (2007): "Automobile Externalities and Policies," *Journal of Economic Literature*, 45, 373–399. [Cited on page 115.]

PEKKARINEN, T. AND J. VARTIAINEN (2016): "Gender Differences in Promotion on a Job Ladder: Evidence from Finnish Metalworkers:," *ILR Review.* [Cited on page 9.]

PEMA, E. AND S. MEHAY (2010): "The Role of Job Assignment and Human Capital Endowments in Explaining Gender Differences in Job Performance and Promotion," *Labour Economics*, 17, 998–1009. [Cited on page 9.]

PEREZ, L., A. TOBÍAS, X. QUEROL, N. KÜNZLI, J. PEY, A. ALASTUEY, M. VIANA, N. VALERO, M. GONZÁLEZ-CABRÉ, AND J. SUNYER (2008): "Coarse Particles From Saharan Dust and Daily Mortality," *Epidemiology*, 19, 800–807. [Cited on pages 63 and 90.]

PESTEL, N. AND F. WOZNY (2019): "Low Emission Zones for Better Health: Evidence from German Hospitals," *IZA Discussion Paper No. 12545.* [Cited on pages 159, 160, 163, 164, 181, and 188.]

Phillips, K. W., K. A. Liljenquist, and M. A. Neale (2009): "Is the Pain Worth the Gain? The Advantages and Liabilities of Agreeing With Socially Distinct Newcomers," *Personality and Social Psychology Bulletin*, 35, 336–350. [Cited on page 39.]

Piermartini, R. and Y. V. Yotov (2016): "Estimating Trade Policy Effects with Structural Gravity," . [Cited on page 135.]

Quddus, M. A., A. Carmel, and M. G. H. Bell (2007): "The Impact of the Congestion Charge on Retail: The London Experience," *Journal of Transport Economics and Policy*, 41, 113–133. [Cited on page 115.]

Querol, X., A. Alastuey, J. Pey, M. Escudero, S. Castillo, A. Orío, M. Pallarés, S. Jiménez, F. Ferreira, F. Marques, J. Monjardino, E. Cuevas, S. Alonso, B. A. nano, P. Salvador, and J. de la Rosa (2013): "Procedimiento para la identificación de episodios naturales de PM10 y PM2.5 y la demostración de de causa en lo referente a las superaciones del valor límite diario de PM10," Tech. rep., Ministerio de Medio Ambiente, y Medio Rural y Mariona - España. [Cited on page 62.]

Ragan, J. F. and C. H. Tremblay (1988): "Testing for Employee Discrimination by Race and Sex," *The Journal of Human Resources*, 23, 123. [Cited on page 10.]

Ransom, M. and R. L. Oaxaca (2005): "Intrafirm Mobility and Sex Differences in Pay," *ILR Review*, 58, 219–237. [Cited on page 9.]

Rivera, N. M. (2017): "The Effectiveness of Temporary Driving Restrictions: Evidence from Air Pollution, Vehicle Flows, and Mass-Transit Users in Santiago," Tech. Rep. 259182, Agricultural and Applied Economics Association. [Cited on page 115.]

Rohlf, A., F. Holub, N. Koch, and N. Ritter (2020): "The effect of clean air on pharmaceutical expenditures," *Economics Letters*, 192. [Cited on pages 112, 161, and 181.]

Roth, S. (2016): "The Contemporaneous Effect of Indoor Air Pollution on Cognitive Performance: Evidence from the UK," Mimeograph, LSE, London, UK. [Cited on page 53.]

Rothstein, D. S. (1997): "Early Career Supervisor Gender and the Labor Market Outcomes of Young Worker," in *Gender and Family Issues in the Workplace*, ed. by F. D. Blau and R. G. Ehrenberg, Russell Sage Foundation. [Cited on page 10.]

Royal Netherlands Meteorological Institute (2019): "European Climate Assessment and Dataset," https://www.ecad.eu/, accessed July 2, 2019. [Cited on page 65.]

Sager, L. (2019): "Estimating the Effect of Air Pollution on Road Safety Using Atmospheric Temperature Inversions," *Journal of Environmental Economics and Management*, 98, 102250. [Cited on page 172.]

Salas, C. (2010): "Evaluating Public Policies with High Frequency Data: Evidence for Driving Restrictions in Mexico City Revisited," Tech. Rep. 374, Instituto de Economia. Pontificia Universidad Católica de Chile. [Cited on page 115.]

Sanders, N. J. (2012): "What Doesn't Kill You Makes You Weaker: Prenatal Pollution Exposure and Educational Outcomes," *Journal of Human Resources*, 47, 826–850. [Cited on page 161.]

Sanders, N. J. and C. Stoecker (2015): "Where Have All the Young Men Gone? Using Sex Ratios to Measure Fetal Death Rates," *Journal of Health Economics*, 41, 30–45. [Cited on page 161.]

Santos Silva, J. and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88, 641–658. [Cited on page 142.]

Sarsons, H. (2018): "Interpreting Signals in the Labor Market: Evidence from Medical Referrals," *mimeo*. [Cited on pages 5 and 9.]

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (forthcoming): "Gender Differences in Recognition for Group Work," *Journal of Political Economy*. [Cited on page 9.]

Schlenker, W. and W. R. Walker (2016): "Airports, Air Pollution, and Contemporaneous Health," *Review of Economic Studies*, 83, 768–809. [Cited on pages 51 and 53.]

Shapiro, J. M. (2006): "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital," *The Review of Economics and Statistics*, 88, 324–335. [Cited on page 171.]

Shapiro, J. S. and R. Walker (2018): "Why is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade," *American Economic Review*, 108, 3814–54. [Cited on page 157.]

Simeonova, E., J. Currie, P. Nilsson, and R. Walker (2019): "Congestion Pricing, Air Pollution and Children's Health," *Journal of Human Resources*, 0218–9363R2. [Cited on pages 112, 115, 161, and 187.]

Sorensen, E. (1986): "Implementing Comparable Worth: A Survey of Recent Job Evaluation Studies," *The American Economic Review*, 76, 364–367. [Cited on page 9.]

Spanish Center for Disease Control, Instituto de Salud Carlos III (2016): "Sistema Centinela de Vigilancia de la Gripe en España," . [Cited on page 66.]

Spanish Ministry for the Ecological Transition (2016): "Air Quality Data 2001-2016," https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/calidad-del-aire/evaluacion-datos/datos/, accessed July 2, 2019. [Cited on page 64.]

——— (2018): "Datos suministrados en el marco del encargo del Ministerio para la Transición Ecológica al Consejo Superior de Investigaciones Científicas (CSIC) para la detección de episodios naturales de aportes transfronterizos de partículas y otras fuentes de contaminación de material particulado, y de formación de ozono troposférico," https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/calidad-del-aire/evaluacion-datos/fuentes-naturales/anuales.aspx, accessed July 2, 2019. [Cited on pages 65 and 103.]

Spanish Ministry of Employment, Migration and Social Security (2018): "Muestra Continua de Vidas Laborales," http://www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/EST211. [Cited on pages 64, 104, and 105.]

Šrám, R. J., B. Binková, J. Dejmek, and M. Bobak (2005): "Ambient Air Pollution and Pregnancy Outcomes: a Review of the Literature," *Environmental Health Perspectives*, 113, 375–382. [Cited on page 166.]

Stafoggia, M., S. Zauli-Sajani, J. Pey, E. Samoli, E. Alessandrini, X. Basagaña, A. Cernigliaro, M. Chiusolo, M. Demaria, J. Díaz, A. Faustini, K. Katsouyanni, A. Kelessis, C. Linares, S. Marchesi, S. Medina, P. Pandolfi, N. Pérez, X. Querol, G. Randi, A. Ranzi, A. Tobias, F. Forastiere, and MED-PARTICLES Study Group (2016): "Desert dust outbreaks in Southern Europe: contribution to daily PM_{10} concentrations and short-term associations with mortality and hospital admissions," *Environmental Health Perspectives*, 124, 413–419. [Cited on pages 63 and 90.]

Swart, E., P. Ihle, H. Gothe, and D. Matusiewicz (2005): *Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven*, Huber, 2 ed. [Cited on page 175.]

UBA (2017): "Stickoxid-Belastung durch Diesel-Pkw noch höher als gedacht," https://www.umweltbundesamt.de/presse/pressemitteilungen/stickoxid-belastung-durch-diesel-pkw-noch-hoeher, accessed: 2019-01. [Cited on page 195.]

——— (2019): "Luftmessnetz: Wo und Wie wird Gemessen?" https://www.umweltbundesamt.de/themen/luftmessnetz-wo-wie-wird-gemessen, accessed: 2019-01. [Cited on page 174.]

U.S. DOT (2015): "Diesel-powered Passenger Cars and Light Trucks," https://www.bts.gov/archive/publications/bts_fact_sheets/oct_2015/entire, accessed: 2020-09. [Cited on page 162.]

van Donkelaar, A., R. V. Martin, C. Li, and R. T. Burnett (2019): "Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors," *Environmental Science & Technology*, 53, 2595–2611. [Cited on page 194.]

Viard, V. and S. Fu (2015): "The effect of Beijing's driving restrictions on pollution and economic activity," *Journal of Public Economics*, 125, 98–115. [Cited on pages 114 and 115.]

WHO (2006): *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide*, Geneva, Switzerland: World Health Organization. [Cited on page 67.]

Wiesbadener Tagblatt (2011): "Stadt muss Umweltzone einführen;," 2011-10-11. [Cited on page 164.]

Williams, A. M. and D. J. Phaneuf (2019): "The Morbidity Costs of Air Pollution: Evidence from Spending on Chronic Respiratory Conditions," *Environmental and Resource Economics*, 1–33. [Cited on pages 159 and 162.]

Wolff, H. (2014): "Keep Your Clunker in the Suburb: Low-Emission Zones and Adoption of Green Vehicles," *The Economic Journal*, 124, F481–F512. [Cited on pages 115, 159, 164, 189, 191, and 192.]

Yates, P. J., W. H. Williams, A. Harris, A. Round, and R. Jenkins (2006): "An Epidemiological Study of Head Injuries in a UK Population Attending an Emergency Department," *Journal of Neurology, Neurosurgery and Psychiatry*, 77, 699–701. [Cited on page 193.]

Ye, J. (2017): "Better safe than sorry? Evidence from Lanzhou's driving restriction policy," *China Economic Review*, 45, 1–21. [Cited on page 115.]

Zhang, W., C.-Y. C. Lin Lawell, and V. I. Umanskaya (2017): "The effects of license plate-based driving restrictions on air quality: Theory and empirical evidence," *Journal of Environmental Economics and Management*, 82, 181–220. [Cited on page 115.]

Zhong, N., J. Cao, and Y. Wang (2017): "Traffic Congestion, Ambient Air Pollution, and Health: Evidence from Driving Restrictions in Beijing," *Journal of the Association of Environmental and Resource Economists*, 4, 821–856. [Cited on pages 112 and 115.]