Università Commerciale "Luigi Bocconi"
PhD School

PhD program in: Public Policy & Administration
Cycle: 33rd
Disciplinary Field (code): SECS-P/06

Essays on Environmental Economics and Policy

Advisor: Valentina Bosetti

PhD Thesis by
Francesco Granella
ID number: 1531383

Year 2023

ESSAYS ON ENVIRONMENTAL ECONOMICS AND POLICY

FRANCESCO GRANELLA

Supervisor: Valentina Bosetti
PhD in Public Policy & Administration, 33rd Cycle
Department of Social and Political Sciences
Università Commerciale "Luigi Bocconi"
Disciplinary Field: SECS-P/06

Francesco Granella - ID number 1531383

CHAPTERS

DETAILED CONTENTS

FOREWORD & ACKNOWLEDGMENTS

700.000 years ago, in the Northwest of Italy, longtime extinct glaciers have carved through the high Alps and left behind smooth rocks and an amphitheater of hills. In the parterre of that opera house I grew up three decades ago, a spectator to the mountains of the Aosta Valley and the Dora river flowing out of it. Those waters, turquoise from sediments of granite and schist, have unexpectedly left a mark on my perspective on humans and the environment. Those mountains - a stark contrast to the gentle hills and flat agricultural land around a young me - would become, in the most literal way, my compass.

FROM THE WATER. Fall in Northwest Italy always brings abundant precipitations and breaks the dry stretches of summer. As the season progresses, the snow pack starts thickening with fresh flakes up high and rain drenches the soil. There is a period, typically in between the months of October and November, when precipitation are rich, and temperatures high enough to limit the accumulation of snow; water flows down the mountainside and in rivers. Hydro-geological risk peaks: rivers risk overflowing, landslides may occur.

In the autumn of 2000, an exceptional cyclonic system dropped 500mm of rain in few hours. The Dora river, as many other in the region, overflew its banks and flooded the city of Ivrea, menacing the millenia-old roman bridge. The villaged where I lived went underwater. While my house suffered only minor damages, the entire surroundings seemed to me, 11 at the time, a bittersweet mix of calm and chaos, silence and worry. Water reached ceiling levels at a friend's house.

In the years that followed, banks where reinforced and new infrastructure projects appeared with the intent to prevent another catastrophe. The floods silently changed my understanding of the environment. They did not leave me with the idea of dangerous nature we ought to protect ourselves from. On the contrary, it became clear that flood damages where the consequence of the urban geography and of the management of, in this case, hydraulic risks.

A decade and a half later I was writing the thesis for my MSc on the long-term consequences of another major flood. I thank my mentor Valentina Bosetti for believing in that sprout of research in environmental economics and cultivating it.

FROM THE MOUNTAIN Where I grew up, it is easy to orient oneself: North is where the mountains are. Two in particular, named Cavallaria and Mombarone, are the most prominent and most visible landmark.

Growing up, I could easily find the way with some mental trigonometry keeping Cavallaria and Mombarone as reference point. At the age of 13 or 14 I once travelled to the Champagne region in France, where the land is flat and the sky high and grey, and found myself - to my amusement - that I was lost and had no clue where North was. Long before I realized it, my eyes were looking for mountain-like shapes to help me navigate.

This dissertation and been a long journey with intense navigation. If I have not lost direction is thanks to many people who have been with me along the way or have shaped me as the person I am. Thank you, you are my mountains and my reference point.

A special mention goes - again, because one is not enough - to my mentor Valentina Bosetti. Thank you for believing in me.

I am grateful for having crossed paths with incredible human beings, for short or long times, and I appreciate every single moment we've spent together. A special thanks to Nicolo' Dalvit, a great friend and distiller of precious advice from econometrics to perspectives on life.

A journey in good company is a good journey. I am grateful for having shared the PhD one with amazing individuals as Benedetta, Chiara, Samir, Selin and Rita. I have learned so much you.

I have found a thriving intellectual community and the sense of family in the researchers at the RFF-CMCC European Institute on Economics and the Environment. I am a better researcher today thanks to them and thanks to the many co-authors I have worked with, in particular Alexandros Cavgias, Lara Aleluia Reis, Bernardo Bastien-Olvera and Frances Moore.

As an upside of the COVID-19 lockdowns, I have found a good friend in Claudio Brenna. Social isolation has been less bitter because of you.

Thanks to my mountaineering partners I shared rope and skin tracks with. A special thought goes to Giovanni.

To Deda e Bebo. For your boundless love.

Thanks to Paula Rettl, a loyal companion for many years and constant source of moral and intellectual inspiration.

Who could I be, who would I be, without my family? Words are scarce and cannot describe how much of what I am is because of them.

INTRODUCTION

This dissertation is concerned with the impacts of environmental stressors on people's health and welfare. In particular, it focuses on air pollution and temperatures anomalies. Air pollution is considered the fifth leading mortality risk factor worldwide (Cohen et al., 2017). Despite impressive improvements in air quality over the last half-century, air pollution remains a global challenge, especially in rapidly urbanizing countries. On an even larger scale, anthropocentric emissions of greenhouse gases have been pushing a shift in the world's climate that is projected to widen in the coming decades. The ability of economies to cope with changing temperatures is paramount to limiting the damages from climate change. The three chapters of this thesis should be read in the framework set by these challenges. In each chapter, a specific question is brought to the surface, and the road to address it is outlined.

Chapter 1 investigates the negative effect of air pollution on physical ability. A large share of the world's population is employed in manual labor. Yet, our understanding of the productivity cots of air pollution for physically intense work remains limited. The chapter identifies in track and field competitions a natural experiment where cognition plays a minor role. Combining half a million competition results with weather and air quality data, it estimates the change in physical performance induced by variations in air pollution.

Chapter 2 considers the methodological tools available to estimate the causal change in air pollution concentrations following a reduction in emissions. It recognizes the challenges to identification posed by fluctuations and trends in atmospheric conditions and proposes a machine learning approach to address them. The chapter then applies this strategy to quantify the reduction in pollution and related health benefits induced by the COVID-19 lockdown of Lombardy, Italy, in spring 2020. This work is a joint effort with Lara Aleluia Reis (RFF-CMCC European Institute on Economics and the Environment), Valentina Bosetti (Bocconi University), and Massimo Tavoni (Politecnico di Milano). The chapter has been published under the same title on Environmental Research Letters.

Chapter 3 is concerned with the persistence of the effects of temperature anomalies on economic growth. If an adverse temperature shock damages the determinants of economic growth, we can expect losses from climate change - a permanent shift in the mean temperature - to be cumulative over time and, therefore, very costly. Despite the primary importance of this question for modeling the climate-economy interactions, data constraints and data-hungry approaches have led to inconclusive answers. This chapter presents a new and more efficient method to test for the persistence of effects; using three different GDP datasets, evidence emerges that temperature effects are indeed persistent. This chapter has been the output of joint work with Bernardo A. Bastien-Olvera and Frances C. Moore of the University of California at Davis, and has been published under the same title on Environmental Research Letters.

1 HETEROGENEOUS EFFECTS OF AIR POLLUTION ON PHYSICAL TASKS: EVIDENCE FROM AMATEUR TRACK AND FIELD

ABSTRACT Although a large share of the world's population is employed in manual labor, our understanding of the productivity costs of air pollution for physically intense work remains limited. This paper estimates the effect of fine particulate matter (PM 2.5) on purely physical tasks by analyzing half a million amateur track and field competition results, a setting where cognition plays a minor role. Exploiting the panel nature of the data and high dimensional fixed effects, I find that a 10 $\mu g/m^3$ increase in PM 2.5 reduces performance by 1% of a standard deviation. The effect grows with the duration of effort, indicating that productivity losses may be larger for occupations requiring low-intensity and sustained effort, such as construction workers.

1.1 INTRODUCTION

A large share of the world's population is hired in manual labor, in both developed and developing economies. At the same time, air pollution is pervasive throughout the globe, often at unhealthy concentrations. In urban centers, blue-collar workers can be exposed to high concentrations of industrial pollution as PM 2.5 easily penetrates indoor for its small diameter (J. He, H. Liu, and Salvo, 2019); in rural areas, unregulated biomass burning is a significant source of harmful airborne pollutants (Rangel and Vogl, 2018; Graff Zivin, T. Liu, et al., 2020; G. He, T. Liu, and Zhou, 2020).

While a growing number of studies find that fine particulate matter (PM 2.5) reduces cognitive performance[1], to date the evidence on the causal effects of ambient air pollution on the physical component of tasks remains limited.[2]

This paper estimates the effects of PM 2.5 on physical tasks where cognition plays a marginal role and task content is easily codified. I assemble a dataset on the universe of track and field competitions held in Italy from 2005 to 2019 and match individual performances with air pollution data. In this environment, young individuals repeatedly perform highly standardized tasks (running, jumping, throwing) under varying environmental conditions. Leveraging the panel structure of the data, I estimate the effect PM 2.5 on performance using a set of high dimensional fixed effects.

I find that an increase in PM 2.5 of 10 $\mu g/m^3$ reduces performance by 1% of a standard deviation, equivalent to a loss of one-third of a percentile in nationwide rankings. Conversely, ozone does not have a discernible effect when accounting for concentrations of particulate matter. The detrimental consequences of PM 2.5 on performance appear at medium level of concentrations between 25 and 50 $\mu g/m^3$. The overall effect is the same for males and females, and greater for high-ability athletes. While the data comes from competitions held in Italy, there are no apparent reasons for the link between air pollution and athlete performance to be specific to the Italian context.

---

1 Evidence covers standardized tasks such high-stake exams (Ebenstein, Lavy, and Roth, 2016; Persico and Venator, 2019; Graff Zivin, T. Liu, et al., 2020) cognitive tests (Bedi et al., 2021), brain games (Nauze and Severnini, 2021), chess matches (Künn, Palacios, and Pestel, 2019), and referee calls in baseball games (Archsmith, Heyes, and Saberian, 2018). Air pollution, in particular PM 2.5, has also been found to interfere with decision making, more broadly defined. For instance, Burkhardt et al., 2019 and Bondy, Roth, and Sager, 2020 find that PM 2.5 increases violent crimes, but not property crimes. Heyes, Neidell, and Saberian, 2016 link increases in PM 2.5 in Manhattan with reduced returns in the New York Stock Exchange. Chen, 2019 provides a detailed summary of the physiological and psychological pathways through which pollution is believed to affect cognitive performance and behavior.

2 Beyond direct productivity losses, exposure to PM 2.5 reduces life expectancy. It is estimated the fifth leading cause of premature mortality worldwide (Cohen et al., 2017).

I then investigate the effect of PM 2.5 by type of physical requirement to inform the productivity losses of common occupations. Short-lasting competitions require explosive strength, whereas longer races require stamina and are more dependent on the pulmonary and cardiovascular systems, which, following the medical literature, bear most of the effects of PM 2.5 (Pope and Dockery, 2006).[3] In line with expectations, I find larger impacts of PM 2.5 on the performance of longer-lasting races. The results suggest that jobs requiring prolonged physical effort incur, under the same conditions, greater productivity losses than jobs requiring short bursts of intense exercise. Most athletes in the sample are below working age, thus the generalization of results to manual laborers should be done with care. Yet, to an extent, results can be informative for jobs with prolonged efforts: for instance, according to the Bureau of Labor Statistics, in the United States stamina is an important ability for as many as 8.5 million workers in the country (Table 9). Explosive strength is important for half a million workers.[4] More generally, 13.7% of all civilian jobs and 45.5% of jobs in Construction and Extraction require heavy work (Table 10) (U.S. Bureau of Labor and Statistics).

This paper enters the stream of literature on the consequences of air pollution on labor productivity. Several works have found that PM 2.5 hampers productivity in a variety of settings where physical effort is important, such as in a packing plant (Chang et al., 2016) or in textile and garment plants (Adhvaryu, Kala, and Nyshadham, 2018; J. He, H. Liu, and Salvo, 2019). Their external validity is however limited by the narrow scope. One exception is offered byFu, Viard, and Zhang, 2021, who expand previous work and estimate nationwide effects on short-run productivity for China's manufacturing sector. They report suggestive evidence that PM 2.5 reduce both cognitive and physical productivity. Nevertheless, generalization of results remain challenging without an understanding of the mechanisms at work. Unpacking the causal links is still work in progress. While the evidence on cognitive effects is growing (*e.g.* Ebenstein, Lavy, and Roth, 2016; Graff Zivin, T. Liu, et al., 2020; Bedi et al., 2021; Carneiro, Cole, and Strobl, 2021; Nauze and Severnini, 2021), less is known about the impacts on physical productivity.

Sports have proven attractive grounds for economists to do research (Kahn, 2000) for their richness of data and, in particular, the availability of readily-observed productivity measures. As Archsmith, Heyes, and

3 The Occupational Information Network (O*NET) of the U.S. Bureau of Labor Statistics describes the abilities, defined as "enduring attributes of the individual that influence performance", required by 923 occupations, quantifying the importance and level required of each ability. It defines explosive strength as "the ability to use short bursts of muscle force to propel oneself (as in jumping or sprinting), or to throw an object". Stamina is defined as "The ability to exert yourself physically over long periods of time without getting winded or out of breath".

4 Important is here defined as a 3 or more on a 1-5 scale from 'Not Important'(1) to 'Extremely Important' (5).

Saberian, 2018 puts it, sports provide a "microcosm for things that might be happening more broadly in society", context rather than an object. For instance, several sports contexts have been used to link air pollution to productivity, providing new understanding of the mechanisms depending on the context, method, and analysis. Data on distance running has been employed to study peer effects (Emerson and B. Hill, 2018) and the gender gap in competitiveness (Frick, 2011). In the literature covering the productivity effects of air pollution, most papers have used data from professional athletes, for whom data is more abundant. Archsmith, Heyes, and Saberian, 2018 finds that baseball umpires are more likely to make incorrect calls when exposed to higher CO and PM 2.5. Lichter, Pestel, and Sommer, 2017 find that the productivity of football players is hampered by particulate matter.

Closest works to this paper are Marcus, 2021, Austin, Heutel, and Kreisman, 2019, and Mullins, 2018. Marcus, 2021 studies the link between ozone and cardiopulmonary performance of school children aged 10 to 15, assessed yearly by the California Department of Education measured with a test of aerobic capacity. She finds that an increase in ozone from 0-25% of the U.S. National Ambient Air Quality Standards to levels above the safety standard increases the share of students with poor aerobic capacity by 5.4 percentage points. Austin, Heutel, and Kreisman, 2019 estimate the changes in cardiopulmonary fitness of students aged 8 to 14 in Georgia, USA, induced by the retrofitting of diesel school buses and the subsequent improvement in air quality inside the vehicles. They find that school district-average VO_2max would improve by 4% if a district retrofitted 100% of its fleet.[5]

Mullins, 2018 estimates the effects of ground ozone, a pollutant often linked to heat and sunlight, on the performance of US collegiate track and field athletes.[6] In contrast to Mullins, 2018, this paper assesses the impacts of PM 2.5, a pollutant that can penetrate indoor and is the fifth leading cause of premature death worldwide due to a combination of near-ubiquity and harming potential (Cohen et al., 2017).[7] In addition, I consider a population of mostly amateurs, whose team membership or income (such as scholarship) are not tied to performance, diluting concerns of positive self-selection on fitness.[8]

5 VO_2max is the maximum rate that oxygen can be taken into and used by the body during exercise (A. V. Hill and Lupton, 1923).

6 Sexton, Wang, and Mullins, 2021 use the same data as Mullins, 2018 to study effect of heat on physical performance.

7 In a robustness check, Mullins, 2018 controls for multiple pollutants, including both coarse and fine particulate matter (PM 10 and PM 2.5), whose coefficient are not statistically significant. However, the latter are correlated as PM 10 is a superset of finer PM 2.5. Including both in the regression, Mullins, 2018 ensures that the coefficient for ozone is not driven by particulate matter. On the other hand, it cannot be ascertained whether the lack of significance for PM 2.5 is explained by lack of causality, or standard errors inflated by the correlation with PM 10.

8 With the exception of a few elite athletes, who are hired for a moderate stipend.

The next section discusses the main characteristics of track and field competitions that are relevant to this study. Section 3.2 describes the data, and Section 1.4 presents the empirical strategy. Section 1.5 discusses results, and Section 1.6 the robustness checks. Section 1.7 concludes.

1.2 TRACK AND FIELD COMPETITIONS AS STANDARDIZED PHYSICAL TESTS

The use of sports data in economics is not novel (Kahn, 2000). Beyond the contributions of Mullins, 2018 discussed above, the most similar work in this regard is Lichter, Pestel, and Sommer, 2017, who quantify the effect of PM 10 on the productivity of professional soccer players, measured as the number of passes per match. However, the interaction of team strategies and individual responses does not allow for separating physical effects from behavioral responses, although they provide suggestive evidence that both factors are at work.[9] Productivity spillovers between players further complicate the attribution of individual productivity (Arcidiacono, Kinsler, and Price, 2017).

Ideally, a researcher could retrieve a pollution-physical productivity function asking subjects to perform a measurable and standardized task at randomly supplied pollution levels. Such an experiment would, however, raise ethical concerns of primary importance.

Track and field is a set of individual sports disciplines that require running, jumping, or throwing in a very standardized setting. Competitions are held on a stadium track, or its inner field, whose characteristics are regulated in detail by international standards (World Athletics, 2019). As an illustration, the inside lane of a running track must be 400 meters long, and each lane must be 1.22m $\pm$ 0.01m wide; equipment, such as hurdles and throwing implements, must respect standards of shape and weight (World Athletics, 2020). Performance of all track events (foot races) are measured electronically, whereas all field events (jumps and throws) are measured manually yet precisely. While regular competitions can be held in indoor tracks, this study is restricted to outdoor contests as air quality in indoor tracks can be worsened in unmeasurable ways by the smoke of starting guns and can differ substantially from outdoor conditions. Road competitions such as marathons are excluded from this study as they take place on non-standardized race courses.[10]

9 Track and field competitions differ from road races as the former take place in standardized stadiums while the latter on unstandardized road courses. Guo and Fu, 2019 find a negative effect of air pollution on the performance of marathon runners in races events in China. However, and self-selection out of a marathon, before or during the race, makes causal identification challenging.

10 The setting and design of road competitions make good environment to address different questions, such as peer effects in productivity (Emerson and B. Hill, 2018).

The cognitive efforts in track and field events are minimal. First, athletes compete individually, irrespective of the performance of other team members.[11] Second, they typically compete in running the fastest, jumping the longest or highest, and throwing the farthest. A notable exception is mid- and long-distance races. In conditions where victory is more important than timing, the stronger athletes might strategically slow down the race pace if they believe they have an edge in a closing sprint. These conditions are most common at the end of the sports season when peak events are held; throughout the season, strategic races are comparatively less common as athletes chase qualifying timings for championships of varying degrees. Section 1.6 shows that results are not driven by mid- and long-distance events.

Males and females are usually equally represented and perform the same or very similar tasks (see Figure 7 in Appendix for a breakdown of types of competitions by gender). This contrasts with other occupational contexts in the literature on pollution and physical performance. For instance, among agricultural laborers studied by Graff Zivin and Neidell, 2012 women are more likely to harvest crops that require less energy. The textile workers examined by J. He, H. Liu, and Salvo, 2019 are predominantly females.

In Italy, track and field competitions are supervised by the Italian Athletics Federation (FIDAL), which guarantees the uniformity and the validity of results through its referees. Athletes are members of clubs, whose catchment area is typically local and are independent of the school system. Entry barriers into the sport are very low, and competitions are comparably accessible across socio-economics backgrounds. However, it should be noted that the average age of track and field competitors is low, in the teens. Individuals positively select into the sport, but conditional on being in the sport, selection into competing is small.

1.3 DATA

1.3.1 Track and field

The analysis uses data on the universe of regular track and field competitions held in Italy from 2005 to 2019. Results are systematically collected by FIDAL in near real-time and are made available on its website.[12] Most outdoor competitions take place from April to September.

Race distances and equipment vary with category and gender to accommodate for physiological differences. For example, the 100-meter

---

11 With the exception of *relay* runs, in which each member of a team runs part of the race. Relays are excluded from this study.

12 Data have been scraped from the FIDAL website at `http://www.fidal.it/`.

dash is typically not run until 16 years old; the equivalent competition for a 14-year old is the 80-meter dash. To ensure comparability across events, age categories, and gender, results are transformed into a standardized score. For every event, years of age, and gender, I trim the top 99 and bottom 1 percent to exclude outliers, then demean and divide by the group standard deviation. The objective of field events is to jump or throw the farthest, whereas in races athletes aim for the shortest time. Therefore, the standardized result of races is reversed in sign so that greater values reflect greater performance for both jumps, throws, and races. The dependent variable is constructed as

$$\widetilde{Y}_{i,age(t),event,gender(i)} = \tag{1}$$
$$= \frac{Y_{i,age(t),event,gender(i)} - \mu_{age(t),event,gender(i)}}{\sigma_{age(t),event,gender(i)}} \cdot \text{Event type}_{event}$$

where $Y_{i,age(t),event,gender(i)}$ is the performance of athlete i on day t on event $event$. $\mu_{age(t),event,gender(i)}$ and $\sigma_{age(t),event,gender(i)}$ are the mean and standard deviation of results in groups defined by age, event and gender.[13] $\text{Event type}_{event}$ is equal to 1 for jumps and throws (field events), to -1 for races (track events).

The standardization leads to a straightforward interpretation of regression results: a change in standardized score $\widetilde{Y}$ is equivalent to a change in unstandardized result Y as percent of a standard deviation in the reference group:

$$\Delta\widetilde{Y}_{i,age(t),event,gender(i)} = \frac{\Delta Y_{i,age(t),event,gender(i)}}{\sigma_{age(t),event,gender(i)}}.$$

FIDAL only records information on the city in which races have been held, though not on the location of the stadium. However, it maintains a geo-localized database of track stadiums in Italy. To precisely assign pollution readings to race days, I assign to each city, whenever possible, the geographic coordinates of track stadiums. In case a city contains more than one stadium, and it is impossible to assign results to a specific one, that city is excluded. Thus, a few stadiums are excluded from the sample. [14] The location of municipalities with track and field events in the final dataset is shown in Figure 1.

The result is an unbalanced panel of 95336 athletes, for more than half a million competition results in 3555 stadium-race days in 137 stadiums. Given the disproportionately large number of young athletes, the average age is 15.2, and about 90% of them take part in 59 competitions or fewer during the period and cities covered by the

13 In a comparable setting, Mullins, 2018 standardize results with respect to world records. However, world records do not exist for many events in which younger athletes participate.

14 The pollution monitoring network is denser in the more polluted and populated North (Figure 6).

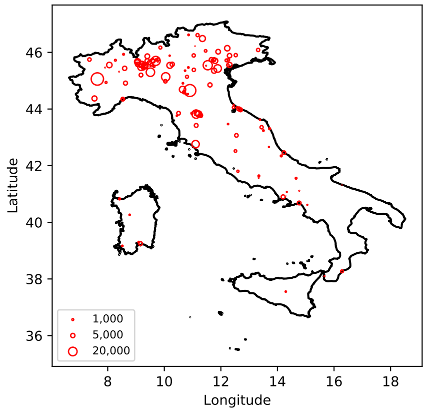


Figure 1: Location of Italian track and field stadiums in the data. Circle size indicates the amount of observations per each stadium.

database.[15] About half of the events are races, 27% are jumps, 23% are throws. Female athletes make up 48% of the sample (Table 1).

1.3.2 Pollution

Daily pollution readings of PM 2.5 and ozone measured at monitoring stations come from AirBase, the European air quality database maintained by the European Environment Agency. Where hourly readings are available, a daily measure of PM 2.5 is constructed as the average of hourly measures from 10 AM to 6 PM, as track and field competitions take place mostly during the afternoon. The maximum reading is used instead for ozone. For every race day, PM 2.5 and ozone readings from monitoring stations within 10 kilometers are interpolated at track stadiums with inverse distance weighting. Hence, pollution in the data varies by stadium and day.

A considerable share of the Italian population is exposed to harmful levels of air pollution. According to the European Environment Agency, 75% of the urban population in Italy was exposed to concentrations of PM 2.5 above EU standards (Ortiz, 2020). The more densely populated Northern regions are some of the most polluted regions in OECD countries. However, track and field competitions take place mostly from April to September, when concentrations are lowest. The average PM 2.5 concentration in the data is 14.4 $\mu g/m^3$, and surpasses the EU annual limit value of 25 $\mu g/m^3$ in about 9% of observations (Figure 2).

15 Data for a large number of athletes aged 35 and older had to be discarded for lacking a precise date of birth.

Table 1: Descriptive statistics

	Mean	Std. Dev.	Median	Minimum	Maximum	N.
Std result	0.03	0.98	0.09	-5.66	3.39	553,171
PM 2.5	14.35	8.36	13.00	0.00	147.04	553,171
Ozone	108.31	28.36	106.40	7.00	247.45	509,494
Female	0.48	0.50	0.00	0.00	1.00	553,171
Age	15.24	3.45	14.78	5.30	40.37	553,171
Temp. max	23.94	5.06	24.32	5.64	37.52	553,171
Precipitation	2.43	5.21	0.19	0.00	48.49	553,171
Wind	2.04	0.76	1.95	0.13	7.95	553,171
Wind assist	0.02	0.55	0.00	-7.50	8.20	553,171
Duration, minutes	0.84	2.35	0.00	0.00	29.70	553,171

Note: Standardized competition results *Std result* are defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. PM 2.5 and ozone are expressed in $\mu g/m^3$; temperature in degree Celsius; precipitation in millimeters; wind in m/s.

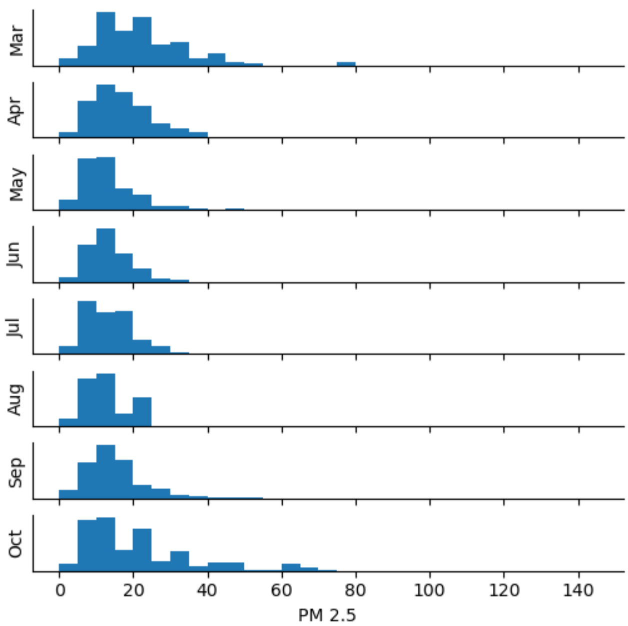


Figure 2: Within-month distribution of PM 2.5. Most competitions occur from April to September, when concentrations of PM 2.5 are lower.

1.3.3 Weather data

The performance of track and field athletes is sensitive to environmental conditions beyond air pollution, such as temperature, relative humidity, precipitation and wind. At the same time, atmospheric conditions are key to the process of pollution formation, transport and dispersal.

I combine performance data and pollution readings with atmospheric conditions from ERA5-Land hourly reanalysis data on a 0.1° by 0.1° grid (Copernicus Climate Change Service, 2019). I construct measures of mean temperature, total precipitation, mean wind speed, and mean relative humidity from 12 PM to 9 PM. Like air pollution, weather conditions are interpolated at stadiums with inverse distance weighting.

Performance in a number of events is particularly susceptible to the wind blowing in favor or against the direction of an athlete.[16] International standards mandate that results in these events cannot be valid as a record on any level if the tailwind exceeds 2 m/s. However, results are still valid for establishing rankings within the competitions. Thus, wind speed during such events is measured inside the stadium with anemometers and recorded with individual results. It can take positive values (tailwind) or negative ones (headwind). For all other events, the variable is set to zero. To distinguish it from the meteorological wind described above, I will refer to this variable as *wind assist*.

1.4 EMPIRICAL STRATEGY

The richness of the data allows identifying the effects of PM 2.5 on track and field competitions using a high-dimensional set of fixed effects. First, I exploit the panel nature of the data and include individual fixed effects. Athletes compete multiple times at varying environmental conditions throughout their career. The analysis relies on variation in performance and air pollution within individuals.

Second, to adjust for the confounding role of atmospheric conditions, I introduce a flexible specification of weather variables. Controls include wind assist, fixed effects for 2° C bins of maximum temperature and their interaction with wind speed, relative humidity, and binned precipitation.[17]

---

16 Namely: races until 200 meters of length, the triple jump and the long jump. The benefit or burden of wind blowing is clear in events where the athlete moves in one direction. When races involve running one or more laps of a track, a stable wind blows cyclically both in favor and against athletes.

17 Temperature bins at extreme temperatures, with fewer observations, are wider. Bins are constructed as: (0 10], (10, 14], (14, 16], (16, 18], (18, 20], (20, 22], (22, 24], (24, 26], (26, 28], (28, 30], (30, 32], (32, 34], (34, 36], (36, 40]. Cumulative precipitation in

Third, concentrations of PM 2.5 are lowest during summer, when the most important competitions are held and the sport season peaks. The relationship between PM 2.5 and performance might be downward biased unless the two trends are accounted for. For this reason, all specifications include fixed effects for year, week, and day-of-the-week.

Finally, stadiums and their locations may correlate in unobserved ways with performance and pollution levels. A large city might host high-level competitions and suffer from high levels of pollution, for instance. I include stadium fixed effects to account for stadiums' constant characteristics, their surroundings, or the competitions they host. Athletes can travel to other cities to compete. I interact stadium fixed effects with fixed effects for athletes' team, a proxy for the city of origin. Given that Italian track and field teams are predominantly local, the interactions capture changes in performance caused by traveling from the team's home city to the stadium, and any potential home advantage. Two thirds of the overall variation in PM 2.5 and half of the variation in performance come from within individuals-stadiums cells, further reducing the risk of confounding effects caused by traveling (Figure 11 in Appendix) .

Most competitions take place in warm months, when solar radiation accelerates chemical reactions to form ozone, a pollutant known to irritate lung airways and increase respiratory problems (Neidell, 2009), and reduce aerobic capacity (Mullins, 2018; Marcus, 2021). Given the negative temporal correlation with PM 2.5, omitting ozone from Equation 2 may lead to underestimation of the true effect of PM 2.5 on performance. All specifications adjust for concentrations of ozone.

The baseline specification then looks like:

$$\begin{aligned}\tilde{Y}_{i,s,t} =& \beta_1 PM2.5_{s,t} + \beta_2 Ozone_{s,t} + Time_t'\gamma_1 + \\ &+ Weather_{t,s}'\gamma_2 + \gamma_3 Wind\ assist_{i,s,t} \\ &+ \alpha_i + S_s + C_{c(i,t)} + S * C_{s,c(i,t)} + \epsilon_{i,s,t}.\end{aligned} \tag{2}$$

The dependent variable $\tilde{Y}_{i,s,t}$ is the standardized results described in Equation 1. Subscript i, s, and t respectively index individuals, stadiums, and time. For ease of notation, I omit subscripts indexing different competitions of the same individual on the same day.[18] The main parameter of interest is β_1. The vector $Time_t$ contains time-specific fixed effects and the vector $Weather_{t,s}$ contains the flexible weather controls. $Wind\ assist_{i,s,t}$ is the wind assist measured inside the stadium with anemometers. α_i indicates individual fixed effects. S_s, $C_{c(i,t)}$ and $S * C_{s,c(i,t)}$ are respectively stadium fixed effects, team

millimiters is binned in the following intervals: no precipitation, (0,1], (1, 5], (5, 10], (10, 100].

18 Only $\tilde{Y}$ and $Wind\ assist$ vary within an individual in a given day.

fixed effects, and their interaction. Standard errors are clustered at the stadium-date level.

1.5 RESULTS

Table 2 presents results for the baseline specification. I find that a 10 $\mu g/m^3$ increase in concentrations reduces performance by 1% of a standard deviation. For the median performance, this is equivalent to the loss of a third of a percentile in nationwide rankings. It should be noted that that most competitions occur in warmer months when pollution levels are relatively low.[19] Indeed 91% of performances in the data happen below 25 $\mu g/m^3$, the annual limit value set by the European Union, and more than half below 15 $\mu g/m^3$.

Column (2) tests whether the result is driven by the correlation between ozone and PM 2.5. Since fewer stadiums are within a 10 km range of an ozone monitoring station, the sample size is slightly reduced. The findings are partially at odds with Mullins, 2018: in an environment with higher levels of both ozone and PM 2.5, I find no statistically discernible effect of ozone, conditional on concentrations of PM 2.5. On the other hand, he finds a discernible negative effects of ozone only for endurance events. I show in Section 1.5.1 that while performance losses attributable to both PM 2.5 and ozone increase with duration of effort, a proxy for reliance on the cardio-pulmonary system, the effect for PM 2.5 is substantially stronger and still discernible for short-lasting events. For comparison with studies on the effects of air pollution on cognitive abilities, Ebenstein, Lavy, and Roth, 2016 find that a 10 $\mu g/m^3$ increase in PM 2.5 is associated with a reduction of 3.9% of a standard deviation in the score of high-stake school exams in Israel. Carneiro, Cole, and Strobl, 2021 estimate the relationship between PM 10, particulate matter smaller than 10 μm and including PM 2.5, and results in Brazil's nationwide university entrance examinations. They find that an increase of 10 $\mu g/m^3$ of PM 10 on the day of examinations leads to a reduction of 8% of a standard deviation in student' scores. When PM 10 is above 20 $\mu g/m^3$, the effect is 13%. According to Roth, 2022, a 10 $\mu g/m^3$ increase in indoor PM 10 reduces test scores London-area university students taking high-stakes exams by approximately 3% of a standard deviation. Bedi et al., 2021 runs grammatical reasoning test in a lab with university students, and find that +10 $\mu g/m^3$ in PM 2.5 reduce scores by 3%.

19 The average effect on performance of increase in PM 2.5 of 10 $\mu g/m^3$ is comparable to the effect of a reduction in maximum daily temperature from 24-26 degrees to 10-14 degrees (Figure 8 in Appendix). As discussed in Section 1.3.3, the daily maximum temperature is a better measurement of the temperature to which athletes are exposed.

Table 2: The impact of PM 2.5 on physical performance. Main specifications.

	(1)	(2)
	Std result	Std result
PM 2.5	-0.0010***	-0.0010**
	(0.0004)	(0.0004)
Ozone		-0.0002
		(0.0001)
Individual FE	Yes	Yes
Time	Yes	Yes
Weather	Yes	Yes
Stadium, Team	Yes	Yes
Observations	553171	507718

Note: The table shows the effects of contemporaneous PM 2.5 on physical performance, measured as track and field competitions results. The unit of analysis is the competition result of an individual. The dependent variable is standardized competition result, defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. PM 2.5 and ozone are expressed in $\mu g/m^3$. *Time* indicates year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for $2°$ C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions. Standard errors are clustered at the stadium-date level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

1.5.1 Aerobic and anaerobic activities

Next, I examine heterogeneous effects by duration of effort, a proxy for the reliance on oxygen intake and on the respiratory system. At low intensities of effort, the human body produces energy through the combustion of oxygen and fuel,[20] releasing carbon dioxide and water as byproducts. This energy production process is termed "aerobic" for the usage of oxygen. While efficient, it is relatively slow as it relies on the circulatory system to deliver oxygen to the working muscles before producing. When significant energy is required for a short burst of activity, muscles fall back on the internal storage of fuel and rapidly but inefficiently produce new fuel. Lactic acid, a byproduct, accumulates until force generation and energy production are inhibited. This energy production process is termed the "anaerobic" pathway as it takes place in the absence of oxygen (Spurway, 1992).

Anaerobic and aerobic pathways are not mutually exclusive, and which supply route is prioritized depends in part on the intensity of the work and partly on the duration of the work. [21] In 400-m track running, which on average lasts about a minute in my data,

20 Glucose, glycogen, fats, and proteins.
21 Intensity and duration are inversely proportional, as the accumulation of lactic acid limits energy production

the anaerobic (aerobic) component is responsible for approximately 40/60% for males and 45/55% for females; in 800-m running (2 and a half minutes in my data) the contribution shifts to 60/40% and 70/30% for males and females respectively (Duffield, Dawson, and Goodman, 2005).

The main hypothesized channel through which PM 2.5 alters the body's normal functioning is through inflammation in the lungs and reduced oxygen intake (Pope and Dockery, 2006).[22] The expectation is that longer-lasting competitions, where oxygen intake is an important element, will have larger losses to PM 2.5.

Following this rationale, I compute heterogeneous effects by the typical duration of a competition, calculated as the average duration of a given event by age and gender. Figure 3 shows that the marginal effect of PM 2.5 on performance is negative and increases in magnitude as the average duration of an event increases. The effect is twice as large for events lasting on average 4 minutes than for very short events (Table 3).[23] The negative effect of ozone on performance also grows with the duration of effort, consistent with Mullins, 2018; however, the magnitude is substantially smaller and statistically discernible only for competitions that last on average 3 minutes or longer. This suggest that ozone affects mostly aerobic activities.

While generalization of results should be done with care, given the young age of individuals considered, the estimates suggest that tasks relying on pulmonary system and oxygen intake bear greater costs of air pollution. The findings can be seen in light of the work done by the US Bureau of Labor Statistics, which for each of almost a thousand occupations defines the importance of different physical abilities, and the level of ability required. According to the Bureau, in the United States *stamina* is an important ability for as many as 8.5 million workers in the country (Table 9); *explosive strength* is important for half a million workers. [24] More generally, 13.7% of all civilian jobs and 45.5% of jobs in Construction and Extraction require heavy work (Table 10) (U.S. Bureau of Labor and Statistics).

22 Some particles can pass from the airways directly into the bloodstream (Brook et al., 2010)

23 The coefficient for duration is positive and significant. Recalling that the identifying variation is within-individual, this means that, on average, individuals perform better, relative to themselves, in longer-lasting events.

24 *Important* is here defined as a 3 or more on a 1-5 scale from 'Not Important'(1) to 'Extremely Important' (5).

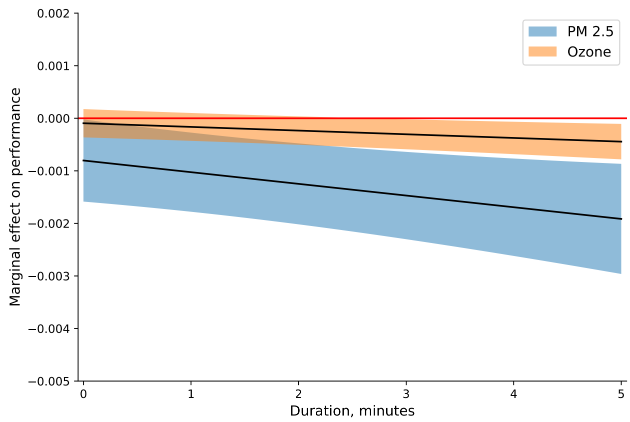


Figure 3: Marginal effect of PM 2.5 and ozone on performance by average event duration.

Table 3: Heterogeneous effects by task requirements.

	(1) Std result	(2) Std result
PM 2.5	-0.0008**	-0.0008**
	(0.0004)	(0.0004)
Duration, minutes	0.0120***	0.0180***
	(0.0017)	(0.0030)
PM 2.5 × Duration, minutes	-0.0003***	-0.0002**
	(0.0001)	(0.0001)
Ozone		-0.0001
		(0.0001)
Ozone × Duration, minutes		-0.0001***
		(0.0000)
Individual FE	Yes	Yes
Time	Yes	Yes
Weather	Yes	Yes
Stadium, Team	Yes	Yes
Observations	553171	507718

Note: The table shows the effects of contemporaneous PM 2.5 and ozone on physical performance, measured as track and field competitions results. The unit of analysis is the competition result of an individual. The dependent variable is standardized competition result, defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. *Duration* is the average duration of competitions (in minutes) for groups defined by age, gender, and event. PM 2.5 and ozone are expressed in $\mu g/m^3$. *Time* indicates year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions. Standard errors are clustered at the stadium-date level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

1.5.2 Gender, ability effects, and nonlinearities

To test whether the performance cost of PM 2.5 differs across individuals, I explore the heterogeneity across gender and ability (Table 4). There exists little large scale causal evidence on the costs of pollution by gender, in particular on physical abilities. It appears that PM 2.5 has no different impact on the performance of females and males, as Columns (1) and (2) show.

Columns (3) and (4) interact PM 2.5 with an indicator for athletes that perform in the top decile at least half of the time when compared to their peers. The latter identifies high ability athletes that systematically perform well. The performance loss caused by PM 2.5 is greater for top athletes, approximately 2.5 times as large. One possible explanation is that low ability athletes have more margin to compensate losses from air pollution.

Economics theory states that protection from air pollution should be set so that the marginal cost of investment matches the marginal benefits. Threshold effects and non-linearities in productivity losses imply that protective investments should not scale linearly as well. I test for non-linearity with multiple specifications, namely: restricted cubic splines with three and four knots; quadratic form of PM 2.5; binning PM 2.5 by half, tercile, and quantile. Results are shown in Figure 4. From all specifications we can deduce that exposure to PM 2.5 appears to be having a non-discernible effect at low concentrations (<25 $\mu g/m^3$), but negative effects on performance are evident at medium concentrations. Results are not driven by high concentrations of PM 2.5. Table 5 shows the estimates for a restricted sample with PM 2.5 less than 50 $\mu g/m^3$ (Column(1)); and less than 75 $\mu g/m^3$ (Column (2)). Estimated coefficients are almost unchanged.

1.6 ROBUSTNESS

As noted in Section 1.1, races on mid and long distances can require a degree of strategy if incentives nudge competitors to run for the win, but not for the timing. In such conditions, athletes may decide to maintain an artificially slow pace throughout the race and bet on their abilities to win a late-race acceleration. This requires runners to carefully evaluate their ability to maintain an optimal pace and the ability to outperform competitors in a final sprint. It is possible that inhalation of PM 2.5 might disrupt the necessary mental processes and reduce performance in these races.

Strategic running inherently reduces performance as measured in seconds. Estimates of the impact of PM 2.5 might be biased away from zero if strategic running is more common on polluted days; for instance, if important championships are held in large and polluted

Table 4: Heterogeneous effects by gender and ability.

	Gender		High ability	
	(1) Std result	(2) Std result	(3) Std result	(4) Std result
PM 2.5	-0.0011***	-0.0011***	-0.0009**	-0.0008**
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Female × PM 2.5	0.0002	0.0004		
	(0.0004)	(0.0004)		
Ozone		-0.0002		-0.0002
		(0.0001)		(0.0001)
PM 2.5 × High ability			-0.0016***	-0.0015**
			(0.0006)	(0.0006)
Individual FE	Yes	Yes	Yes	Yes
Time	Yes	Yes	Yes	Yes
Weather	Yes	Yes	Yes	Yes
Stadium, Team	Yes	Yes	Yes	Yes
Observations	553171	507718	553171	507718

Note: The table shows the effects of contemporaneous PM 2.5 on physical performance, measured as track and field competitions results. The unit of analysis is the competition result of an individual. The dependent variable is standardized competition result, defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. *High ability* is an indicator for athletes that perform in the top decile at least 50% of the time. *Time* indicates year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions. Standard errors are clustered at the stadium-date level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

cities. While such a scenario is plausible, the amount of bias should be limited once stadium and time fixed effects are included in the regression.

To address the remaining doubts, and to ensure results do not pick up a cognitive effect, I estimate the main specifications excluding all race competitions of distance over 400 meters and report results in Table 6. Table 12 in Appendix further addresses strategic behavior in multi-stage competitions including qualifiers. Results are unaltered, confirming that strategic races do not drive the observed impacts of PM 2.5.

Individuals may avoid competing in locations with high pollution levels if they fear their health is at risk (Graff Zivin and Neidell, 2013). Although unlikely given the low concentrations during spring and summer, they might choose whether and where to compete depending on factors, such as weather conditions, that correlate with pollution. The inclusion of individual fixed effects assures that the identifying variation does not come from the selection of less performing athletes into high pollution days. Nonetheless, I test whether concentrations of PM 2.5 predict participation in competitions. Table 7 reports the results of a regression of the log number of participants in a given

Table 5: The effect of air pollution on performance: excluding high concentrations.

	PM 2.5 < 50	PM 2.5 < 75
	(1)	(2)
	Std result	Std result
PM 2.5	-0.0008**	-0.0010**
	(0.0004)	(0.0004)
Ozone	-0.0002	-0.0002
	(0.0001)	(0.0001)
Individual FE	Yes	Yes
Time	Yes	Yes
Weather	Yes	Yes
Stadium, Team	Yes	Yes
Observations	505347	507508

Note: The table shows the effects of contemporaneous PM 2.5 on physical performance, measured as track and field competitions results. The unit of analysis is the competition result of an individual. The dependent variable is standardized competition result, defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. Column (1) reports results after excluding events with PM 2.5 greater or equal to 50 $\mu g/m^3$. Column (2) reports results after excluding events with PM 2.5 greater or equal to 75 $\mu g/m^3$. *Time* indicates year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions. Standard errors are clustered at the stadium-date level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

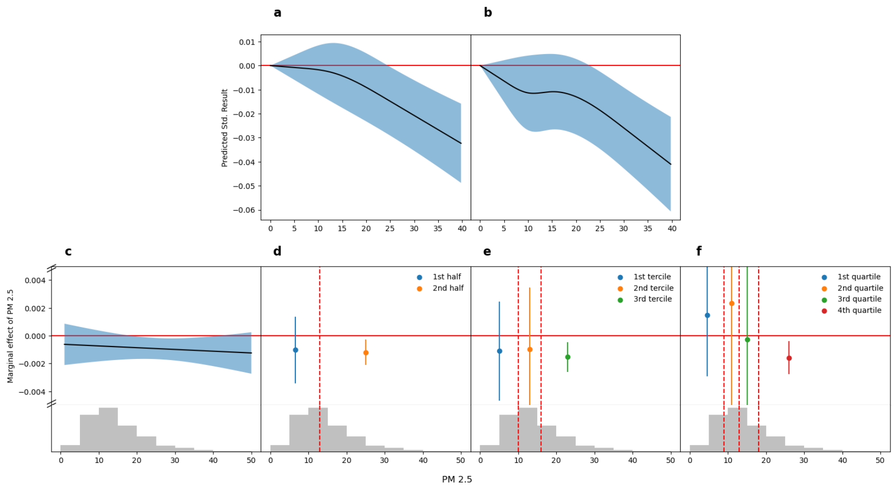


Figure 4: Nonlinear effects of PM 2.5 on performance. Panel **a** and **b** show the predicted performance estimated with a restriced cubic spline with three and four knots, respectively. Knot locations are based on Harrell's (2001) recommended percentiles. Panel **c**, **d**, **e**, and **f** report the marginal effects of PM 2.5 on performance for difference specifications: quadratic (**c**), by sample half (**d**), tercile (**e**), and quartile **(f)**. Histograms at the bottom report the distribution of PM 2.5 in the sample period. The few (0.05%) observations larger than 50 $\mu g/m^3$ have been excluded from the graphs for clarity.

stadium-date on PM 2.5, progressively adjusting for stadium, time of the year, and weather. If anything, on days with higher pollution, *more* athletes take part to competitions (Column (1)). However, once the invariable characteristics of stadiums are accounted for, neither PM 2.5 nor ozone predict participation to contests (Columns (2), (3) and (4)).

Finally, to further assess the robustness of results I perform a placebo test replacing contemporaneous concentrations of PM 2.5 and ozone with those observed in the same city one year later. Future concentrations do not predict competitions results (Table 8). This is reassuring that previous results are not driven by unmodeled seasonality patterns.

Table 6: Excluding events where strategic behavior is possible.

	(1)	(2)
	Std result	Std result
PM 2.5	-0.0010***	-0.0009**
	(0.0004)	(0.0004)
Ozone		-0.0002
		(0.0001)
Individual FE	Yes	Yes
Time	Yes	Yes
Weather	Yes	Yes
Stadium, Team	Yes	Yes
Observations	468162	428623

Note: The table shows the effects of contemporaneous PM 2.5 on physical performance, measured as track and field competitions results. The unit of analysis is the competition result of an individual. The dependent variable is standardized competition result, defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. The sample excludes all race competitions of distance over 400 meters. PM 2.5 and ozone are expressed in $\mu g/m^3$. *Time* dummies include year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions.

Table 7: Testing for presence of avoidance behavior.

	(1) Log(Partecipants)	(2) Log(Partecipants)	(3) Log(Partecipants)	(4) Log(Partecipants)
PM 2.5	0.0042*	-0.0012	0.0037	0.0008
	(0.0026)	(0.0024)	(0.0024)	(0.0029)
Ozone	0.0048***	0.0023***	0.0004	-0.0006
	(0.0008)	(0.0008)	(0.0009)	(0.0016)
Time	No	No	Yes	Yes
Weather	No	No	No	Yes
Stadium	No	Yes	Yes	Yes
Observations	3246	3246	3246	2926

Note. The table tests whether concentrations of PM 2.5 and ozone predict participation to competitions. The dependent variable is the log-number of participants to competitions in a given stadium-date. *Time* indicates year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium* includes stadium fixed effects. Standard errors are clustered at the stadium-date level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: The impact of PM 2.5 on physical performance. Placebo test with future air pollution measures.

	(1) Std result	(2) Std result
PM 2.5, 1-yr lead	-0.0005	-0.0006
	(0.0004)	(0.0004)
O3, 1-yr lead		0.0002
		(0.0001)
Individual FE	Yes	Yes
Time	Yes	Yes
Weather	Yes	Yes
Stadium, Team	Yes	Yes
Observations	469297	430911

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: The dependent variable is standardized competition result, which is the competition results minus the average result of a group defined by age, gender, and event, and dividing by the standard deviation of results of the same group (*e.g.*, 17-old, male, long jump). *PM 2.5, 1-yr lead* and *O3, 1-yr lead* are the concentrations of PM 2.5 and ozone observed one year later. PM 2.5 and ozone are expressed in $\mu g/m^3$. *Time* dummies include year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions.

1.7 CONCLUSIONS

A body of studies has assessed the effect on worker's productivity of environmental stressors such as pollution and temperature. To overcome limits to portability of results, a growing number of works studies the impacts on standardized tasks as it allows comparison of outcomes between individuals with the same assignment. However, generalization does not always follow from standardization, as the mechanisms at work often remain fuzzy. Moreover, most works focus on the effects on cognition and the evidence on the productivity effects through physical channels remains limited.

This paper offers new evidence on the impacts of PM 2.5 leveraging on a large dataset of track and field competitions, a set of highly standardized and primarily physical activities. The simplicity of the tasks involved - running, jumping, throwing - and their well-understood physiology make extension to other physical activities more transparent. The richness of the data allows for assessing the link between short-term exposure to PM 2.5 and performance, and then to explore one particular driver of the effects: the duration of continuous effort and, implicitly, the reliance on stamina.

I find that an increase in PM 2.5 of 10 $\mu g/m^3$ reduces performance by 1% of a standard deviation after including a battery of fixed effects, including individual fixed effects and a flexible modeling of weather. The impact of PM 2.5 on performance grows as the duration of competitions - and the dependence on the pulmonary system - increase. The results suggest that jobs requiring exertion of muscle force continuously over time incur, under the same conditions, greater productivity losses than jobs requiring short burst of intense exercise. While track and field competitions differ from most physical work in intensity and participants, the analysis explores heterogeneity that might extend to common manual jobs. The findings highlight potentially unequal costs of air pollution across the hundreds of millions of workers worldwide employed physical labor, adding to current concerns over distributional consequences of environmental stressors (Hsiang, Oliva, and Walker, 2019).

REFERENCES

He, Jiaxiu, Haoming Liu, and Alberto Salvo (Jan. 2019). "Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China". en. In: *American Economic Journal: Applied Economics* 11.1, pp. 173–201. ISSN: 1945-7782. DOI: 10.1257/app.20170286. URL: https://www.aeaweb.org/articles?id=10.1257/app.20170286.

Rangel, Marcos A. and Tom S. Vogl (Dec. 2018). "Agricultural Fires and Health at Birth". In: *The Review of Economics and Statistics* 101.4, pp. 616–630. ISSN: 0034-6535. DOI: 10.1162/rest_a_00806. URL: https://doi.org/10.1162/rest_a_00806 (visited on 01/28/2021).

Graff Zivin, Joshua, Tong Liu, et al. (Nov. 2020). "The unintended impacts of agricultural fires: Human capital in China". en. In: *Journal of Development Economics* 147, p. 102560. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2020.102560. URL: http://www.sciencedirect.com/science/article/pii/S0304387820301358 (visited on 11/05/2020).

He, Guojun, Tong Liu, and Maigeng Zhou (Mar. 2020). "Straw burning, PM2.5, and death: Evidence from China". en. In: *Journal of Development Economics*, p. 102468. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2020.102468. URL: http://www.sciencedirect.com/science/article/pii/S0304387820300432 (visited on 03/21/2020).

Ebenstein, Avraham, Victor Lavy, and Sefi Roth (Oct. 2016). "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution". In: *American Economic Journal: Applied Economics* 8.4, pp. 36–65. DOI: 10.1257/app.20150213. URL: http://www.aeaweb.org/articles?id=10.1257/app.20150213.

Persico, Claudia L. and Joanna Venator (Aug. 2019). "The Effects of Local Industrial Pollution on Students and Schools". en. In: *Journal of Human Resources*, 0518–9511R2. ISSN: 0022-166X, 1548-8004. DOI: 10.3368/jhr.56.2.0518-9511R2. URL: http://jhr.uwpress.org/content/early/2019/08/02/jhr.56.2.0518-9511R2 (visited on 01/28/2021).

Bedi, Arjun S. et al. (May 2021). "Particle Pollution and Cognition: Evidence from Sensitive Cognitive Tests in Brazil". In: *Journal of the Association of Environmental and Resource Economists* 8.3, pp. 443–474. ISSN: 2333-5955. DOI: 10.1086/711592. URL: https://www.journals.uchicago.edu/doi/10.1086/711592 (visited on 10/19/2021).

Nauze, Andrea La and Edson R. Severnini (May 2021). *Air Pollution and Adult Cognition: Evidence from Brain Training*. en. Tech. rep. National Bureau of Economic Research. DOI: 10.3386/w28785. URL: https://www.nber.org/papers/w28785 (visited on 05/10/2021).

Künn, Steffen, Juan Palacios, and Nico Pestel (Sept. 2019). *Indoor Air Quality and Cognitive Performance*. IZA Discussion Papers 12632. Institute of Labor Economics (IZA). URL: https://ideas.repec.org/p/iza/izadps/dp12632.html.

Archsmith, James, Anthony Heyes, and Soodeh Saberian (2018). "Air Quality and Error Quantity: Pollution and Performance in a High-Skilled, Quality-Focused Occupation". In: *Journal of the Association of Environmental and Resource Economists* 5.4, pp. 827–863. DOI: 10.1086/698728. eprint: https://doi.org/10.1086/698728. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/698728.

Burkhardt, Jesse et al. (2019). "The effect of pollution on crime: Evidence from data on particulate matter and ozone". In: *Journal of Environmental Economics and Management* 98, p. 102267. ISSN: 0095-

0696. DOI: https://doi.org/10.1016/j.jeem.2019.102267. URL: http://www.sciencedirect.com/science/article/pii/S0095069619301901.

Bondy, Malvina, Sefi Roth, and Lutz Sager (May 2020). "Crime Is in the Air: The Contemporaneous Relationship between Air Pollution and Crime". In: *Journal of the Association of Environmental and Resource Economists* 7.3, pp. 555–585. ISSN: 2333-5955. DOI: 10.1086/707127. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/707127 (visited on 03/27/2020).

Heyes, Anthony, Matthew Neidell, and Soodeh Saberian (Oct. 2016). *The Effect of Air Pollution on Investor Behavior: Evidence from the S&P 500*. Working Paper 22753. National Bureau of Economic Research. DOI: 10.3386/w22753. URL: http://www.nber.org/papers/w22753.

Chen, Xi (Feb. 2019). "Smog, Cognition and Real-World Decision-Making". In: *International Journal of Health Policy and Management* 8.2, pp. 76–80. ISSN: 2322-5939. DOI: 10.15171/ijhpm.2018.105. URL: https://www.ijhpm.com/article_3558.html (visited on 01/28/2021).

Cohen, Aaron J et al. (May 2017). "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015". en. In: *The Lancet* 389.10082, pp. 1907–1918. ISSN: 0140-6736. DOI: 10.1016/s0140-6736(17)30505-6. URL: http://www.sciencedirect.com/science/article/pii/S0140673617305056.

Pope, C. Arden II and Douglas W. Dockery (June 2006). "Health Effects of Fine Particulate Air Pollution: Lines that Connect". In: *Journal of the Air & Waste Management Association* 56.6, pp. 709–742. ISSN: 1096-2247. DOI: 10.1080/10473289.2006.10464485. URL: https://doi.org/10.1080/10473289.2006.10464485.

Chang, Tom et al. (Aug. 2016). "Particulate Pollution and the Productivity of Pear Packers". en. In: *American Economic Journal: Economic Policy* 8.3, pp. 141–169. ISSN: 1945-7731. DOI: 10.1257/pol.20150085. URL: https://www.aeaweb.org/articles?id=10.1257/pol.20150085.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham (2018). *Management and shocks to worker productivity: evidence from air pollution exposure in an Indian garment factory*. Tech. rep.

Fu, Shihe, V Brian Viard, and Peng Zhang (Nov. 2021). "Air Pollution and Manufacturing Firm Productivity: Nationwide Estimates for China". In: *The Economic Journal* 131.640, pp. 3241–3273. ISSN: 0013-0133. DOI: 10.1093/ej/ueab033. URL: https://doi.org/10.1093/ej/ueab033 (visited on 12/08/2021).

Carneiro, Juliana, Matt Cole, and Eric Strobl (Apr. 2021). "The Effects of Air Pollution on Students' Cognitive Performance: Evidence from Brazilian University Entrance Tests". In: *Journal of the Association of Environmental and Resource Economists*. ISSN: 2333-5955. DOI: 10.

1086/714671. URL: https://www.journals.uchicago.edu/doi/10.1086/714671 (visited on 04/28/2021).

Kahn, Lawrence M. (Sept. 2000). "The Sports Business as a Labor Market Laboratory". en. In: *Journal of Economic Perspectives* 14.3, pp. 75–94. ISSN: 0895-3309. DOI: 10.1257/jep.14.3.75. URL: https://www.aeaweb.org/articles?id=10.1257/jep.14.3.75 (visited on 01/31/2021).

Emerson, Jamie and Brian Hill (June 2018). "Peer effects in marathon racing: The role of pace setters". en. In: *Labour Economics* 52, pp. 74–82. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2018.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0927537118300307 (visited on 10/19/2021).

Frick, Bernd (JUN 2011). "Gender differences in competitiveness: Empirical evidence from professional distance running". English. In: *Labour Economics* 18.3, SI, 389–398. ISSN: 0927-5371. DOI: {10.1016/j.labeco.2010.11.004}.

Lichter, Andreas, Nico Pestel, and Eric Sommer (2017). "Productivity effects of air pollution: Evidence from professional soccer". In: *Labour Economics* 48.C, pp. 54–66. DOI: 10.1016/j.labeco.2017.06.. URL: https://ideas.repec.org/a/eee/labeco/v48y2017icp54-66.html.

Marcus, Michelle (Jan. 2021). "Pollution at Schools and Children's Aerobic Capacity". URL: https://michellemmarcus.files.wordpress.com/2021/01/marcus_aerocap_012820.pdf.

Austin, Wes, Garth Heutel, and Daniel Kreisman (June 2019). "School bus emissions, student health and academic performance". en. In: *Economics of Education Review* 70, pp. 109–126. ISSN: 0272-7757. DOI: 10.1016/j.econedurev.2019.03.002. URL: http://www.sciencedirect.com/science/article/pii/S0272775719301530.

Mullins, Jamie T. (2018). "Ambient air pollution and human performance: Contemporaneous and acclimatization effects of ozone exposure on athletic performance". en. In: *Health Economics* 27.8, pp. 1189–1200. ISSN: 1099-1050. DOI: https://doi.org/10.1002/hec.3667. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3667 (visited on 06/03/2021).

Hill, A. V. and H. Lupton (Jan. 1923). "Muscular Exercise, Lactic Acid, and the Supply and Utilization of Oxygen". In: *QJM* os-16.62, pp. 135–171. DOI: 10.1093/qjmed/os-16.62.135.

Sexton, Steven, Zhenxuan Wang, and Jamie T. Mullins (Mar. 2021). "Heat Adaptation and Human Performance in aWarming Climate". URL: https://people.umass.edu/jmullins/pdfs/Sexton,%20Wang,%20&%20Mullins%20(2021)%20Heat%20Adaptation%20and%20Human%20Performance%20in%20a%20Warming%20Climate.pdf.

Guo, Mengmeng and Shihe Fu (Oct. 2019). "Running With a Mask? The Effect of Air Pollution on Marathon Runners' Performance". en. In: *Journal of Sports Economics* 20.7, pp. 903–928. ISSN: 1527-0025.

DOI: 10.1177/1527002518822701. URL: https://doi.org/10.1177/1527002518822701.

Arcidiacono, Peter, Josh Kinsler, and Joseph Price (Jan. 2017). "Productivity Spillovers in Team Production: Evidence from Professional Basketball". In: *Journal of Labor Economics* 35.1, pp. 191–225. ISSN: 0734-306X. DOI: 10.1086/687529. URL: https://www.journals.uchicago.edu/doi/10.1086/687529 (visited on 10/19/2021).

World Athletics (2019). *Track and Field Facilities Manual*. Tech. rep.

– (2020). *Competition and Technical Rules*. Tech. rep.

Graff Zivin, Joshua and Matthew Neidell (Dec. 2012). "The Impact of Pollution on Worker Productivity". en. In: *American Economic Review* 102.7, pp. 3652–3673. ISSN: 0002-8282. DOI: 10.1257/aer.102.7.3652. URL: https://www.aeaweb.org/articles?id=10.1257/aer.102.7.3652.

Ortiz, Alberto (2020). *Air quality in Europe : 2020 report*. Luxembourg: Publications Office of the European Union. ISBN: 9789294802927. URL: https://www.eea.europa.eu/publications/air-quality-in-europe-2020-report/at_download/file.

Copernicus Climate Change Service (2019). *ERA5-Land hourly data from 2001 to present*. DOI: 10.24381/CDS.E2161BAC.

Neidell, Matthew (Mar. 2009). "Information, Avoidance Behavior, and Health The Effect of Ozone on Asthma Hospitalizations". en. In: *Journal of Human Resources* 44.2, pp. 450–478. ISSN: 0022-166X, 1548-8004. DOI: 10.3368/jhr.44.2.450. URL: http://jhr.uwpress.org/content/44/2/450 (visited on 05/21/2021).

Roth, Sefi (2022). *The Contemporaneous Effect of Indoor Air Pollution on Cognitive Performance: Evidence from the UK*. Tech. rep.

Spurway, NC (1992). "Aerobic exercise, anaerobic exercise and the lactate threshold". In: *British Medical Bulletin* 48.3, pp. 569–591.

Duffield, Rob, Brian Dawson, and Carmel Goodman (2005). "Energy system contribution to 400-metre and 800-metre track running". In: *Journal of sports sciences* 23.3, pp. 299–307.

Brook, Robert D. et al. (June 2010). "Particulate Matter Air Pollution and Cardiovascular Disease". In: *Circulation* 121.21, pp. 2331–2378. DOI: 10.1161/cir.0b013e3181dbece1.

Graff Zivin, Joshua and Matthew Neidell (Sept. 2013). "Environment, Health, and Human Capital". In: *Journal of Economic Literature* 51.3, pp. 689–730. DOI: 10.1257/jel.51.3.689. URL: http://www.aeaweb.org/articles?id=10.1257/jel.51.3.689.

Hsiang, Solomon, Paulina Oliva, and Reed Walker (Feb. 2019). "The Distribution of Environmental Damages". en. In: *Review of Environmental Economics and Policy* 13.1, pp. 83–103. ISSN: 1750-6816. DOI: 10.1093/reep/rey024. URL: https://academic.oup.com/reep/article/13/1/83/5304824.

A APPENDIX

A.1 Figures

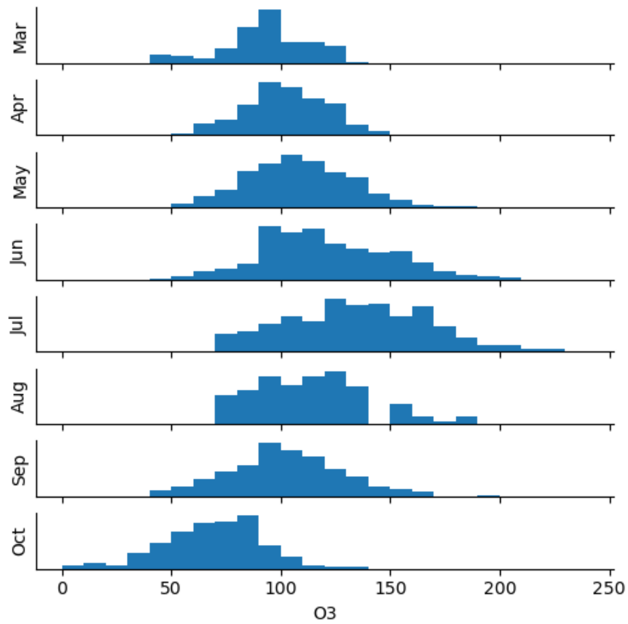


Figure 5: Within-month distribution of ozone.

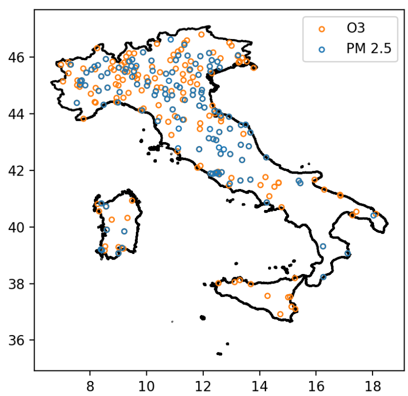


Figure 6: Location of pollution monitors in 2013. The monitor network is dense in the more populated and polluted North.

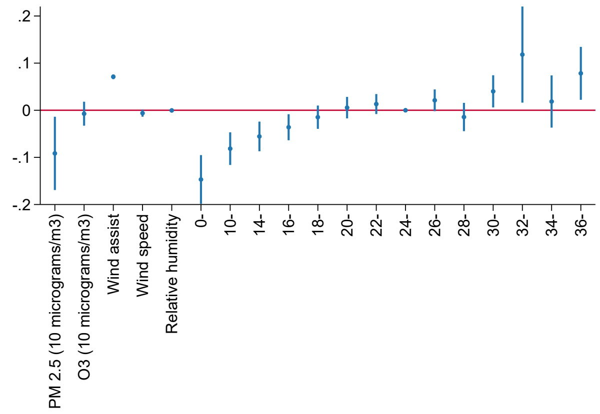


Figure 8: The impact of PM 2.5 on physical performance. Comparison with the effect of temperature. Reference temperature bin is [24-26) degrees Celsius. The dependent variable is standardized competition result, which is the competition results minus the average result of a group defined by age, gender, and event, and dividing by the standard deviation of results of the same group (*e.g.*, 17-old, male, long jump). The regression includes fixed effects for year, week, and day-of-the-week; stadium fixed effects, team fixed effects, and their interactions. Standard errors are clustered at the stadium-date level.

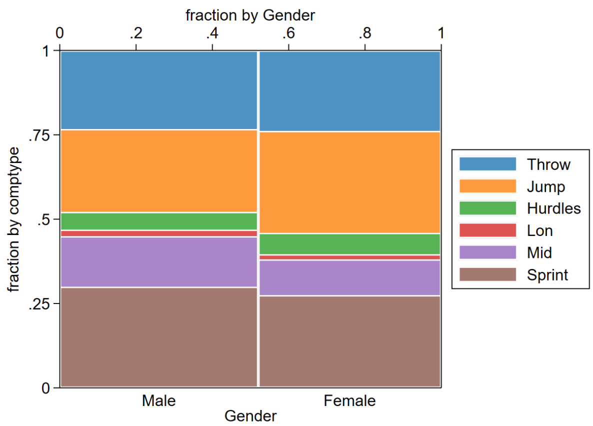


Figure 7: Share of observations by gender and type of event: races (sprints, hurdles, mid and long distance), jumps and throws.

A.2 Tables

Table 9: Top occupations with high importance of 'Stamina', by number of workers

Occupation	Size	Importance	Level
Laborers and Freight, Stock, and Material Movers, Hand	2821700	3.12	4.12
Waiters and Waitresses	2023200	3	3.62
Construction Laborers	1285200	3.12	4.12
Maids and Housekeeping Cleaners	1212800	3	3.88
Landscaping and Groundskeeping Workers	1117800	3	3.5
Light Truck Drivers	1035800	3	3.25
Nannies	992400	3	3.38
Carpenters	942900	3.12	4.06
Police and Sheriff's Patrol Officers	671200	3	4.12
Farmworkers and Laborers, Crop, Nursery, and Greenhouse	526300	3	3.81

Size: number of workers in the occupation. Importance: the degree of importance a particular descriptor is to the occupation. The possible ratings range from 'Not Important'(1) to 'Extremely Important' (5). Level: This rating indicates the degree, or point along a continuum, to which a particular descriptor is required or needed to perform the occupation.

Table 10: Percentage of civilian jobs requiring different strength levels in selected occupations, 2016.

Occupation	Sedentary	Light work	Medium work	Heavy work
All jobs	13.30%	24.40%	45.00%	13.70%
Management	27.5	36.9	31.7	-
Architecture and engineering	24.9	25.8	41.4	-
Community and social service	24.7	36.2	34.9	-
Legal	40.3	28.9	30.8	-
Education, training, and library	-	48.9	40.7	5.2
Arts, design, entertainment, sports, and media	22.7	33.1	36.3	-
Healthcare support	-	23.3	48.1	21.9
Food preparation and serving related	-	22.2	67.2	9.8
Building and grounds cleaning and maintenance	-	13.8	69.7	15.7
Personal care and service	-	29	49.7	16.9
Sales and related	9.8	22.9	58.4	-
Office and administrative support	31.1	33.5	27.7	5.1
Construction and extraction	-	-	37.3	45.5
Installation, maintenance, and repair	-	-	49.3	35.4
Production	-	12	63.2	17.2
Transportation and material moving	2.8	10.2	47.4	32.3

Source: Bureau of Labor Statistics, U.S. Department of Labor. Dash indicates no jobs in this category or data did not meet publication criteria of the BLS.

Table 11: Decomposition of variance of PM 2.5 and performance.

	Standard Deviation	
	PM 2.5	Standardized result
Overall	8.36	0.98
Within individual	7.05	0.62
Between individuals	5.65	0.83
Within stadium	7.43	0.97
Between stadiums	4.81	0.25
Within individual-stadium	5.50	0.52

Note: The panel is unbalanced. Standardized competition result is defined as competition results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group.

Table 12: The impact of PM 2.5 on physical performance. Excluding events where strategic behavior is possible. Races over distances greater than 400 meters are excluded and for each athlete only the best result in a given day in a given event is included. For instance, qualifying rounds with poorer results than finals are excluded.

	(1)	(2)
	Std result	Std result
PM 2.5	-0.0010***	-0.0009**
	(0.0004)	(0.0004)
Ozone		-0.0002
		(0.0001)
Individual FE	Yes	Yes
Time	Yes	Yes
Weather	Yes	Yes
Stadium, Team	Yes	Yes
Observations	463175	423859

Note: The table shows the effects of contemporaneous PM 2.5 on physical performance, measured as track and field competitions results. The unit of analysis is the competition result of an individual. The dependent variable is standardized competition result, defined as results minus the average result of a group defined by age, gender, and event (*e.g.*, 17-old, female, long jump), divided by the standard deviation of results of the same group. Races over distances greater than 400 meters are excluded and for each athlete only the best result in a given day in a given event is included. PM 2.5 and ozone are expressed in $\mu g/m^3$. *Time* dummies include year, week, and day-of-the-week fixed effects. *Weather* includes wind assist, as well as fixed effects for 2° C bins of maximum daily temperature and their interaction with wind, relative humidity, and binned precipitation. *Stadium, Team* includes stadium fixed effects, team fixed effects, and their interactions.

A.3 Tasks vs occupations

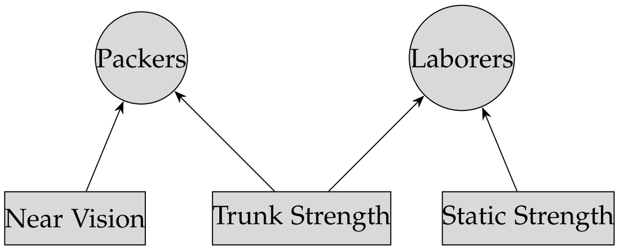


Figure 9: Example of jobs (circles) requiring an overlapping set of abilities (squares).

Suppose that we can approximate the productivity of J and Q with linear functions: $P^J = \lambda_1 A(e) + \lambda_2 B(e)$ and $P^Q = \lambda_3 B(e) + \lambda_4 C(e)$, with e representing the level of the environmental stressor. An estimate of the productivity effect of e on the output of J, $\frac{\partial P^J}{\partial e} = \lambda_1 \frac{\partial A}{\partial e} + \lambda_2 \frac{\partial B}{\partial e}$ can provide little information on the productivity effect $\frac{\partial P^Q}{\partial e} = \lambda_3 \frac{\partial B}{\partial e} + \lambda 4 \frac{\partial C}{\partial e}$. However, suppose we can observe $\frac{\partial B}{\partial e}$, the moderating effect on B alone. We can say the effect is consequential for both J and Q to the degrees λ_2 and λ_3 they rely on ability B.

Proxies for λ_2 and λ_3 can be The Occupational Information Network (O*NET) database contains hundreds of standardized and job-specific descriptors on nearly 1,000 jobs, covering the entire US economy. The database, which is freely available to the public, is continually updated with input from a wide range of workers in each occupation.

2 COVID-19 LOCKDOWN ONLY PARTIALLY ALLEVIATES HEALTH IMPACTS OF AIR POLLUTION IN NORTHERN ITALY

with

Lara Aleluia Reis, *RFF-CMCC European Institute on Economics and the Environment*

Valentina Bosetti, *Bocconi University* and *RFF-CMCC European Institute on Economics and the Environment*

Massimo Tavoni, *Politecnico di Milano* and *RFF-CMCC European Institute on Economics and the Environment*

Published on *Environmental Research Letters*.

ABSTRACT Evaluating the reduction in pollution caused by a sudden change in emission is complicated by the confounding effect of weather variations. We propose an approach based on machine learning to build counterfactual scenarios that address the effect of weather and apply it to the COVID-19 lockdown of Lombardy, Italy. We show that the lockdown reduced background concentrations of $PM_{2.5}$ by 3.84 $\mu g/m^3$ (16%) and NO_2 by 10.85 $\mu g/m^3$ (33%). Improvement in air quality saved at least 11% of the years of life lost and 19% of the premature deaths attributable to COVID-19 in the region during the same period. The analysis highlights the benefits of improving air quality and the need for an integrated policy response addressing the full diversity of emission sources.

1 INTRODUCTION

Exposure to airborne pollutants is detrimental to human health. Fine particulate matter ($PM_{2.5}$) increases mortality rates and hospitalizations due to respiratory and cardiovascular disease (Pope and Dockery, 2006; Ebenstein, Fan, et al., 2017; Deryugina et al., 2019). Additionally, it leads to a decline in physical and cognitive productivity (Graff Zivin and Neidell, 2012; Ebenstein, Lavy, and Roth, 2016; Xin Zhang, X. Chen, and Xiaobo Zhang, 2018; J. He, H. Liu, and Salvo, 2019; Kahn and P. Li, 2020). Similarly, exposure to nitrogen dioxide (NO_2) leads to an increase in hospital admissions and premature mortality (Mills et al., 2015; Amini et al., 2019; Duan et al., 2019).

The design of effective pollution abatement policies requires a comprehensive understanding of the relationship between reductions of emissions and concentrations. However, the processes of formation, transport, and dispersion of pollutants are complex phenomena, introducing considerable uncertainty on the effect of policies on air quality. Moreover, impact assessments need to address the confounding effect of annual and daily weather variations, a significant driver of pollutant concentrations.

This paper provides novel evidence on the change in concentrations of PM 2.5 following a composite reduction in emissions across different sources. Specifically, we exploit the dramatic decrease in Italy's mobility and economic activity in response to the COVID-19 outbreak from late February to early May. We provide causal estimates of the change in $PM_{2.5}$ and NO_2 over more than two months for Lombardy, one of the most polluted regions among Organisation for Economic Co-operation and Development countries, and one of the first areas outside China that imposed a strict lockdown.

Using a machine-learning algorithm, we address the confounding effect of weather and build a counterfactual scenario of the pollution concentrations that would have occurred if the COVID-19 pandemic had not broken out and no lockdown had been implemented. Finally, we compute the years of life saved and the number of premature deaths avoided by the improvement in air quality. We compare these numbers against the years of life lost and premature deaths due to COVID-19 in the region over the same period.

Ex-post studies can provide valuable estimates of the sensitivity of concentrations to emissions. However, a host of confounding factors can seriously hinder policy evaluation. In particular, the concentration of airborne pollutants is highly dependent on atmospheric conditions. Formation, transport, dispersion, and even emission of pollutants are directly or indirectly affected by the weather (Kroll et al., 2020). For instance, severe haze events in Beijing follow periodic cycles governed by meteorological conditions, especially wind patterns (Guo et al., 2014). Unless the confounding impact of weather is accounted for, the estimated change in concentrations following intervention will be biased.

A common approach to impact evaluation of pollution control policies is comparing areas that were affected by a policy and areas that were not (*e.g.*, G. He, Pan, and Tanaka, 2020 and Cole, Elliott, and B. Liu, 2020 for the case of COVID-19 lockdowns). However, even when differences in weather have been accounted for, unaffected and comparable areas may not always exist. For the problem at hand, a precise separation between affected and unaffected regions is not possible, considering the ubiquitous adoption of measures to control the spreading of COVID-19.

We turn the complex correlation of weather and pollution to our favor, predicting concentrations as a function of weather variables and season with machine learning. We follow a simple strategy, similar to Petetin et al., 2020, that does not require the availability of comparable but unaffected regions. For each air pollution monitoring station in Lombardy, we train an extreme gradient boosting regressor (Friedman, 2001), a tree-based machine learning algorithm, over daily concentrations from 2012 to 2019 and predict concentrations for the first four months of 2020. We show in Supplementary information that this approach is more reliable than linear regression models. To account for any constant error in our prediction, including inter-annual trends (Silver et al., 2020), we adopt a difference-in-differences strategy. We identify the average impact of the lockdown on air pollution concentrations as the difference between the prediction error before and during the lockdown.

We find that, despite the unprecedented halt in mobility and economic activity, the concentrations of major pollutants only partially decreased as a consequence of the lockdown. Background concentrations of $PM_{2.5}$ and NO_2 decreased by 3.84 $\mu g/m^3$ (16%) and 10.85 $\mu g/m^3$ (33%), respectively. Nonetheless, the improvement in air quality saved at least 11% of the years of life lost and 19% of the premature deaths attributable to COVID-19 in the region during the same period.

This paper contributes to several active strands of literature in air pollution research. First, it speaks to works on the assessment of pollution control policies, and in particular, to the growing corpus of research employing machine learning and fine-grained data. The paper illustrates an innovative procedure to quantify the implications of a change in emissions on outdoor concentrations of pollutants, isolating the effect of weather variability. While existing studies applying a similar approach restrict the analysis to no more than a few days, we show the conditions under which the procedure can be applied to longer time windows, the length of weeks or months. We illustrate the approach through a specific event - the lockdown of Lombardy, in Northern Italy - but it can be generalized wherever spatially and temporally detailed data on air pollution concentrations and atmospheric conditions are available.

Second, this paper is relevant to pollution control policies in the domain of study. Lombardy is a high-income, densely populated region, home to approximately 10 million people, and one of the most polluted in OECD countries. The European Commission has repeatedly referred Italy to the Court of Justice of the European Union over persistently high levels of NO_2 and PM_{10}, mainly in Lombardy and the rest of the Po Valley (European Commission v. Italian Republic, 2012; European Commission v. Italian Republic, 2019; European Commission v. Italian Republic, 2020). This study sheds light on the sectoral contributions to emissions of $PM_{2.5}$ and NO_2, offering tools to regulators and policymakers.

Finally, our study relates to the literature on source apportionment to different sectors, particularly agriculture, a topic of increasing relevance (Lelieveld et al., 2015). During the study period, agricultural production continued unaffected, and on average 11.6 $\mu g/m^3$ (39%) of PM_{10} in Milan, the largest city, were attributable to agriculture. We acknowledge that missing sufficient data on 2020 sectoral emissions and on the composition of $PM_{2.5}$, source apportionment to different sectors remains elusive. Were the data available, our machine learning approach could be used to exactly estimate changes in the composition of $PM_{2.5}$.

2 SECTORAL EMISSIONS DURING LOCKDOWN

The timing and nature of the lockdown of Lombardy and Italy are discussed in detail in the Supplementary information. We highlight here two key moments. On February 21, 2020, the first outbreak of COVID-19 in Italy was identified in the south of Lombardy. Within 24 hours, 11 municipalities in the region went under strict lockdown: schools were closed, all non-essential economic activities had to stop, and a stay-at-home order was in place. Teaching activities in the rest of Lombardy also were suspended. On March 8, authorities extended the lockdown to the rest of Lombardy; and to the rest of Italy on the following day. Lockdown measures were kept in place almost unaltered until May 4.

The progressive spreading of the virus in Northern Italy and the tightening of containment measures have substantially reduced mobility and economic activity. As mobile phone data reveals, the movement of individuals in Lombardy has followed a two-step response, following the first outbreak of COVID-19 cases in lower Lombardy (February 21) and the lockdown of the entire country (March 9) (Figure 10a). By mid-March, mobility dropped by three-fourths, according to data compiled by Google and Apple (Google, 2020; Apple, 2020). Under lockdown, all non-essential industrial production halted. As a consequence, energy demand in Northern Italy steadily decreased since March 9, as businesses shut down, bottoming to 50% of pre-lockdown levels after two weeks (Figure 10b).

However, not all major sources of emissions, especially those releasing precursors of $PM_{2.5}$, have been affected by restrictions. The lockdown forced most people to home isolation; it is sensible to hypothesize that emissions from residential buildings increased as a consequence. On the other hand, emissions from non-residential buildings might have decreased. Although data to confirm this is lacking, it is plausible that emissions from heating systems have not been affected substantially.

During the transition between winter and spring, agriculture becomes an important source of secondary $PM_{2.5}$ in Lombardy (INEMAR, 2017). The dispersal of animal liquids on open fields is a common (though regulated) practice that releases ammonia in the atmosphere, a precursor to secondary $PM_{2.5}$. Public authorities have not restricted agricultural activities during lockdown in the interest of securing food supplies. These practices have continued virtually unchanged compared to previous years (personal exchange with public officials at the regional office for agriculture).

The agricultural sector is responsible for almost all emissions of ammonia (NH_3) in the region (INEMAR, 2017), a precursor to particulate matter as it combines into ammonium nitrates and ammonium sulfates. Data on the decomposition of background PM_{10} in Milan shows that ammonium nitrates and ammonium sulfates accounted for almost 40% of PM_{10} concentrations during the lockdown (see Figure 13 in Supplementary information). This corroborates the evidence that restrictive measures did not meaningfully alter agricultural emissions.

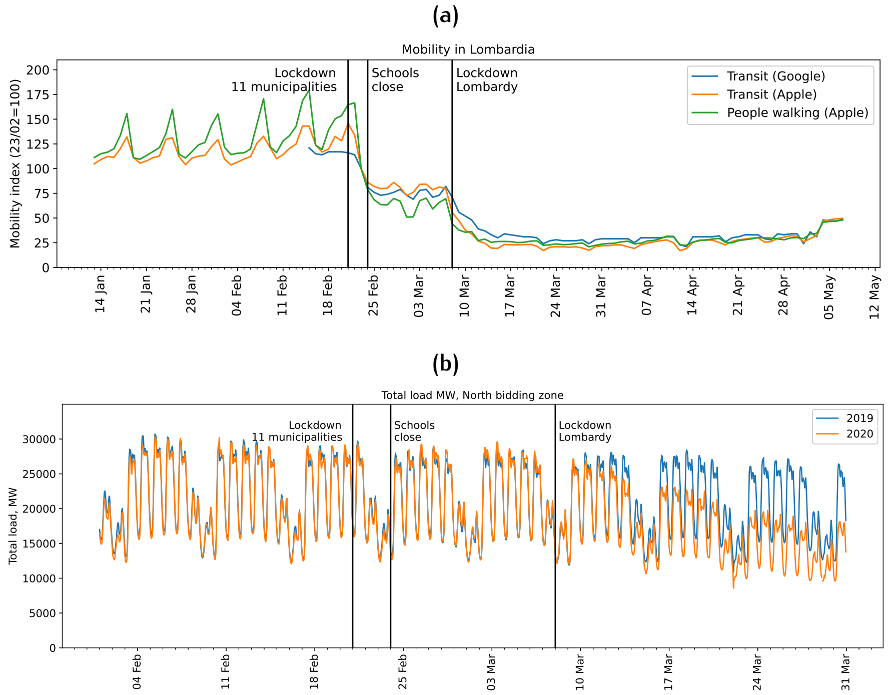


Figure 10: **Proxies of sectoral emissions.** **a**, Mobility indices for Milan Lombardy based on mobile phone data. Indices equal 100 on February 23. *Source*: Google, 2020; Apple, 2020. **b**, Total load of energy demand in Northern Italy in MW, 2019 vs 2010. The time series of 2019 has been shifted to match the day of the week. *Source*: TERNA, 2020.

3 METHODS

3.1 Machine learning

To identify the causal effect of the lockdown on concentrations without directly observing emissions, we build a synthetic counterfactual. We train a machine learning algorithm that can reproduce pollution concentrations on a business-as-usual scenario, and then predict concentrations during the lockdown. The difference between observed concentrations and the counterfactual, or prediction error, is the effect of the intervention. To account for potential systemic bias in the counterfactual, we adopt a difference-in-differences strategy. We identify the average impact of the lockdown on concentrations as the difference between the average prediction error before and during the lockdown. This approach does not require identifying comparable regions whose concentrations follow a business-as-usual trend.

We first assemble a dataset of air pollution, atmospheric conditions, and calendar variables for the period 2012 to 2020 for the Italian region of Lombardy. Pollution concentrations are measured at 83 monitoring stations. Data on daily minimum and maximum temperature, average wind speed and wind direction, average relative humidity, daily cumulative precipitation, and atmospheric soundings come from 227 weather stations.

For every monitoring station, we build the counterfactual using an extreme gradient boosting regressor, a tree-based model (Friedman, 2001).[1] Next, monitor by monitor, we train the algorithm on data from 2012 through 2019 and predict concentrations of $PM_{2.5}$ and NO_2 in 2020. We use the pre-lockdown period from January 1 to February 22, which was not included in the training set, to assess the validity of the counterfactual.

As our ultimate goal is a reliable prediction of pollutant concentrations from January through early May 2020, cross-validation is performed over four folds, each one consisting of the months from January to April for 2016, 2017, 2018, and 2019. The more common cross-validation on random subsamples, or *folds*, gives equal weight to all seasons. However, with such validation strategy it cannot be ruled out that an algorithm make good average predictions, while over-predicting in one season and under-predicting in the opposite one. Suppose, for instance, that the predictions of a learner are positively biased in spring, negatively biased in fall, and unbiased in winter and summer. In this case, testing predictions on the pre-lockdown period (in wintertime) does not give correct estimates of the bias during the lockdown (in springtime). For this reason, we perform cross-validation over the months for which we want predictions to be reliable. Model parameters are selected to maximize the cross-validated RMSE.

The identification strategy relies on two assumptions. First, input variables should not be themselves affected by the intervention; otherwise, estimated effects will be biased towards zero. To this end, we exploit the sensitivity of concentrations to meteorological conditions and build the counterfactual as a function of weather and season. While emissions are affected by weather (*e.g.*, lower emissions from heating systems on warmer days), our identification assumption is not violated as the weather is not affected by emissions. On the other hand, the algorithm implicitly learns the patterns of emissions as the weather varies and seasons pass.

Second, emissions that would have materialized absent the lockdown, and once weather has been accounted for, should be equal to emissions in the training period. One might be concerned that differences in technology (such as upgrading of the vehicle fleet) or economic activity between the training and prediction sample violate this assumption (Silver et al., 2020). We address this concern adopting a difference-in-differences strategy that excludes any constant prediction bias from the estimated effects of the lockdown. As long as the variation of observed values around the true counterfactual mean is

1 We use the python package `xgboost` (T. Chen and Guestrin, 2016).

well reproduced, estimates will be valid. Furthermore, the learner is cross-validated on data from 2016 through 2019; thus, recent years are given more weight.

We estimate the average effect of the lockdown with the following equation:

$$y_{it} - \widehat{y}_{it} = \alpha + \beta \mathrm{Lockdown}_t + \epsilon_{it} \tag{3}$$

where y_{it} is concentration measured at monitor i on day t, $\widehat{y}_{it}$ is the predicted value, and $\mathrm{Lockdown}$ is a dummy equal to 1 during the lockdown and 0 prior to it. α captures any time-invariant bias of the predictor; β is the parameter of interest; and ϵ_{it} is a random term. The preferred specification then distinguishes treatment effects by type of monitoring station.[2] Since concentrations are consequential to the extent that they reflect exposure, we weight observations by population within 20 kilometers from monitors.[3] We leave estimates of unweighted regressions, which yield qualitatively similar results, to the Supplementary information. To our knowledge, there is little guidance in the literature on how to estimate standard errors in this context properly. Thus, where reasonable, we cluster standard errors by monitor; where the number of clusters is small, we use robust heteroskedasticity-standard errors.

2 Namely background, industrial, and traffic monitoring stations.

3 Territory within 20 kilometers of two or more monitors is assigned to the closest monitor. The construction of population weights is described in more detail in Supplementary information.

3.2 Data sources

We assemble a dataset of air pollution, atmospheric conditions and calendar variables for the period 2012 to 2020 for the Italian region of Lombardy. The region is the home to about 10 million people and is the first contributor to national GDP by size. Its natural geography is conducive to low winds and stable air masses throughout the cold season. Mountain ranges to the North, West and South effectively block transboundary air streams extending wintertime thermal inversions and aggravating pollution events. For exceeding recommended air quality thresholds, Italy has been fined and subject to infringement procedures by the European Commission. We describe the data sources and pollution trends in Lombardy.

Air pollution

Data for air pollution is collected, checked, and published by *ARPA Lombardia*, the regional environmental agency.[4] We obtain readings for NO_2 and total $PM_{2.5}$ for background, traffic, and industrial stations as available. Hourly readings are averaged to daily readings. We exclude all monitoring stations that are not functioning during the lockdown or have been set up after 2015. Background stations account for about 60% of pollution monitors, traffic stations for about 30%, and the remaining 10% is located in industrial areas.

Average yearly concentrations of $PM_{2.5}$ in Milan, the region's capital, are systematically above the safety levels established by the WHO (10 $\mu g/m^3$); from December to the end of February, daily concentrations average above 40 $\mu g/m^3$. Average levels of NO_2 during the period are also well above WHO safety standards.

Weather data

Data on weather conditions at weather stations throughout the region are also elaborated and made available by *ARPA Lombardia*. We retrieve the daily minimum and maximum temperature; average wind speed and wind direction; average relative humidity; and daily cumulative precipitation. We further include a host of atmospheric sounding indices measured at Milano Linate airport and made available by the University of Wyoming, namely Showalter index, Lifted index, SWEAT index, K index, and Cross Totals, and Vertical Totals indices. All atmospheric variables enter as predictors in the form of contemporaneous and lagged values. Although monitor data and atmospheric soundings have gone through quality checks at the source, we winsorize all atmospheric predictors at 1 and 99 percentiles to bound the influence of extreme values.

4 Both air pollution and weather data are publicly available at https://www.dati.lombardia.it/stories/s/auv9-c2sj.

Additional predictors

The ratio of $PM_{2.5}$ to PM_{10} in Lombardy is typically altered in presence of pollution transported from long distances. For instance, a mass of dust from the Caspian Sea reached Northern Italy in late March, substantially altering the ratio. We assume the $PM_{2.5}$ to PM_{10} ratio is independent of the lockdown and include it among predictors as the concentration of $PM_{2.5}$ is affected by such shocks. Additional predictors are calendar variables to capture trends over time and seasons. We include year, month, week of the year, day of the month, day of the week in the form of continuous variables as well as dummy variables. We further include sine functions of time to mimic seasonality.

Population weights

Population weights for monitoring stations reflect the population within 20 kilometers of monitors (Figure 3 in Supplementary information). Population data on a 1 km by 1 km grid comes from the Italian National Statistical Office (ISTAT).[5] Grid cells within less than 20 kilometers from two or more monitors are assigned to the closest one.

3.3 Health impact assessment

To compute the number of avoided deaths and years of life saved by the reduction in PM2.5, we follow Fowlie, Rubin, and Walker, 2019 and take all-cause mortality relative risk (RR) ratios for $PM_{2.5}$ from two influential studies, Krewski et al., 2009 and Lepeule et al., 2012. In addition, we use the RR ratio recommended by the WHO (Henschel, Chan, Organization, et al., 2013) and adopted by the European Environment Agency (European Environment Agency, 2019). For NO_2, we only use the WHO recommendations. The calculation of avoided deaths and years of life saved from concentration-response functions is described in Supplementary Information A.1.

The more conservative estimates are based on Krewski et al., 2009, who report an hazard ratio 1.056 for an increase of 10 $\mu g/m^3$ of $PM_{2.5}$. Lepeule et al., 2012 estimate instead a larger hazard ratio of 1.14 for the same change in concentrations. The WHO recommends estimating the long-term impact of exposure to $PM_{2.5}$ in adult populations using an RR of 1.062 for 10 $\mu g/m^3$; it recommends an RR of 1.055 for 10 $\mu g/m^3$ of NO_2 above 20 $\mu g/m^3$ in adult populations.

5 The data is available at https://www.istat.it/it/files//2015/04/GEOSTAT_grid_POP_1K_IT_2011-22-10-2018.zip. Last accessed on July 23, 2020.

4 RESULTS AND DISCUSSION

4.1 Accuracy of predictions

To assess the accuracy of predictions, we test the counterfactual against observed values during the pre-lockdown period from January 1 to February 22, which has not been used for training. Table 13 reports mean values of Pearson's correlation coefficient (Corr), mean bias (MB), normalized mean bias (nMB), and root mean square error (RMSE). As we ultimately compute the difference-in-differences between observed values and the counterfactual, we also report the centered RMSE (cRMSE) and the normalized centered RMSE (ncRMSE).[6] For completeness, the table also includes statistics for the training set.

The correlation between observed and predicted values in the pre-lockdown period is 0.87 and 0.88 for $PM_{2.5}$ and NO_2, respectively. The counterfactual overestimates observed values by 1.34 $\mu g/m^3$ ($PM_{2.5}$) and 4.7 $\mu g/m^3$ (NO_2), thus motivating the use of a difference-in-differences strategy. The centered RMSE is 30% ($PM_{2.5}$) and 27% (NO_2) of mean observed concentrations. A graphical summary of model predictive performance, Taylor diagrams, can be found in Supplementary information.

In air pollution forecasting, machine learning techniques are typically used to predict concentrations an hour to few days ahead, and studies that can be used as benchmark are scarce. To the best of our knowledge, Petetin et al., 2020 is the only work whose methodology and length of forecast are comparable. They use machine learning to build a counterfactual for NO_2 concentrations in Spain during the COVID-19 lockdown. They report a normalized mean bias of 2% to 7%, depending on the type of station, a correlation coefficient of 0.71 to 0.75, and normalized RMSE of 28% to 32%. Compared to their study, our algorithm better mimics variation around the mean, than the mean itself. However, in our estimation strategy, any constant bias is captured by the constant in Equation 3.

6 The centered RMSE is computed as $\left[1/N \sum (\hat{y}_i - \bar{\hat{y}} - y_i + \bar{y})^2\right]^{1/2}$.

Table 13: Accuracy of predictions, average values across monitors

Pollutant	Dataset	Corr	MB	nMB	RMSE	cRMSE	ncRMSE
NO2	Train	1	.004	0	.276	.275	.008
NO2	Test	.875	-4.672	-.159	9.961	8.088	.261
PM2.5	Train	.999	0	0	.443	.443	.015
PM2.5	Test	.871	-1.335	-.049	8.764	8.476	.295

Notes: Corr: Pearson's correlation coefficient. *MB*: Mean bias, where negative values indicate observed values below predicted values. *nMB*: Normalized mean bias. *RMSE*: Root mean squared error. *nRMSE*: Normalized RMSE. *cRMSE*: Centered RMSE. *ncRMSE*: Normalized centered RMSE. Mean bias, RMSE and centered RMSE are expressed in $\mu g/m^3$. Mean bias, RMSE and centered RMSE are normalized dividing by mean observed concentrations. The centered RMSE is computed as $\left[1/N \sum(\hat{y}_i - \bar{\hat{y}} - y_i + \bar{y})^2\right]^{1/2}$.

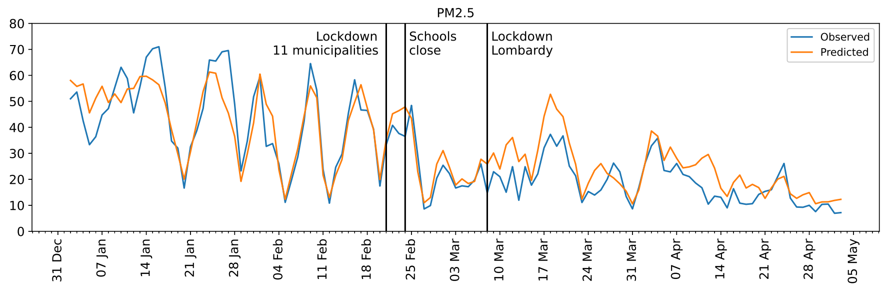
4.2 Effect of the lockdown on air pollution

Following the lockdown, air quality in Lombardy improved only partially. Figure 11 plots the population-weighted observed and counterfactual values for $PM_{2.5}$ (Figure 11a) and NO_2 (Figure 11b). NO_2 at background stations reached levels below the yearly limit set by the WHO Air Quality Guidelines. However, background concentrations of $PM_{2.5}$ still exceeded the daily limit of 25 $\mu g/m^3$ every one in four days.

The counterfactual well mimics observed values in the pre-lockdown period, corroborating the validity of the statistical approach. In contrast, a gap between observed and counterfactual values is evident as restrictions are tightened. We show in Supplementary information that the method outperforms a linear regression.

Suggestive evidence of the effect of the lockdown on concentrations of NO_2, which in Lombardy largely originate from motor vehicles, is visible from the week of February 25, consistent with the reduction in mobility documented in Figure 10a. The effect on $PM_{2.5}$ only appears as non-essential economic activities are halted in Lombardy and the rest of Italy, and is smaller in magnitude.

(a)



(b)

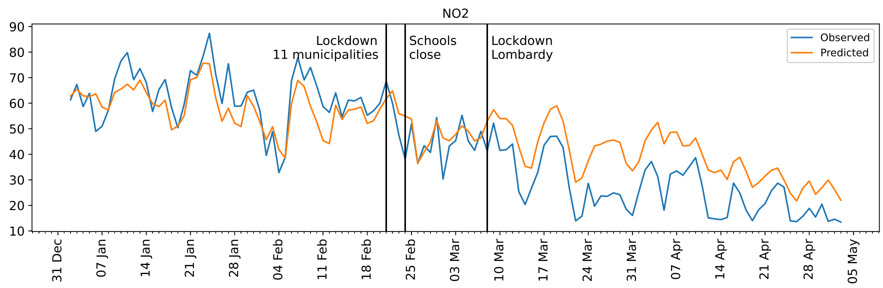


Figure 11: **Population-weighted average of observed and counterfactual values.** **a**, $PM_{2.5}$. **b**, NO_2. Population is measured within 20 kilometers of a monitoring station. Territory within less than 20 kilometers from two or more monitors is assigned to the closest one.

The lockdown may have affected $PM_{2.5}$ concentrations mainly through two channels: the reduction of primary $PM_{2.5}$ emissions, such as black and organic carbon, and reduction of precursors of secondary $PM_{2.5}$. We remark that NO_2 is a precursor of secondary $PM_{2.5}$; a reduction in NO_2 may, therefore, lead to a decline in $PM_{2.5}$. However, as data on $PM_{2.5}$ composition is insufficient, we cannot quantify the contribution of NO_2 to the reductions in $PM_{2.5}$ concentrations.[7] Therefore we treat both pollutants independently.

We estimate a population-weighted version of Equation 3 in Methods and report results in Table 14. Results of unweighted regressions are qualitatively similar and can be found in Supplementary information. From February 22 to May 4, the lockdown has on average reduced daily concentrations of $PM_{2.5}$ and NO_2 by 5.32 $\mu g/m^3$ and 13.56 $\mu g/m^3$. That is a reduction of 21.8% and 35.6%, respectively, from the average levels that would have been observed had not the epidemic broken out.

Next, our preferred specification distinguishes effects of the lockdown by type of monitor. Background monitors are located where concentrations are representative of the ambient exposure of the general population; industrial monitors are located in the proximity of industrial sites or industrial sources; traffic monitors are located near a major road.

Population-weighted average background concentrations of $PM_{2.5}$ decreased by 3.84 $\mu g/m^3$ from 24.42 $\mu g/m^3$ (Table 15).[8] The reduction was almost twice as large in monitored industrial sites and near major roads. Background concentrations of NO_2 dropped by 10.85 $\mu g/m^3$ from 33.22 $\mu g/m^3$, by 10.66 $\mu g/m^3$ near monitored industrial sites and by 15.85 $\mu g/m^3$ more at major roads.

7 At the time of writing, data on composition of $PM_{2.5}$ has not been released. Data on composition of PM_{10} is available only for 3 monitoring station.

8 The very low number of monitors by type makes clustered standard errors inappropriate. We thus use robust standard errors.

Table 14: Population-weighted regression

	$\Delta_{Observed,Counterfactual}$	
	(1)	(2)
	PM 2.5	NO2
Lockdown	-5.32***	-13.56***
	(1.08)	(1.21)
Constant	0.73	2.59
	(1.37)	(1.67)
Average baseline concentration	24.39	38.14
Observations	3555	10084

Notes: Regression weighted by population within 20 kilometers of a monitoring station. Territory within less than 20 kilometers from two or more monitors is assigned to the closest one. The dependent variable is the difference between the observed values and the counterfactual. *Lockdown* is a dummy variable equal to 0 from January 1, 2020 to February 22, and equal to 1 after February 22, 2020. *Average baseline concentration* is the population-weighted average of counterfactual values during the lockdown, less the constant in case the latter is statistically significant at 10%. Standard errors, in brackets, are clustered by monitor. * p<0.1, ** p<0.05, *** p<0.01.

Table 15: Heterogeneous effects by type of monitoring station

	$\Delta_{Observed,Counterfactual}$					
	PM 2.5			NO2		
	Background	Industrial	Traffic	Background	Industrial	Traffic
Lockdown	-3.84***	-7.39***	-7.28***	-10.85***	-10.66***	-15.85***
	(0.97)	(1.54)	(1.20)	(0.64)	(0.96)	(0.75)
Constant	-1.26	5.18***	2.79**	0.21	7.29***	4.04***
	(0.84)	(1.37)	(1.07)	(0.49)	(0.84)	(0.63)
Average baseline concentration	24.42	27.99	27.77	33.22	31.93	46.67
Number of monitors	18	2	10	53	6	24
Observations	2117	244	1194	6483	731	2870

Notes: Regression weighted by population within 20 kilometers of a monitoring station. Territory within less than 20 kilometers from two or more monitors is assigned to the closest one. The dependent variable is the difference between the observed values and the counterfactual. *Lockdown* is a dummy variable equal to 0 from January 1, 2020 to February 22, and equal to 1 after February 22, 2020. *Average baseline concentration* is the population-weighted average of counterfactual values during the lockdown, less the constant in case the latter is statistically significant at 10%. Robust standard errors are in brackets. * p<0.1, ** p<0.05, *** p<0.01.

4.3 Human health benefits

As the reduction in road transport and the slowing of economic activity reduced toxic emissions, the burden of pollutants on human health eased. For calculations, we use the estimated change in concentrations at background stations. Avoided deaths and YLS should be considered a lower-bound estimate of total health benefits avoided deaths.

The reduction in $PM_{2.5}$ prevented 10.2 to 24.8 premature deaths per 100,000 individuals and saved 72.1 to 175.9 years of life per 100,000 individuals, depending on the concentration-response function (Table 16). The reduction in NO_2 prevented 28.8 premature deaths and saved 203.7 years of life per 100,000 individuals. Given the high correlation between concentrations of $PM_{2.5}$ and NO_2, the concentration-response function of these pollutants are interdependent. It is recommended that avoided deaths and YLS be not aggregated across pollutants, lest incurring in partial double counting.

As a comparison, in Italy in 2016 for every 100,000 individuals, there have been 96.6 premature deaths attributable to PM2.5 and 24.1 attributable to NO_2, or 23.8 and 5.9 premature deaths in three months, respectively (European Environment Agency, 2019). Since most of the premature deaths happen in the more polluted North of Italy, including Lombardy, the lockdown has temporarily reduced the cost of pollution by a substantial amount.

We compare the results against the number of deaths and the years of life lost (YLL) related to COVID-19 in Lombardy during the same period, computed from patient-level data.[9] In Lombardy, from February 22 to May 3 2020, every 100,000 people 155 died after testing positive for COVID-19 and 1891 years of life have been directly lost to the virus. Avoided deaths from the reduction in $PM_{2.5}$ are 6.5% to 16% of COVID-19 deaths; YLS are 3.8% to 9.3% of YLL to COVID-19. Avoided deaths from the reduction in NO_2 are 18.6% of COVID-19 deaths; YLS are 10.8% of YLL to COVID-19.

9 Data on the individual COVID-19 patients has been shared by regional health officers under an institutional agreement.

Table 16: Avoided premature deaths and years of life saved per 100,000 in Lombardy due to improved air quality during lockdown.

	Pollutant	Source of HR	Hazard ratio	Avoided deaths
Avoided deaths	NO2	EEA/WHO	1.055	28.8
	PM 2.5	EEA/WHO	1.062	11.3
	PM 2.5	Krewski et al. (2009)	1.056	10.2
	PM 2.5	Lepeule et al. (2012)	1.14	24.8
Years of life saved	NO2	EEA/WHO	1.055	203.7
	PM 2.5	EEA/WHO	1.062	79.7
	PM 2.5	Krewski et al. (2009)	1.056	72.1
	PM 2.5	Lepeule et al. (2012)	1.14	175.9

In Lombardy, from February 22 to May 3 2020, every 100,000 people 155 died after testing positive for COVID-19 and 1891 years of life have been directly lost to the virus. The hazard ratio is the ratio of two concentration-response functions, or hazard rates, between a high and a low concentration differing by 10 $\mu g/m^3$. Avoided premature deaths are calculated using the population-weighted change in concentrations at background stations.

5 CONCLUSIONS

The dramatic reduction in emissions of airborne pollutants that has come with the response to COVID-19 provides a unique natural experiment to assess the sensitivity of pollutants concentrations and health to emissions. We estimate a substantial yet partial improvement in air quality in Lombardy following the outbreak, and suggest that the improvement originates primarily from the reduction of road transport; and to a lesser degree from the reduction in industrial activity. Important sources of emissions as heating systems and agriculture have not been substantially affected by the outbreak.

The methodology used to build the counterfactual does not require identifying comparable but unaffected regions, but relies on the assumption of emissions absent the lockdown following historical variation around the mean. The approach is not limited to this case study, but can be applied in a variety of settings due to the increasing and reliable availability of pollution and weather data.

Finally, we are nowhere near suggesting the pandemic has been beneficial for the affected communities, yet the health benefits from improved air quality are noticeable. While global pandemics are rare phenomena, exposure to unhealthy levels of toxic air pollutants is the rule, including in affluent regions of the world such as the one considered here. This paper has emphasized some of the health benefits of cleaner air, but also highlighted the variety of emissions sources and the need for a broader policy response to solve Europe's biggest environmental health risk.

REFERENCES

Pope, C. Arden II and Douglas W. Dockery (June 2006). "Health Effects of Fine Particulate Air Pollution: Lines that Connect". In: *Journal of the Air & Waste Management Association* 56.6, pp. 709–742. ISSN: 1096-2247. DOI: 10.1080/10473289.2006.10464485. URL: https://doi.org/10.1080/10473289.2006.10464485.

Ebenstein, Avraham, Maoyong Fan, et al. (Sept. 2017). "New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy". en. In: *Proceedings of the National Academy of Sciences* 114.39, pp. 10384–10389. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1616784114. URL: https://www.pnas.org/content/114/39/10384 (visited on 05/13/2020).

Deryugina, Tatyana et al. (Dec. 2019). "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction". en. In: *American Economic Review* 109.12, pp. 4178–4219. ISSN: 0002-8282. DOI: 10.1257/aer.20180279. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20180279&&from=f.

Graff Zivin, Joshua and Matthew Neidell (Dec. 2012). "The Impact of Pollution on Worker Productivity". en. In: *American Economic Review* 102.7, pp. 3652–3673. ISSN: 0002-8282. DOI: 10.1257/aer.102.7.3652. URL: https://www.aeaweb.org/articles?id=10.1257/aer.102.7.3652.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth (Oct. 2016). "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution". In: *American Economic Journal: Applied Economics* 8.4, pp. 36–65. DOI: 10.1257/app.20150213. URL: http://www.aeaweb.org/articles?id=10.1257/app.20150213.

Zhang, Xin, Xi Chen, and Xiaobo Zhang (Sept. 2018). "The impact of exposure to air pollution on cognitive performance". en. In: *Proceedings of the National Academy of Sciences* 115.37, pp. 9193–9197. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1809474115. URL: https://www.pnas.org/content/115/37/9193.

He, Jiaxiu, Haoming Liu, and Alberto Salvo (Jan. 2019). "Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China". en. In: *American Economic Journal: Applied Economics* 11.1, pp. 173–201. ISSN: 1945-7782. DOI: 10.1257/app.20170286. URL: https://www.aeaweb.org/articles?id=10.1257/app.20170286.

Kahn, Matthew E. and Pei Li (July 2020). "Air pollution lowers high skill public sector worker productivity in China". en. In: *Environmental Research Letters* 15.8, p. 084003. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab8b8c. URL: https://doi.org/10.1088%2F1748-9326%2Fab8b8c (visited on 10/12/2020).

Mills, Inga C et al. (2015). "Quantitative systematic review of the associations between short-term exposure to nitrogen dioxide and mortality and hospital admissions". In: *BMJ open* 5.5.

Amini, Heresh et al. (2019). "Short-term associations between daily mortality and ambient particulate matter, nitrogen dioxide, and the air quality index in a Middle Eastern megacity". In: *Environmental Pollution* 254, p. 113121.

Duan, Yanran et al. (2019). "Effect of changes in season and temperature on cardiovascular mortality associated with nitrogen dioxide air pollution in Shenzhen, China". In: *Science of The Total Environment* 697, p. 134051.

Kroll, Jesse H. et al. (Aug. 2020). "The complex chemical effects of COVID-19 shutdowns on air quality". In: *Nature Chemistry* 12.9, pp. 777–779. DOI: 10.1038/s41557-020-0535-z.

Guo, Song et al. (2014). "Elucidating severe urban haze formation in China". In: *Proceedings of the National Academy of Sciences* 111.49, pp. 17373–17378. ISSN: 0027-8424. DOI: 10.1073/pnas.1419604111. eprint: https://www.pnas.org/content/111/49/17373.full.pdf. URL: https://www.pnas.org/content/111/49/17373.

He, Guojun, Yuhang Pan, and Takanao Tanaka (July 2020). "The short-term impacts of COVID-19 lockdown on urban air pollution in China". In: *Nature Sustainability*. DOI: 10.1038/s41893-020-0581-y.

Cole, Matthew A., Robert J. R. Elliott, and Bowen Liu (May 2020). *The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach*. en. Tech. rep. Department of Economics, University of Birmingham. URL: https://ideas.repec.org/p/bir/birmec/20-09.html (visited on 05/04/2020).

Petetin, Hervé et al. (2020). "Meteorology-normalized impact of the COVID-19 lockdown upon NO 2 pollution in Spain". In: *Atmospheric Chemistry and Physics* 20.18, pp. 11119–11141.

Friedman, Jerome H (2001). "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics*, pp. 1189–1232.

Silver, Ben et al. (July 2020). "The impact of COVID-19 control measures on air quality in China". In: *Environmental Research Letters* 15.8, p. 084021. DOI: 10.1088/1748-9326/aba3a2. URL: https://doi.org/10.1088%2F1748-9326%2Faba3a2.

European Commission v. Italian Republic (Dec. 2012). *Judgment of the Court (First Chamber), 19 December 2012, C-68/11, ECLI:EU:C:2012:815, Court of Justice of the European Union*.

– (2019). *Application of 26 July 2019, C-573/19, Court of Justice of the European Union*.

– (Nov. 2020). *Judgment of the Court (Grand Chamber) of 10 November 2020, C-644/18, ECLI:EU:C:2020:895, Court of Justice of the European Union*.

Lelieveld, J. et al. (Sept. 2015). "The contribution of outdoor air pollution sources to premature mortality on a global scale". In: *Nature* 525.7569, pp. 367–371. DOI: 10.1038/nature15371.

Google (2020). *COVID-19 Community Mobility Reports*. Data retrieved from https://www.google.com/covid19/mobility/. Last accessed on June 27, 2020.

Apple (2020). *COVID-19 Mobility Trends Reports*. Data retrieved from https://www.apple.com/covid19/mobility. Last accessed on September 22, 2020.

INEMAR (2017). *INEMAR - Inventario Emissioni Aria*. Data retrieved from http://inemar.arpalombardia.it/inemar/webdata/main.seam?cid=22157. URL: https://www.inemar.eu/xwiki/bin/view/Inemar/WebHome.

TERNA (2020). *Transparency Report*. Data retrieved from https://www.terna.it/it/sistema-elettrico/transparency-report/download-center. Last accessed on June 28, 2020.

Chen, Tianqi and Carlos Guestrin (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Fowlie, Meredith, Edward Rubin, and Reed Walker (May 2019). "Bringing Satellite-Based Air Quality Estimates Down to Earth". In: *AEA Papers and Proceedings* 109, pp. 283–88. DOI: 10.1257/pandp.20191064. URL: https://www.aeaweb.org/articles?id=10.1257/pandp.20191064.

Krewski, Daniel et al. (2009). *Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality*. 140. Health Effects Institute Boston, MA. URL: https://www.healtheffects.org/system/files/Krewski140.pdf.

Lepeule, Johanna et al. (2012). "Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009". In: *Environmental health perspectives* 120.7, pp. 965–970. URL: https://dash.harvard.edu/bitstream/handle/1/10436317/3404667.pdf?sequence=1&isAllowed=y.

Henschel, Susann, Gabrielle Chan, World Health Organization, et al. (2013). "Health risks of air pollution in Europe-HRAPIE project: New emerging risks to health from air pollution-results from the survey of experts". In.

European Environment Agency (2019). *Air quality in Europe - 2019 report*.

A SUPPLEMENTARY MATERIAL

A.1 Years of life saved

Concentration-response functions are typically estimated with log-linear regressions of mortality risk on pollutants of the form $\ln(y) = \alpha + \beta C$, so that $y = Ae^{\beta C}$. The change in mortality risk from y' to y'' is

$$\begin{aligned} y' - y'' &= A(e^{\beta C'} - e^{\beta C''}) \\ &= Ae^{\beta C'}(1 - e^{\beta(C''-C')}) \\ &= y'(1 - \frac{1}{e^{\beta(C'-C'')}}) \end{aligned}$$

with $A = e^{\alpha}$. Here y' is the baseline mortality risk and $e^{\beta(C'-C'')}$ is the RR. The β coefficient is not typically reported, but is easily found as $\beta = \ln(RR)/10$.

For each gender g and age group a above 30, we multiply the change in mortality risk from the baseline by the number of individuals in Lombardy of that gender and age group ($N_{g,a}$).[10] This gives us the number of avoided deaths for a year-long reduction in pollutants. We then multiply this number by gender- and age-specific life expectancy to obtain the YLS.

$$\begin{aligned} \text{Avoided Deaths}_{g,a} &= y'_{g,a} \cdot (1 - \frac{1}{e^{\beta(C'-C'')}}) \cdot N_{g,a} \cdot \frac{1}{6} \\ \text{YLS}_{g,a} &= \text{Avoided Deaths}_{g,a} \cdot \text{Life Expectancy}_{g,a} \\ \text{YLS} &= \sum_g \sum_a \text{YLS}_{g,a} \end{aligned}$$

It should be noted that we are assuming that avoided deaths and years of life saved by a two-month improvement in air quality are equivalent to a sixth of the benefits of a year-long improvement. In addition, we assume that the gains are linear in reductions of concentrations.[11]

Gender- and age-group specific baseline mortality risk, population size and life expectancy come from mortality tables for Lombardy compiled by the Italian National Statistical Office (ISTAT). Avoided deaths and YLS are computed using the lockdown on pollution $(C' - C'')$ estimated at background stations.

10 The benefits from reductions in NO_2 are set to zero for values below 20 $\mu g/m^3$, as recommended by Henschel, Chan, Organization, et al., 2013.
11 This is in line with Henschel, Chan, Organization, et al., 2013, who recommend a linear concentrations-response function.

A.2 Accuracy of linear regression for construction of counterfactuals

We show a linear regression model does not perform as well as the machine learning algorithm used for the main results. For every monitoring station, we regress daily concentrations on a vector of daily weather summaries, namely daily cumulative precipitation, average temperature, average wind speed and average wind direction, in 2012 through 2019 (Equation 4). We then use the estimated coefficients to predict concentrations in 2020 before and throughout the lockdown (Equation 5). Finally, we assess the accuracy of predictions during the pre-lockdown period from January 1 to February 21, 2020. Precipitation, temperature, wind speed and direction on day t at any given pollution monitor are interpolated with inverse distance weight from the three closest weather stations within 0.2 degrees from the monitor.

$$y_{t_{2012-2019}} = \alpha + \beta' Weather_{t_{2012-2019}} + \epsilon_{t_{2012-2019}} \tag{4}$$

$$\hat{y}_{t_{2020}} = \hat{\alpha} + \hat{\beta}' Weather_{t_{2020}} \tag{5}$$

Observed and predicted population-weighted average concentrations are displayed in Figure 12. While approximating pre-lockdown values on average, the predictions fail to capture a non-negligible portion of the variability. The validity of predictions based on linear regressions is especially poor for $PM_{2.5}$. The same conclusions can be drawn examining average accuracy measures for linear regression predictions in Table 17.

Table 17: Accuracy of liner regression predictions, average values across monitors

Pollutant	Dataset	Corr	MB	nMB	RMSE	cRMSE	ncRMSE
NO2	Train	0.71	0	0	9.7	9.7	0.33
NO2	Test	0.7	-5.09	-0.16	13.22	11.45	0.37
PM2.5	Train	0.63	0	0	12.21	12.21	0.53
PM2.5	Test	0.59	0.35	0.01	14.43	14.21	0.5

Notes: Corr: Pearson's correlation coefficient. *MB*: Mean bias, where negative values indicate observed values below predicted values. *nMB*: Normalized mean bias. *RMSE*: Root mean squared error. *nRMSE*: Normalized RMSE. *cRMSE*: Centered RMSE. *ncRMSE*: Normalized centered RMSE. Mean bias, RMSE and centered RMSE are expressed in $\mu g/m^3$. Mean bias, RMSE and centered RMSE are normalized dividing by mean observed concentrations. The centered RMSE is computed as $\left[1/N \sum(\hat{y}_i - \bar{\hat{y}} - y_i + \bar{y})^2\right]^{1/2}$.

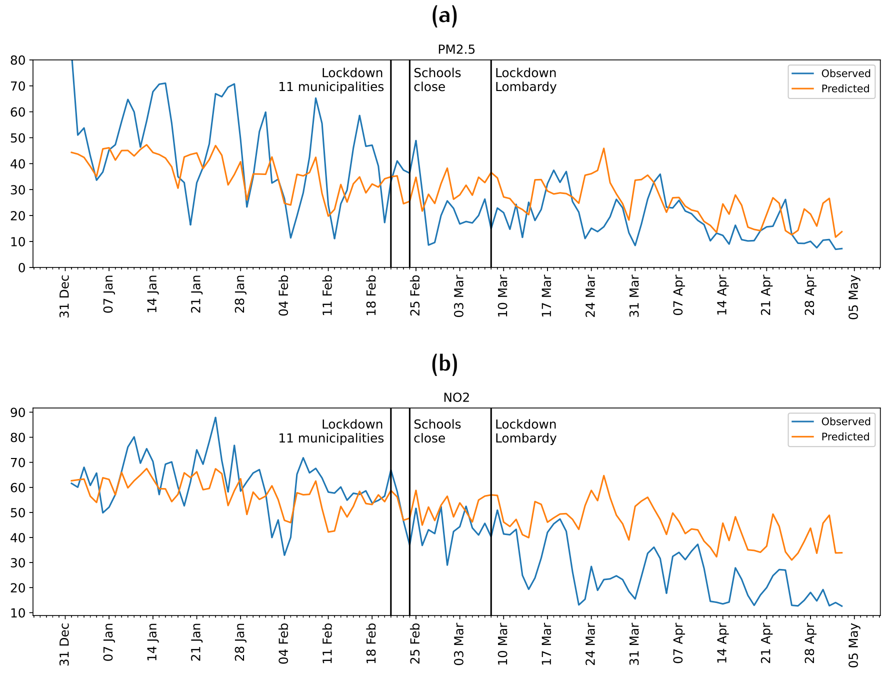


Figure 12: **Population-weighted average of observed and counterfactual values built with linear regression models.** **a**, $PM_{2.5}$. **b**, NO_2. Population is measured within 20 kilometers of a monitoring station. Territory within less than 20 kilometers from two or more monitors is assigned to the closest one.

A.3 Supplementary tables

Table 18: Pollution monitors by type.

Pollutant	Type of monitor	Number of municipalities	Number of monitors
NO2	Background	50	53
NO2	Industrial	6	6
NO2	Traffic	20	24
PM2.5	Background	18	18
PM2.5	Industrial	2	2
PM2.5	Traffic	10	10

Background stations measure pollutions concentrations that are representative of the average exposure of the general population, or vegetation. Industrial stations are located in close proximity to an industrial area or an industrial source. Traffic stations are located in close proximity to a single major road.

Table 19: Unweighted regression

	$\Delta_{Observed,Counterfactual}$	
	(1)	(2)
	PM 2.5	NO2
Lockdown	-4.37***	-9.19***
	(0.41)	(0.65)
Constant	1.19**	0.73
	(0.47)	(0.53)
Average baseline concentration	25.58	38.14
Observations	3555	10084

Notes: Unweighted regression. The dependent variable is the difference between the observed values and the counterfactual. *Lockdown* is a dummy variable equal to 0 from January 1, 2020 to February 22, and equal to 1 after February 22, 2020. *Average baseline concentration* is the average of counterfactual values during the lockdown, less the constant in case the latter is statistically significant at 10%. Standard errors, in brackets, are clustered by monitor. * p<0.1, ** p<0.05, *** p<0.01.

Table 20: Heterogeneous effects by type of monitoring station - unweighted regression

	$\Delta_{Observed,Counterfactual}$					
	PM 2.5			NO2		
	Background	Industrial	Traffic	Background	Industrial	Traffic
Lockdown	-3.70***	-7.63***	-4.92***	-7.53***	-7.50***	-13.39***
	(0.39)	(1.33)	(0.53)	(0.19)	(0.58)	(0.39)
Constant	0.79*	5.10***	1.11*	-0.05	2.89***	1.98***
	(0.34)	(1.20)	(0.46)	(0.15)	(0.48)	(0.31)
Average baseline concentration	25.21	27.91	26.09	33.22	27.53	44.61
Number of monitors	18	2	10	53	6	24
Observations	2117	244	1194	6483	731	2870

Notes: Unweighted regression. The dependent variable is the difference between the observed values and the counterfactual. *Lockdown* is a dummy variable equal to 0 from January 1, 2020 to February 22, and equal to 1 after February 22, 2020. *Average baseline concentration* is the average of counterfactual values during the lockdown, less the constant in case the latter is statistically significant at 10%. Robust standard errors are in brackets. * p<0.1, ** p<0.05, *** p<0.01.

A.4 Supplementary figures

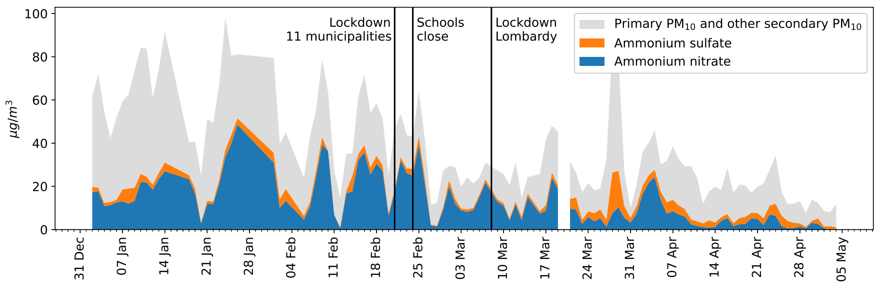


Figure 13: Composition of background PM_{10} in Milan, Lombardy. *Source:* ARPA Lombardia.

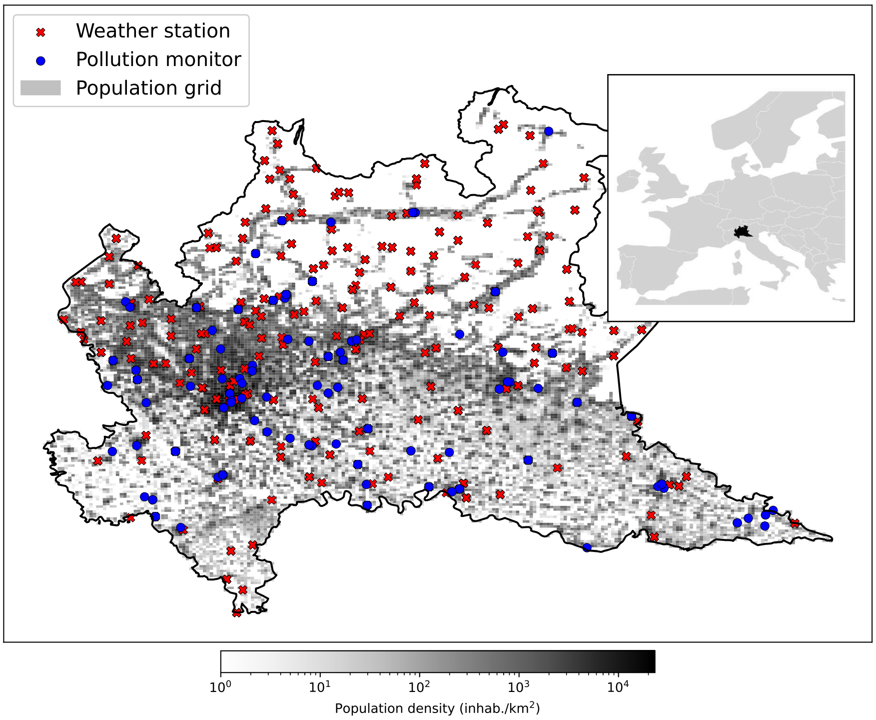
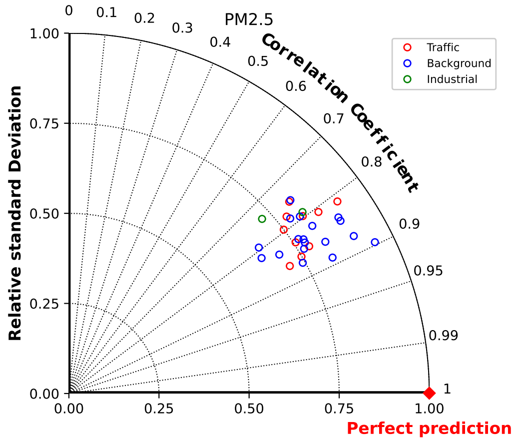


Figure 14: Location of pollution monitors and weather stations in Lombardy over a 1 km by 1 km population grid.

(a) $PM_{2.5}$



(b) NO_2

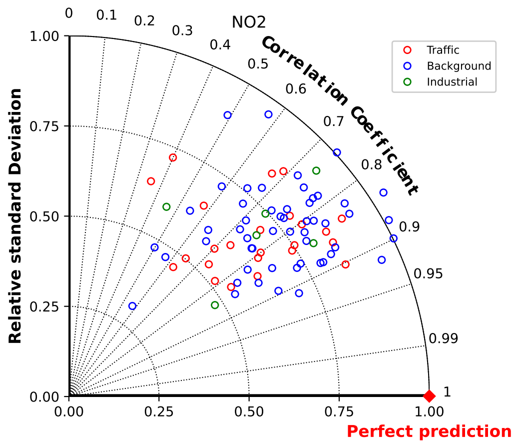


Figure 15: Taylor diagrams are a practical way to display different dimensions of model predictive performance (Taylor, 2001) . Each circle represents the prediction of a model, that is, in this case, a monitoring station for the pre-lockdown period from January 1 to February 22. Isocurves from the origin outward measure the standard deviation of a model's predictions relative to the standard deviation of the observed values. The azimut measures Pearson's correlation coefficient. The ideal model prediction has a relative standard deviation of 1 and a correlation coefficient of 1, and is marked by the red diamond. We do not show the RMSE, as is practice in Taylor diagrams, because it is graphically incompatible with the relative standard deviation.

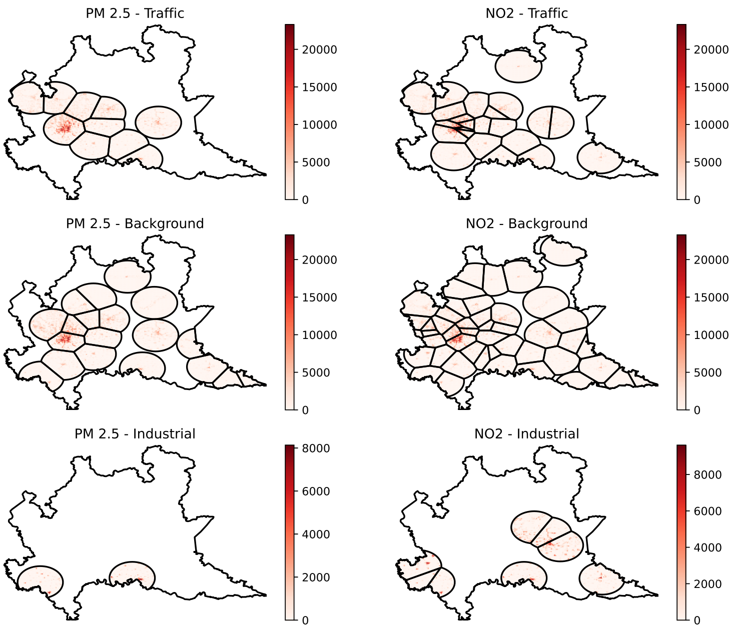


Figure 16: Each polygon circumscribes the territory nearest to a monitor and within 20 kilometers from it. Color represents population in grid cells of 1 km^2.

A.5 Lockdown of Lombardy

Italy has witnessed one of the first major outbreak of COVID-19 outside China. The virus has first been identified in two Chinese tourists who had arrived at Milano Malpensa Airport and, on January 31st, tested positive for the virus when visiting Rome (ANSA, 2020a). For the next three weeks, only a handful of cases had been identified and all had a direct link with known hot-spots, such as a student returning from vacation in Wuhan and a couple of tourists from Taiwan (ANSA, 2020b; ANSA, 2020c).

However, on February 21st, the first non-imported cases and the first death related to COVID-19 in the country were confirmed in lower Lombardy. By the end of the day, 17 individuals had been tested positive, 15 of which in Lodi and surroundings, in lower Lombardia, and 2 in the neighboring region of Veneto. The largest hotspot had been identified in the hospital of Codogno, where 5 members of the medical staff and 3 patients had tested positive to COVID-19. On the same day, the Minister of Health announced severely restrictive measures on 11 municipalities and over 50 000 people. Until further notice, schools and all public and sporting events were suspended; non-essential production, commercial activities and public offices had to close doors; self-isolation at home was mandated and enforced; access to the municipalities was monitored by police and armed forces (Presidente del Consiglio dei Ministri, 2020a; ANSA, 2020d; La Repubblica, 2020; Guidelli, 2020) Also, self-isolation for two weeks was mandatory for whoever in the country had had contacts with confirmed cases. Violations of lockdown areas and self-isolation could be sanctioned with fines and up to a three months prison sentence (Presidente della Repubblica, 2020; Ministro della Salute, 2020a).

Over the next two days, local governments all over the country imposed restrictions of heterogeneous degrees, with strictest measures in the regions of Lombardia and Veneto. In Lombardia, the regional government suspended all teaching activities in schools and universities, prohibited public events, and suspended religious gatherings; pubs had to close by 6 pm (Ministro della Salute, 2020b).

Local measures were soon followed by the intervention of the central government. On February 25th, the Prime Minister signed a Law Degree to expand and incorporate containment efforts in hotspot regions of Northern Italy. The decree closed schools and universities (originally until March 15th) and recommended remote working in Emilia Romagna, Friuli Venezia Giulia, Lombardia, Veneto, Liguria, and Piemonte (Presidente del Consiglio dei Ministri, 2020b).

A week later, on March 1st, the government extends previous measures and prescribes non-restrictive ones over non-affected regions(Presidente del Consiglio dei Ministri, 2020c); on March 4th, it announces all schools and universities in the countries will close.

By March 7th, 5883 cases had been confirmed in Italy and 233 COVID-19-related deaths recorded (Protezione Civile, 2020). Despite containment measures, the growing number of confirmed cases and deaths pressured the Italian government to impose stricter controls. With a Law Decree on March 8th, Italy became the first country in Europe to impose a lockdown over Lombardia and 14 provinces of the northern and central regions of Piemonte, Emilia-Romagna, Veneto, and Marche. The restrictive measures were soon extended to the rest of the country on the following day. The decree imposed compulsory social distancing and self-isolation at home and the halt of all non-essential economic activities (Presidente del Consiglio dei Ministri, 2020d; Presidenza del Consiglio dei Ministri, 2020).

The list of sectors and activities deemed essential had been furthered narrowed on March 23rd; most notably, construction works were stopped, and all public offices had to close (Presidente del Consiglio dei Ministri, 2020e). The lockdown then continued under virtually unaltered conditions until May 3rd.

A.6 Averaging wind speed and direction

Consider two vectors $s' = [s_1, \ldots, s_h, \ldots, s_{24}]$ and $d' = [d_1, \ldots, d_h, \ldots, d_{24}]$ containing hourly data on wind speed and direction, respectively. Speed and direction at hour h are s_h and d_h. To calculate average wind speed and average wind direction we:

1. Convert wind direction from degrees to radians
   $r = d \cdot \pi/180$

2. Calculate the average of East-West and North-South speed components and invert sign.
   $\bar{s}^{EW} = -\frac{1}{24} \sum s_i \cdot \sin(r_i)$
   $\bar{s}^{NS} = -\frac{1}{24} \sum s_i \cdot \cos(r_i)$

3. Calculate average wind speed
   $S = \sqrt{\bar{s}^{EW^2} + \bar{s}^{EW^2}}$

4. Calculate average wind direction
   $\bar{r} = \arctan2(\bar{s}^{NS}, \bar{s}^{EW})$

5. Convert radians to degrees
   $\bar{d} = \bar{r} \cdot 180/\pi$

$$D = \begin{cases} \bar{d} + 180 & \text{if } \bar{d} < 180 \\ \bar{d} & \text{if } \bar{d} = 0 \\ \bar{d} - 180 & \text{if } \bar{d} > 180. \end{cases}$$

   D is the average wind direction, and S is the average wind speed.

A.7 Effects on economic activity and morbidity

We report here an estimate of the theoretical gains in GDP and lost workdays from air pollution-related illness due to the improvement in air quality, *but absent the pandemic*.

To calculate the aggregated productivity gains, we employ the results of Dechezleprêtre, Rivers, and Stadler, 2019, who use thermal inversions to identify the causal impact of air pollution on economic activity. They estimate that a one $\mu g/m^3$ increase in $PM_{2.5}$ concentration leads to a 0.8% decrease in regional annual GDP. Accordingly, the average reduction of $PM_{2.5}$ by 3.84 $\mu g/m^3$ for two months corresponds, for simplicity ignoring the exponential growth process, to $3.84 \cdot 0.8/6 = +0.512\%$ in regional annual GDP.

We compute the number of lost workdays from air pollution-related illnesses as in Vandyck et al., 2018. They assume a fixed ratio of 547 avoided lost workdays per avoided premature mortality. The multiplier was derived from the WHO-HRAPIE recommendations, based on earlier work, and applied in the context of the EU Clean Air Package. Following this methodology, we calculate that 5579.4 to 13565.8 lost workdays have been avoided by reducing $PM_{2.5}$ concentrations; and 15753.6 by the decrease in NO_2 concentrations.

REFERENCES

Henschel, Susann, Gabrielle Chan, World Health Organization, et al. (2013). "Health risks of air pollution in Europe-HRAPIE project: New emerging risks to health from air pollution-results from the survey of experts". In.

Taylor, Karl E (2001). "Summarizing multiple aspects of model performance in a single diagram". In: *Journal of Geophysical Research: Atmospheres* 106.D7, pp. 7183–7192.

ANSA (2020a). "Coronavirus: coppia Taiwan in Toscana dal 26 al 29 gennaio". In: https://www.ansa.it/toscana/notizie/2020/02/09/coronavirus-coppia-taiwan-in-toscana-dal-26-al-29-febbraio_99d50b7f-62fc-4822-ba9d-cbc9e447ebe1.html. Last accessed on July 28, 2020.

– (Feb. 2020b). "Coronavirus: contagiato è reggiano, era in Cina in vacanza". In: https://www.ansa.it/emiliaromagna/notizie/2020/02/07/coronavirusitaliano-in-buone-condizioni_a8e7dccc-ad92-4fe7-ad06-6e5d00d9c57a.html. Last accessed on July 28, 2020.

– (2020c). "Coronavirus: coppia Taiwan in Toscana dal 26 al 29 gennaio". In: https://www.ansa.it/toscana/notizie/2020/02/09/coronavirus-coppia-taiwan-in-toscana-dal-26-al-29-

febbraio_99d50b7f-62fc-4822-ba9d-cbc9e447ebe1.html. Last accessed on July 28, 2020.

Presidente del Consiglio dei Ministri (Feb. 2020a). *Decreto del Presidente del Consiglio dei Ministri 25 febbraio 2010. Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19.* Gazzetta Ufficiale Serie Generale n.47 del 25-02-2020. URL: https://www.gazzettaufficiale.it/eli/id/2020/02/25/20A01278/sg.

ANSA (2020d). "Coronavirus, sedici nuovi contagiati. Due sono in Veneto. Altri 8 casi a Codogno, 5 operatori sanitari". In: https://www.ansa.it/lombardia/notizie/2020/02/21/coronavirus-sedici-nuovi-contagiati.-due-sono-in-veneto.-altri-8-casi-a-codogno-5-operatori-sanitari_c7a81b85-4370-46b8-a2db-9a7a4df05d90.html. Last accessed on July 28, 2020.

La Repubblica (2020). "Coronavirus, in dieci comuni lombardi: 50 mila persone costrette a restare in casa. Quarantena all'ospedale milanese di Baggio". In: https://milano.repubblica.it/cronaca/2020/02/21/news/coronavirus_codogno_castiglione_d_adda_contagiati_misure_sicurezza-249154447/. Last accessed on July 28, 2020.

Guidelli, Matteo (2020). "Coronavirus in Italia: scattato l'isolamento dei focolai, 43 varchi e 500 uomini". In: *ANSA.it*. https://www.ansa.it/canale_saluteebenessere/notizie/sanita/2020/02/23/coronavirus-in-italia-aumentano-i-contagi.-scuole-chiuse-in-lombardia-e-veneto.-stop-al-carnevale-di-venezia_40e888b6-8418-406a-9771-7a6c3c0b45fd.html. Last accessed on July 28, 2020.

Presidente della Repubblica (2020). *Decreto-legge 23 febbraio 2020, n. 6 Misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19 (20G00020). Decreto-Legge convertito con modificazioni dalla L. 5 marzo 2020, n. 13 (in G.U. 09/03/2020, n. 61).* Gazzetta Ufficiale Serie Generale n.45 del 23-02-2020.

Ministro della Salute (2020a). *Ordinanza 21 febbraio 2020 Ulteriori misure profilattiche contro la diffusione della malattia infettiva COVID-19. (20A01220).* Gazzetta Ufficiale Serie Generale n.44 del 22-02-2020.

– (2020b). *Ordinanza 23 febbraio 2020 Misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19. Regione Lombardia. (20A01273).* Gazzetta Ufficiale Serie Generale n.47 del 25-02-2020.

Presidente del Consiglio dei Ministri (2020b). *Decreto del Presidente del Consiglio dei Ministri 25 febbraio 2020 Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19. (20A01278).* Gazzetta Ufficiale Serie Generale n.47 del 25-02-2020.

– (2020c). *Decreto del Presidente del Consiglio dei Ministri 1 marzo 2020 Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza* *epidemiologica da COVID-19. (20A01381)*. Gazzetta Ufficiale Serie Generale n.52 del 01-03-2020.

Protezione Civile (2020). *Dati COVID-19 Italia - Public repository of the Protezione Civile. Last accessed on July 28, 2020.* URL: https://github.com/pcm-dpc/COVID-19.

Presidente del Consiglio dei Ministri (2020d). *Decreto del Presidente del Consiglio dei Ministri 8 marzo 2020 Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19. (20A01522)*. Gazzetta Ufficiale Serie Generale n.59 del 08-03-2020.

Presidenza del Consiglio dei Ministri (2020). *IoRestoaCasa, misure per il contenimento e gestione dell'emergenza epidemiologica*. http://www.governo.it/it/iorestoacasa-misure-governo.

Presidente del Consiglio dei Ministri (2020e). *Decreto del Presidente del Consiglio dei Minsitri 22 marzo 2020 Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19, applicabili sull'intero territorio nazionale. (20A01807)*. Gazzetta Ufficiale Serie Generale n.76 del 22-03-2020.

Dechezleprêtre, Antoine, Nicholas Rivers, and Balazs Stadler (2019). *The economic cost of air pollution: Evidence from Europe*. Tech. rep. OECD Publishing.

Vandyck, Toon et al. (Nov. 2018). "Air quality co-benefits for human health and agriculture counterbalance costs to meet Paris Agreement pledges". In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-06885-9.

3 PERSISTENT EFFECT OF TEMPERATURE ON GDP IDENTIFIED FROM LOWER FREQUENCY TEMPERATURE VARIABILITY

with
Bernardo A. Bastien-Olvera, *University of California, Davis*
Frances C. Moore, *University of California, Davis*

Published on *Environmental Research Letters*.

ABSTRACT It is well established that temperature variability affects a range of outcomes relevant to human welfare, including health, emotion and mood, and productivity across a number of economic sectors. However, a critical and still unresolved empirical question is whether temperature variation has a long-lasting effect on economic productivity and, therefore, whether damages compound over time in response to long-lived changes in temperature expected with climate change. Several studies have identified a relationship between temperature and GDP, but empirical evidence as to the persistence of these effects is still weak. This paper presents a novel approach to isolate the persistent component of temperature effects on output using lower frequency temperature variation. The effects are heterogeneous across countries but collectively, using three different GDP datasets, we find evidence of persistent effects, implying temperature affects the determinants of economic growth, not just economic productivity. This, in turn, means that the aggregate effects of climate change on GDP may be far larger and far more uncertain than currently represented in integrated assessment models used to calculate the social cost of carbon.

1 INTRODUCTION

A large body of evidence now exists showing a relationship between temperature fluctuations and economic productivity (Burke, S. M. Hsiang, and Miguel, 2015; Dell, Jones, and Olken, 2012; Kalkuhl and Wenz, 2020; Carleton and S. M. Hsiang, 2016). Temperature has been shown to influence output at global (Burke, S. M. Hsiang, and Miguel, 2015; Dell, Jones, and Olken, 2012), national (Deryugina and S. Hsiang, 2017; Schlenker and Roberts, 2009), and regional scales (Kalkuhl and Wenz, 2020), affecting a wide range of sectors in both high-income and low-income countries. The persistence of these impacts has first-order implications for the magnitude of climate change damages: if temperature fluctuations affect the determinants of economic growth (e.g., depreciation of capital or the total factor productivity growth rate), then they have a persistent impact on the level of economic output. In this case, climate change damages are cumulative and may be orders of magnitude larger than currently represented in models used for the cost-benefit analysis of climate change, which mostly assume non-persistent damages (for example, when temperature variations affect the productivity of labor or capital) with a few recent exceptions (Dietz and Stern, 2015; Gazzotti et al., 2021; Glanemann, Willner, and Levermann, 2020; Hänsel et al., 2020; Frances C. Moore and Diaz, 2015; Moyer et al., 2014; Ricke et al., 2018; Estrada, Tol, and Gay-García, 2015).

Despite its importance for determining the aggregate costs of climate change, evidence on the persistence of the impacts of temperature shocks is sparse and contradictory (Piontek et al., 2021). Dell, Jones, and Olken, 2012 show that persistent and non-persistent effects can produce identical contemporaneous effects on the growth rate but can be distinguished using lagged temperature effects. Using global national accounts data, they fit a reduced-form model with lagged temperature terms and find evidence that effects of temperature shocks in poorer countries do not revert within ten years, implying large negative effects of higher temperatures on economic growth, at least in the medium-term. Burke et al. (Burke, S. M. Hsiang, and Miguel, 2015) use a similar dataset to find robust evidence for a non-linear, hill-shaped relationship between contemporaneous temperature and GDP growth. However, evidence for persistent impacts on the economy is weaker since the sum of lagged effects has large standard errors with confidence intervals that include both zero and very large negative effects. In a model-selection exercise based on cross-validation, Newell et al (Newell, Prest, and Sexton, 2021) show that total climate damages are highly sensitive to the question of persistence and to the functional form of empirical models used to estimate effects, but also find that out-of-sample cross-validation tests are insufficiently powerful to disambiguate between alternate models of impact persistence. At a smaller spatial scale, Deryugina and S. Hsiang, 2017 found evidence of persistent but declining effects during the first ten years after a temperature shock in individual U.S. counties. Deryugina and S. Hsiang, 2017, and Colacito, Hoffmann, and Phan, 2019 found that increases in summer and fall temperature could have persistent effects on the gross state product of U.S. states.

A major empirical challenge is that estimating the sum of lagged effects, particularly for a non-linear function, can produce large standard errors and high uncertainty. For instance, in the quadratic specification used by Burke et al, identifying cumulative effects over ten years requires estimating and summing 20 regression coefficients (Burke, S. M. Hsiang, and Miguel, 2015). The uncertainty around this statistic depends on the variance and covariance of all 20 parameter estimates. More recent empirical investigations of climate impacts on economic growth have focused on resolving detail at the subnational scale (Kalkuhl and Wenz, 2020; Colacito, Hoffmann, and Phan, 2019; Damania, Desbureaux, and Zaveri, 2020), or on resolving impacts on the production process (Letta and Tol, 2018). While they suggest some persistence in temperature effects, the uncertainty around this key question relevant to understanding the aggregate costs of climate change remains largely unresolved.

2 METHODS

Here, we propose a statistical test to identify the presence of persistent effects of temperature on output using lower-frequency temperature variation. We first use a simulation exercise to demonstrate the power of the test to discriminate between cases with and without persistent effects of temperature. Second, we implement this test on individual country-level temperature and economic growth time-series. The test complements previous approaches that have used either lagged temperatures or out-of-sample tests to attempt to resolve the question of impact persistence but which, as described above, have mostly produced ambiguous results.

The essence of the approach is that persistent and transient impacts on economic output can be distinguished using temperature variation occurring at different frequencies. Internal variability of the climate system gives rise to oscillations at different timescales. This is an intrinsic characteristic of non-linear dynamic systems like the Earth's climate (Lorenz, 1963). While some of these fluctuations, such as El Nino Southern Oscillation with a period of 2 to 7 years, are well understood (Imbers et al., 2013), spectral analysis of atmospheric time series reveal fluctuations at all possible frequencies (Hasselmann, 1976; Michael E. Mann, Byron A. Steinman, and Sonya K. Miller, 2020). Figure 17 Panel A shows this variability in the US temperature time

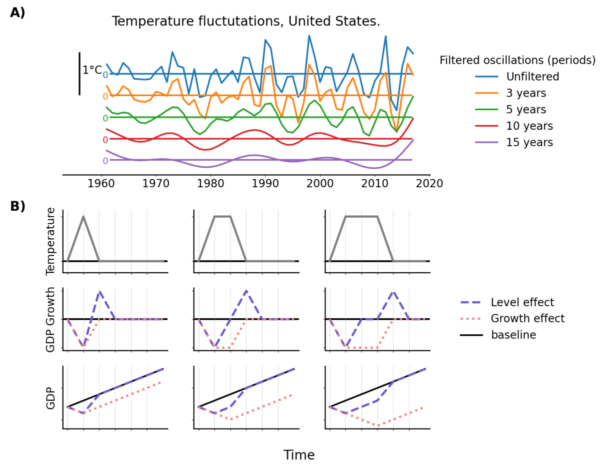


Figure 17: Temperature fluctuations (demeaned and detrended) and their effects on GDP. **Panel A**. US population-weighted temperature fluctuations after detrending and filtering higher-frequency variation (Matsuura and Willmott, 2018a). The top blue line shows the US temperature time series. Lower lines show the filtered time series, removing successively more higher-frequency variation. The time series are spread across the y-axis for visual purposes; all time series oscillate around zero because they were demeaned and detrended before filtering. **Panel B**. Upper panel: Temperature shocks at decreasing frequencies. Mid panel: Effects of those shocks on GDP growth under levels and growth models. Lower panel: Effects of temperature shocks on GDP.

series between 1960 and 2017 [24,25]. We use a low-pass filter to successively remove high-frequency variation and obtain temperature time series that preserve only lower-frequency oscillations.

Temperature variability at different timescales will produce distinct economic dynamics depending on the persistence of economic impacts. This is illustrated in Figure 1b, which shows the change in GDP growth and GDP level expected under temperature shocks of different durations and alternate models of economic impact. Dell, Jones, and Olken, 2012 derive a simple equation for a model that includes both non-persistent *level effects* (β) and persistent *growth effects* (γ), given baseline growth rate g:

$$g_t = g + \gamma T_t + \beta \Delta T_t \tag{6}$$

Where T_t is the deviation in temperature from some mean value in period t and T_t is the change in temperature between period t and $t-1$.[1] Although it is likely that some economies experience both levels

1 We inherit the taxonomy of "levels effects" and "growth effects" from Dell, Jones, and Olken, 2012. While we focus on GDP growth, the terms originate in reference to effects on GDP. A level effect alters the level of GDP, and when temperature reverts to the baseline so does production. A growth effect alters the growth rate, thus its effects are cumulative and persistent.

and growth effects simultaneously, we use two stylized cases in Figure 17b to illustrate how the timescale of temperature variation interacts with the models of economic impact. In the pure "level effects" model we set the growth effect to zero (i.e. $\gamma = 0$) so that:

$$g_t = g + \beta \Delta T_t \tag{7}$$

In the "growth effects" model, we set the level effect to zero (i.e. $\beta = 0$) so that:

$$g_t = g + \gamma T_t \tag{8}$$

A one-year temperature shock equally reduces GDP in a level effects model and in a growth effects model (Figure Figure 17b, left column). However, when temperature returns to the baseline so does GDP in the level model, but not in the growth model (Figure Figure 17b, bottom-left panel). The two models thus produce distinct long-term effects on GDP: growth effects on GDP keep accumulating as the duration of the temperature excursion increases, but level effects disappear when temperature returns to its baseline. It is this effect of past temperature shocks on the future level of GDP, occurring because temperature affects the determinants of economic growth, that we refer to in this manuscript as "persistent" impacts.

Note that the effects illustrated in Figure 17 Panel B do not include any variation in the impact of temperature shocks as a function of the shock duration. The question of whether longer-period temperature excursions, more analogous to the type of permanent warming expected from climate change, produce either larger (via compounding effects and intensification) or smaller (via adaptation) impacts compared to shorter temperature shocks has been widely debated (Frances C Moore and Lobell, 2014; Burke and Emerick, 2016; Mérel and Gammans, 2021; Taraz, 2017; Kolstad and Frances C Moore, 2020). The question of persistence - whether the level of GDP is affected by *past* temperature shocks - is distinct from this issue however. The distinction between persistent vs non-persistent impacts arises because of how temperature affects the economy; non-persistent effects arise through temporary effects on productivity (crop yield losses from extreme heat are one example) whereas persistent effects arise from impacts on factors that have a long-lived effect on economic production (destruction of capital in extreme events for instance).

Adaptation or intensification would somewhat alter the shape of the responses shown in the right column of Figure 17b, but the levels and growth models would still produce qualitatively different dynamics, particularly in response to temperature shocks of different lengths.

The duration of temperature excursions from a mean value is key to identifying the presence of growth effects (Figure 17b top row). The correlation between temperature and GDP growth in a growth effects model does not depend on the duration of the temperature anomaly, but breaks down in a level effects model as the length of the excursion grows (Figure 17b, middle row). This happens because there are no level effects if temperature is constant but away from the baseline.

Therefore it should be possible, in principle, to detect the presence of persistent effects in empirical data using different timescales of temperature variability. It is a common practice in signal processing problems to decompose time series into a sum of periodic components with varying frequencies, amplitudes and phases (**?**), widely used in a variety of fields like audio processing, electrical engineering, and climate science (Bergland, 1969; Ghil et al., 2002; Smith, 2007). This approach allows the time-series to be reconstructed using a specific subsets of desired frequencies. A low-pass filter is a version of the time series that only preserves low frequency components. Following studies in the climate literature (Michael E Mann, Byron A Steinman, and Sonya K Miller, 2014), we use a low-pass filter to remove inter-annual variations and obtain temperature time series that preserve only lower-frequency oscillations. If changes in temperature do not influence the underlying determinants of growth (levels only model), the estimated effect of low-frequency temperature anomalies on GDP growth should converge towards zero from the estimated effect of unfiltered temperature data. In contrast, if changes in temperature alter the determinants of growth (presence of persistent effects), the correlation between temperature and GDP growth should be detectable after the temperature data is filtered.

Figure 18 demonstrates this effect in a simulation exercise. It shows results from time series regressions of simulated economic growth on simulated temperature at different levels of filtering under two stylized cases - one in which there are only non-persistent damages (i.e. the level effects-only model, purple line) and one with only persistent damages (i.e. the growth effects-only model, pink line), following equations 7 and 8 respectively. Additionally, to illustrate one of the many possible combinations, another semi-transparent line shows a simulation with mixed growth and level effects with opposite signs.

The random temperature time series used in the simulations preserve the frequency distribution of the Earth's natural oscillations by matching the spectral decomposition on 1500 years of pre-industrial global temperatures based on the Last Millennium Reanalysis (Tardif et al., 2019). Using this decomposition we generate 10,000 random 350 year temperature time series that preserve this frequency distribution but with random phase shifts (Figwer, 1997) and then simulate economic dynamics for each temperature time series under the two alternate impacts models using equations 7 and 8, and the combined effect using equation 6, adding an independent and identically distributed (iid) noise component. We regress the simulated economic growth data on temperature after filtering out varying ranges of frequencies from the temperature time series, and adjusting the regression estimate to avoid a small bias introduced by the changing amplitude of temperature variations at lower frequency filters (see Methods).

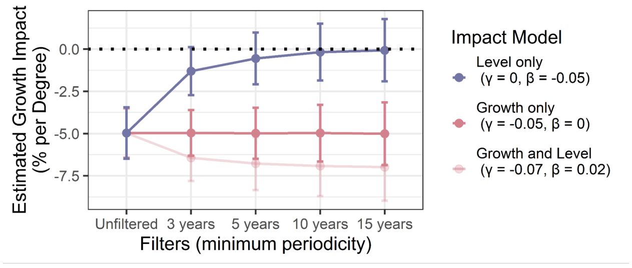


Figure 18: Simulation exercise demonstrating the divergence of regression results with increasing frequency filters under two alternate models of temperature impacts on economic production, a non-persistent "level only" model (purple) and a fully persistent "growth only" model (pink). A third semi-transparent pink line shows a combined model with opposite signs of growth and level effects.

Figure 18 shows the mean value of the estimated coefficients and its confidence interval for all the simulations. Without any filtering using only contemporaneous temperatures, growth and level impacts are indistinguishable, as originally pointed out by (Dell, Jones, and Olken, 2012). But filtering out high frequencies in the temperature data produces divergent effects: the estimated effect under the growth only model remains detectable while the coefficients in the level model attenuates markedly. In other words, the different patterns in Figure 18 mean that these two possible worlds - one with and one without persistent temperature impacts - could potentially be distinguished using this method. In essence, a statistical test on the coefficient for the filtered data is a test for the presence of growth effects, and is independent of the presence, or sign, of level effects.

While previous literature used lagged temperature estimates to test for growth effects, we show through a simulation that using a low-pass filter is more efficient in distinguishing between levels and growth effects at the medium to long term in a context where data is limited to 70 years. Supplementary Figure 17 compares the coefficients estimated with the filtering approach (left panel) and the sum of the lagged coefficients for a full distributed lag model (middle) and a more parsimonious version that reduces the number of estimated coefficients by imposing smoothness on the lag structure (right). The distributed lag model is as powerful at distinguishing levels from growth effects when the number of lags and the length of filtering are small. However, filtering grows more efficient for greater number of lags and longer filters, as the distributed lag model becomes increasingly noisy. This suggests that the low-pass filtering test can be a helpful complement to existing approaches using lagged temperature in investigating the persistence of effects over the medium to long run in data scarce contexts.

We use our test to investigate the persistence of temperature effects on economic production. We use GDP data from the World Bank covering 217 countries from 1961 to 2017 (World Bank, 2021), merging this dataset with population-weighted temperature and rainfall data from University of Delaware (Matsuura and Willmott, 2018a; Matsuura and Willmott, 2018b). To identify whether country-level temperature impacts have persistent effects we performed the following regression for each country and length of filter:

$$g_t = \theta_f T_{t,f} + \pi_f P_{t,f} + \epsilon_t \tag{9}$$

Where $T_{t,f}$ and $P_{t,f}$ are the population-weighted temperature and rainfall in year t after demeaning, detrending, and filtering out frequencies higher than f. The filters f are low-pass filters that filter-out any oscillations with periods shorter than 3, 5, 10, and 15 years, or f=unfiltered when no filter was applied. The low-pass filter algorithm requires data that spans at least twice the upper bound periodicity, which results in some countries not having estimates for all the levels of filtering due to missing data at earlier time periods. Country-specific quadratic time trends are removed from all variables (growth, temperature and rainfall) prior to analysis to address concerns of non-stationarity in the weather and economic time-series. Excluding rainfall from Equation 9 would bias the estimate of θ, since rainfall is known to correlate with both GDP growth (Kotz, Levermann, and Wenz, 2022) and temperature (Fischer and Knutti, 2016). However, we restrict the analysis to temperature and leave the discussion of results on precipitation to the Supplementary Material.

Given the lack of strong prior empirical evidence for the persistence of temperature effects, or strong theoretical or empirical evidence regarding drivers of heterogeneity in the response, the analysis focuses at the country level to give more flexibility and allow estimates to differ across countries. On the other hand, this comes at the cost of larger statistical uncertainty. We analyze the evidence for persistence across all countries at the global scale by separately pooling the positive and negative estimates of f and estimating the following regression model

$$\hat{\theta}_{f,c} = F_f + \epsilon_{f,c} \tag{10}$$

Where the value of the temperature coefficient estimate in country c at filtering level f is regressed on a vector of indicators of the level of filtering, clustering standard errors at the continent level.

3 RESULTS AND DISCUSSION

The behavior of the estimates $\hat{\theta}_f$ for each country contains information about the persistence of temperature effects on the economy. In particular, non-zero low-frequency estimates signal presence of growth effects, as shown in the simulations (Figure 7). We find that 39 countries have low-frequency estimates that are statistically different from zero at the 90% confidence level (of which 18 might be expected as false positives given the number of comparisons). Further, looking across all countries there is not strong evidence for systematic trends in coefficients towards zero at lower frequency variation, as would be expected if impacts operated only through non-persistent level effects.

Figure 19A shows the values of f for all countries at different levels of filtering, binned into two broad categories: a converging-towards-zero effect (blues), where the absolute value of f decreases at lower frequencies (as expected by the presence of level effects only, or by the combination of a level effect and a smaller growth effect), and a not-converging-towards-zero effect (oranges), where the absolute value of f increases at lower frequencies (explained only by the presence of persistent effects). In addition, there is a third category we describe as "unclassified" (grey) where the absolute value of f increases but changes sign between the unfiltered and the most filtered estimates. This behavior could be explained by levels and growth effects of opposite signs; yet, these countries are conservatively not classified as either converging or not converging. Within the two groups of converging and not converging countries, we further identify subsets of countries where the filtered estimates are either statistically larger (i.e. intensifying; dark orange) or smaller (converging; dark blue) from the unfiltered estimates.

Figure 19B tracks features of countries' estimates that are key to detect the presence of growth and levels effects. The left column divides countries based on the statistical significance of the unfiltered estimate, the middle column shows the statistical significance of the countries' most filtered estimate, and the right column shows whether estimates show converging, not converging, or intensifying effects. Among the 27 countries whose unfiltered estimate is statistically different from zero, the 15-year filtered coefficients of 18 countries are not statistically different from zero, meaning only level effects were detected in those countries (purple lines in Figure 19B). Presence of growth effects (Figure 19B, pink lines) is detected in the remaining 9 countries and in 30 other countries whose unfiltered estimate was not statistically different from zero.

The middle column of Figure 19B shows that in 18 countries where growth effects have been detected the filtered estimates are statistically larger than the unfiltered estimates (i.e. "intensifying effect"), a pattern that is consistent with level and growth effects of opposite sign. Among the remaining 137 countries that do not attain conventional statistical significance of the most filtered estimate, more countries have non-converging estimates (n=65, orange lines) than converging estimates (n=27, blue lines). There is no country where the filtered estimate is significantly smaller than the unfiltered estimate.

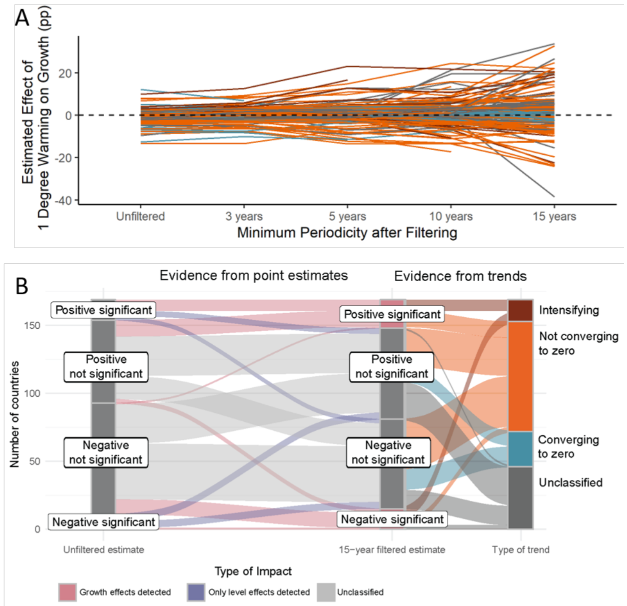


Figure 19: Panel A. Country-level estimates of the temperature effect on economic growth. For visualization purposes only, each line connects the estimated coefficients from regressions at different levels of filtering of the temperature data. Lines are color coded depending on the trend from the unfiltered to the most filtered estimate: orange when the absolute value of coefficients increases with filtering ("Not converging to zero"); dark orange when the difference between unfiltered and most filtered is significant at 10% ("Intensifying"); blue when the absolute value of coefficients decreases with filtering ("Converging to zero"), and dark blue when the trend is statistically significant at 10% ("*Converging to zero"; not found in this results); grey when the most filtered estimate is larger than the unfiltered but with opposite sign. The graph only shows countries with estimates below the 99th percentile for readability. Panel B. The left-hand side of the chart displays the number of countries for which there is evidence of growth effects, in pink, and evidence of level effects, in purple. The right-hand side classifies 15-year filtered estimates by the type of trend using the same color code as Panel A.

We performed the same analysis using two alternative economic growth datasets that span a longer time period but include fewer countries. Firstly we used the Barro-Ursua dataset, with annual data on economic growth of 43 countries starting as early as 1790 to 2009, developed to examine the persistence of macroeconomic shocks (Barro and Ursúa, 2008; Barro and Ursua, 2010). Secondly, we use the Maddison Project database that standardizes country-level GDP per capita for 170 countries for several centuries (Bolt and Van Zanden, 2020). Due to the sparsity of temperature and rainfall records pre-1900, we use only post-1900 data for both datasets. Supplementary Figure 4 replicates Figure 3 for these two alternate datasets covering different subsets of countries and much longer time-periods than the World Bank data. We again fail to find strong evidence that estimates systematically converge towards zero using lower frequency variation, as would be expected if impacts to the economy operated through non-persistent levels effects.

Pooling estimates from all countries, we are able to evaluate evidence, at the global level, for converging estimates at lower frequency filters. We thus estimate equation 10. Where the temperature coefficient estimate $\hat{\theta}_{f,c}$ in country c at filtering level f is regressed on a vector with the levels of filtering F, clustering standard errors at the continent level to allow for cross-country correlation and weighting the observations by the inverse standard error. Patterns such as divergence or convergence towards zero as filtering increases would cancel out if, as it shown in Figure 19, upper panel, there are both positive and negative effects. We therefore perform the analysis separately for countries with positive and negative unfiltered estimates. If only non-persistent level effects were present, we would expect to see the negative (positive) estimates converging towards zero, resulting in a positive (negative) coefficient estimate on the filtering variables F.

Figure 20 shows the cumulative estimated effect for each level of filtering, and shows that, across all countries, we do not see evidence for this attenuating effect. Instead, the regression results show evidence of persistent effect where the average value estimated using lower frequency temperature variation is similar to the value estimated using unfiltered data (See Supplementary Table 1).

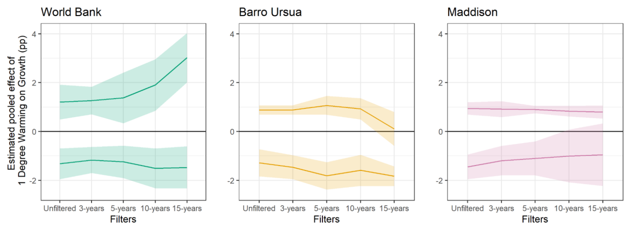


Figure 20: Pooled estimates of countries with positive and negative unfiltered coefficients across different levels of filtering using three alternative datasets.

Finally, Supplementary Figure 5 examines evidence for heterogeneity in the marginal effect of temperature between countries, specifically whether they are associated with either per capita GDP or mean temperature. Using only estimates significantly different from zero at the unfiltered and 15-year filter levels (i.e. countries for which evidence of persistent effects is strongest), we find some evidence that impacts are negatively correlated with countries' mean temperature as found in previous studies [1], but no systematic differences in the estimated effects between rich and poor countries (Supplementary Figure **8** shows a similar pattern resulting from a distributed lag non-linear model under a panel analysis).

4 DISCUSSION

The question of the persistence of climate damages is a first-order problem for climate change economics. Studies that allow climate change to affect the determinants of economic growth tend to produce far larger aggregate climate change costs than studies that impose only level effects on production (Frances C. Moore and Diaz, 2015; Ricke et al., 2018; Estrada, Tol, and Gay-García, 2015; Piontek et al., 2021). In response to the permanent shifts in temperature expected with climate change, persistent impacts operate via effects on the growth rate compound over time, producing far larger aggregate damages over the long time frames relevant for assessing climate change costs. Yet, impacts have been modeled as non-persistent by the numerous integrated assessment studies that since the 1990s have calculated climate damages and evaluated optimal climate policy.

In contrast with previous literature that models non-linear effects of temperature on growth, we analyze the temperature-growth relationship with country-level regressions. The smaller temperature ranges allow us to accurately model the effects using a linear approximation. (see Supplementary Material and Supplementary Figure 2). In addition, instead of using high-frequency, year-to-year temperature variation to estimate climate impacts on the economy, here we use lower frequency variation. Our identification strategy focuses on the persistent effect of temperature by adjusting for time trends and country-specific dynamics (via demeaning and detrending) but uses lower-frequency temperature variability instead of lags to distinguish between growth and levels effects. Using a low-pass filter instead of lags avoids adding noise terms together that could prevent identifying medium run persistent effects (see Supplementary Material and Supplementary Figure 1).

Applying this test to three different datasets of economic growth, we fail to find strong evidence of only non-persistent effects. There are two key pieces of evidence. First, we found statistically significant persistent temperature impacts on economic growth in 22% (19%; 8%) of the countries using the World Bank (Maddison Project; Barro-Ursua) dataset. Significant effects in these regressions implies the persistence of temperature impacts at least over the 15-year period of our lowest-frequency regressions. Secondly, we examine how regression estimates change using lower frequency temperature variation. The lack of persistent effects, as posited by the vast majority of IAM studies estimating climate damages, would imply convergence of these estimates towards zero. But we fail to find evidence of such convergence. At the individual country level, only 15% (21%; 34%) of countries have effects that converge towards zero. For many more countries, the estimated effects either do not converge towards zero or intensify over time, an effect that could be due to adaptation or coping dynamics, competing growth and levels effects with different signs, or a reduction in attenuation bias with longer filter lengths (though this effect is likely small, as described more fully in Supplementary Figure 6). Pooling evidence from across all countries produces stable effect sizes with lower frequency variation for all three datasets, at least over the 10-15 year period. Therefore, the evidence suggests a sensitivity of aggregate economic output to temperature shocks persisting over at least the 10-15 year time frame and a conspicuous absence of evidence for fully non-persistent levels impacts.

Like previous work, we find both positive and negative effects of temperature on different countries. It should be remarked that decade-long temperature excursions used to estimate the effects here are very small in amplitude (the median amplitude for 15-year filtered temperature is 0.11°C). While Figure 3 shows the effect of 1°C increase in temperature, the actual magnitude of temperature variation over this time-scale is much smaller and it is an open question whether these effect sizes can be extrapolated to much larger changes in temperature expected with climate change.

This highlights a fundamental empirical challenge in estimating the effects of climate change. Climate change will produce large ($\approx$ 2 to 4°C) and sustained changes in temperature. The historical record contains both large but short temperature excursions and much smaller but longer temperature variation. Previous papers (Burke, S. M. Hsiang, and Miguel, 2015; Dell, Jones, and Olken, 2012) have examined the effect of high frequency variation, raising the question of whether these estimates can be extrapolated to longer-lasting temperature changes (e.g. due to effects of adaptation, compounding effects, or the dynamics of persistent vs transient economic impacts). Here we instead focus on the opposite - lower-frequency but much smaller variation (at least in the filtered estimates). This gives more confidence that effects estimated are representative of impacts of sustained temperature change, at least over the medium run, while raising questions about whether these can be extrapolated to much larger levels of warming expected with climate change.

Finally, we note that our approach is not able to distinguish between a levels effect that continues compounding over the 15 year time-frame of our lowest-frequency estimates but then subsequently reverses, and a "pure" growth effect in which there is no subsequent reversal. Differentiating these two types of effects is a question of what happens in time-frames longer than 15-years, which is an inherently difficult empirical question due to the relatively short time span of data available. However, either interpretation of the filtered results (i.e. 15 years of continuously worsening levels effects followed by reversal or a fully persistent effect) implies persistence of damages over time periods longer than a decade. Either interpretation would imply larger aggregate climate damages than the standard approach to representing climate change costs in integrated assessment models, which assumes no persistence or compounding effects.

While providing evidence of persistent impacts of temperature shocks on growth, our framework does not isolate the mechanisms by which they arise. Past studies have modeled persistent impacts as resulting from a slow-down in total factor productivity growth (Frances C. Moore and Diaz, 2015; Moyer et al., 2014), changes to the capital depreciation rate (Frances C. Moore and Diaz, 2015), or impacts to the stock of natural capital (Bastien-Olvera and Frances C. Moore, 2021). Other studies leave the mechanism of growth rate impacts unspecified (Glanemann, Willner, and Levermann, 2020; Ricke et al., 2018). Letta and Tol, 2018 investigate this question and suggest impacts arise through effects on total factor productivity growth, but more work is needed to understand exactly how these impacts manifest.

A consistent and unsurprising finding from past work is that allowing for persistent damages, because of their compounding nature, vastly increases the uncertainty in climate change impact projections. For instance, Newell, Prest, and Sexton, 2021 estimate confidence intervals on damage estimates that allow for growth-rate effects orders of magnitude larger than those that restrict impacts to only the level of GDP. Similarly, in a recent modeling study, Kikstra et al., 2021 show that the persistence of economic damages is the most important parameter determining aggregate climate change costs. Our findings do not show strong evidence for the presence of only non-persistent impacts and instead suggest compounding effects over at least a decadal time frame. Therefore, restricting modeling of climate change damages to only non-persistent levels effects likely greatly under-states both the uncertainty and the downside risk associated with climate change.

REFERENCES

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel (Nov. 2015). "Global non-linear effect of temperature on economic production". en. In: *Nature* 527.7577, pp. 235–239. ISSN: 1476-4687. DOI: 10.1038/nature15725. URL: https://www.nature.com/articles/nature15725 (visited on 11/09/2021).

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (July 2012). "Temperature Shocks and Economic Growth: Evidence from the Last Half Century". en. In: *American Economic Journal: Macroeconomics* 4.3, pp. 66–95. ISSN: 1945-7707. DOI: 10.1257/mac.4.3.66. URL: https://www.aeaweb.org/articles?id=10.1257/mac.4.3.66 (visited on 11/15/2021).

Kalkuhl, Matthias and Leonie Wenz (Sept. 2020). "The impact of climate conditions on economic production. Evidence from a global panel of regions". en. In: *Journal of Environmental Economics and Management* 103, p. 102360. ISSN: 0095-0696. DOI: 10.1016/j.jeem.2020.102360. URL: https://www.sciencedirect.com/science/article/pii/S0095069620300838 (visited on 06/07/2021).

Carleton, Tamma A. and Solomon M. Hsiang (Sept. 2016). "Social and economic impacts of climate". In: *Science* 353.6304, aad9837. DOI: 10.1126/science.aad9837. URL: https://www.science.org/doi/10.1126/science.aad9837 (visited on 08/08/2022).

Deryugina, Tatyana and Solomon Hsiang (Nov. 2017). *The Marginal Product of Climate*. NBER Working Papers 24072. National Bureau of Economic Research, Inc. URL: https://ideas.repec.org/p/nbr/nberwo/24072.html.

Schlenker, Wolfram and Michael J. Roberts (Sept. 2009). "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change". In: *Proceedings of the National Academy of Sciences* 106.37, pp. 15594–15598. DOI: 10.1073/pnas.0906865106. URL: https://www.pnas.org/doi/10.1073/pnas.0906865106 (visited on 08/08/2022).

Dietz, Simon and Nicholas Stern (2015). "Endogenous Growth, Convexity of Damage and Climate Risk: How Nordhaus' Framework Supports Deep Cuts in Carbon Emissions". en. In: *The Economic Journal* 125.583, pp. 574–620. ISSN: 1468-0297. DOI: 10.1111/ecoj.12188. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12188 (visited on 08/08/2022).

Gazzotti, Paolo et al. (June 2021). "Persistent inequality in economically optimal climate policies". en. In: *Nature Communications* 12.1, p. 3421. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23613-y. URL: https://www.nature.com/articles/s41467-021-23613-y (visited on 08/08/2022).

Glanemann, Nicole, Sven N Willner, and Anders Levermann (Jan. 2020). "Paris Climate Agreement passes the cost-benefit test". eng.

In: *Nature communications* 11.1, p. 110. ISSN: 2041-1723. DOI: 10.1038/s41467-019-13961-1. URL: https://europepmc.org/articles/PMC6985261 (visited on 08/08/2022).

Hänsel, Martin C. et al. (Aug. 2020). "Climate economics support for the UN climate targets". en. In: *Nature Climate Change* 10.8, pp. 781–789. ISSN: 1758-6798. DOI: 10.1038/s41558-020-0833-x. URL: https://www.nature.com/articles/s41558-020-0833-x (visited on 08/08/2022).

Moore, Frances C. and Delavane B. Diaz (Feb. 2015). "Temperature impacts on economic growth warrant stringent mitigation policy". en. In: *Nature Climate Change* 5.2, pp. 127–131. ISSN: 1758-6798. DOI: 10.1038/nclimate2481. URL: https://www.nature.com/articles/nclimate2481 (visited on 08/08/2022).

Moyer, Elisabeth J. et al. (June 2014). "Climate Impacts on Economic Growth as Drivers of Uncertainty in the Social Cost of Carbon". In: *The Journal of Legal Studies* 43.2, pp. 401–425. ISSN: 0047-2530. DOI: 10.1086/678140. URL: https://www.journals.uchicago.edu/doi/10.1086/678140 (visited on 08/08/2022).

Ricke, Katharine et al. (Oct. 2018). "Country-level social cost of carbon". en. In: *Nature Climate Change* 8.10, pp. 895–900. ISSN: 1758-6798. DOI: 10.1038/s41558-018-0282-y. URL: https://www.nature.com/articles/s41558-018-0282-y (visited on 08/08/2022).

Estrada, Francisco, Richard S. J. Tol, and Carlos Gay-García (July 2015). "The persistence of shocks in GDP and the estimation of the potential economic costs of climate change". en. In: *Environmental Modelling & Software* 69, pp. 155–165. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2015.03.010. URL: https://www.sciencedirect.com/science/article/pii/S1364815215000900 (visited on 08/08/2022).

Piontek, Franziska et al. (July 2021). "Integrated perspective on translating biophysical to economic impacts of climate change". en. In: *Nature Climate Change* 11.7, pp. 563–572. ISSN: 1758-6798. DOI: 10.1038/s41558-021-01065-y. URL: https://www.nature.com/articles/s41558-021-01065-y (visited on 08/08/2022).

Newell, Richard G., Brian C. Prest, and Steven E. Sexton (July 2021). "The GDP-Temperature relationship: Implications for climate change damages". en. In: *Journal of Environmental Economics and Management* 108, p. 102445. ISSN: 0095-0696. DOI: 10.1016/j.jeem.2021.102445. URL: https://www.sciencedirect.com/science/article/pii/S0095069621000280 (visited on 08/08/2022).

Colacito, Riccardo, Bridget Hoffmann, and Toan Phan (2019). "Temperature and Growth: A Panel Analysis of the United States". en. In: *Journal of Money, Credit and Banking* 51.2-3, pp. 313–368. ISSN: 1538-4616. DOI: 10.1111/jmcb.12574. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12574 (visited on 08/08/2022).

Damania, R., S. Desbureaux, and E. Zaveri (July 2020). "Does rainfall matter for economic growth? Evidence from global sub-national data (1990-2014)”. en. In: *Journal of Environmental Economics and Management* 102, p. 102335. issn: 0095-0696. doi: 10.1016/j.jeem.2020.102335. url: https://www.sciencedirect.com/science/article/pii/S0095069620300589 (visited on 08/08/2022).

Letta, Marco and Richard S. J. Tol (June 2018). “Weather, Climate and Total Factor Productivity”. In: *Environmental and Resource Economics* 73.1, pp. 283–305. doi: 10.1007/s10640-018-0262-8.

Lorenz, Edward N. (Mar. 1963). “Deterministic Nonperiodic Flow”. EN. In: *Journal of the Atmospheric Sciences* 20.2, pp. 130–141. issn: 0022-4928, 1520-0469. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2. url: https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963_020_0130_dnf_2_0_co_2.xml (visited on 08/08/2022).

Imbers, J. et al. (2013). “Testing the robustness of the anthropogenic climate change detection statements using different empirical models”. en. In: *Journal of Geophysical Research: Atmospheres* 118.8, pp. 3192–3199. issn: 2169-8996. doi: 10.1002/jgrd.50296. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50296 (visited on 08/08/2022).

Hasselmann, K. (Jan. 1976). “Stochastic climate models Part I. Theory”. In: *Tellus* 28.6, pp. 473–485. issn: 0040-2826. doi: 10.3402/tellusa.v28i6.11316. url: https://doi.org/10.3402/tellusa.v28i6.11316 (visited on 08/08/2022).

Mann, Michael E., Byron A. Steinman, and Sonya K. Miller (Jan. 2020). “Absence of internal multidecadal and interdecadal oscillations in climate model simulations”. en. In: *Nature Communications* 11.1, p. 49. issn: 2041-1723. doi: 10.1038/s41467-019-13823-w. url: https://www.nature.com/articles/s41467-019-13823-w (visited on 08/08/2022).

Matsuura, Kenji and Cort J Willmott (2018a). “Terrestrial air temperature: 1900–2017 gridded monthly time series”. In: *Electronic. Department of Geography, University of Delaware, Newark, DE* 19716.

Moore, Frances C and David B Lobell (2014). “Adaptation potential of European agriculture in response to climate change”. In: *Nature Climate Change* 4.7, pp. 610–614.

Burke, Marshall and Kyle Emerick (Aug. 2016). “Adaptation to Climate Change: Evidence from US Agriculture”. In: *American Economic Journal: Economic Policy* 8.3, pp. 106–40. doi: 10.1257/pol.20130025. url: http://www.aeaweb.org/articles?id=10.1257/pol.20130025.

Mérel, Pierre and Matthew Gammans (2021). “Climate Econometrics: Can the Panel Approach Account for Long-Run Adaptation?” In: *American Journal of Agricultural Economics* 103.4, pp. 1207–1238.

Taraz, Vis (2017). “Adaptation to climate change: Historical evidence from the Indian monsoon”. In: *Environment and Development Economics* 22.5, pp. 517–545.

Kolstad, Charles D and Frances C Moore (2020). "Estimating the economic impacts of climate change using weather observations". In: *Review of Environmental Economics and Policy*.

Bergland, G (1969). "Fast Fourier transform hardware implementations–An overview". In: *IEEE Transactions on Audio and Electroacoustics* 17.2, pp. 104–108.

Ghil, Michael et al. (2002). "Advanced spectral methods for climatic time series". In: *Reviews of geophysics* 40.1, pp. 3–1.

Smith, Julius Orion (2007). *Mathematics of the discrete Fourier transform (DFT): with audio applications*. Julius Smith.

Mann, Michael E, Byron A Steinman, and Sonya K Miller (2014). "On forced temperature changes, internal variability, and the AMO". In: *Geophysical Research Letters* 41.9, pp. 3211–3219.

Tardif, Robert et al. (2019). "Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling". In: *Climate of the Past* 15.4, pp. 1251–1273.

Figwer, Jaroslaw (1997). "A new method of random time-series simulation". In: *Simulation Practice and Theory* 5.3, pp. 217–234.

World Bank (2021). *World Bank open data*. available at: https://data.worldbank.org/.

Matsuura, Kenji and Cort J Willmott (2018b). "Terrestrial precipitation: 1900–2017 gridded monthly time series". In: *Electronic. Department of Geography, University of Delaware, Newark, DE* 19716.

Kotz, Maximilian, Anders Levermann, and Leonie Wenz (Jan. 2022). "The effect of rainfall changes on economic production". en. In: *Nature* 601.7892, pp. 223–227. ISSN: 1476-4687. DOI: 10.1038/s41586-021-04283-8. URL: https://www.nature.com/articles/s41586-021-04283-8 (visited on 08/18/2022).

Fischer, E. M. and R. Knutti (Nov. 2016). "Observed heavy precipitation increase confirms theory and early models". en. In: *Nature Climate Change* 6.11, pp. 986–991. ISSN: 1758-6798. DOI: 10.1038/nclimate3110. URL: https://www.nature.com/articles/nclimate3110 (visited on 08/18/2022).

Barro, Robert J and José F Ursúa (2008). *Macroeconomic crises since 1870*. Tech. rep. National Bureau of Economic Research.

Barro, Robert J and Jose F Ursua (2010). *Barro-Ursúa macroeconomic data*. https://scholar. harvard. edu/barro/publications/barro-ursua-macroeconomic-data.

Bolt, Jutta and Jan Luiten Van Zanden (2020). "Maddison style estimates of the evolution of the world economy. A new 2020 update". In: *Maddison-Project Working Paper WP-15, University of Groningen, Groningen, The Netherlands*.

Bastien-Olvera, Bernardo A. and Frances C. Moore (Feb. 2021). "Use and non-use value of nature and the social cost of carbon". en. In: *Nature Sustainability* 4.2, pp. 101–108. ISSN: 2398-9629. DOI: 10.1038/s41893-020-00615-0. URL: https://www.nature.com/articles/s41893-020-00615-0 (visited on 04/23/2021).

Kikstra, Jarmo S et al. (2021). "The social cost of carbon dioxide under climate-economy feedbacks and temperature variability". In: *Environmental Research Letters* 16.9, p. 094037.

A SUPPLEMENTARY MATERIAL

A.1 Further details on the methods

The general approach of using lower frequency temperature variation to better understand the magnitude and dynamics of climate change impacts is well established in the climate impacts literature. Several papers contrast impacts estimated using high-frequency weather variation with those estimated using lower-frequency variation, either average temperature differences over long intervals (i.e. "long differences") or multi-decadal moving averages, to identify the effects of adaptation on the levels of climate damages (Burke and Emerick, 2016; Mérel and Gammans, 2021; Taraz, 2017; Kolstad and F. C. Moore, 2020). Most notably, Hsiang, 2016 presents panel regressions of US temperature and corn yield data, successively filtering out higher-frequency temperature and yield variation and argues that the stability of regression estimates using longer temperature variation indicates agricultural adaptation to warming is either slow or ineffective.

While conceptually similar to our empirical approach, the question this literature addresses is distinct in that, because the dependent variable in each case is a level outcome (typically crop yields), these papers address how adaptation does or does not attenuate the level of climate damages as a function of the longevity of temperature variation. Since our dependent variable is a growth rate, the question addressed is whether the effects of short-term temperature shocks on the level of GDP persist, and therefore whether damages compound over time in response to sustained periods of warming. Most importantly, even if the estimated growth effect attenuates to zero at lower frequencies (i.e. the purple line in Figure 2), this is still consistent with an effect of long-term warming on the level of GDP, for instance as modeled in the damage function of most cost-benefit integrated assessment models (Diaz and F. Moore, 2017).

For the simulation exercise (Figure 17), we first generated 10,000 random 350-year temperature time series that preserve the internal dynamics and characteristic periodicity intrinsic to the climate system. This dynamic was retrieved by performing a fast Fourier transform (FFT) of 1500 years of global mean surface temperature data prior to anthropogenic influence, obtained from the Last Millennium Reanalysis project (Tardif et al., 2019). Simulated temperature time-series were generated using the spectral profile given by this FFT but with randomly chosen phases, generating 10,000 random counterfactual time series that might have arisen from the Earth's natural variability.

For each of the 10,000 temperature time series we generated two alternative economic growth time series that reflected the two climate impacts scenarios that we hope to distinguish: levels and growth. Following Dell, Jones, and Olken, 2012, the levels model is given by

$g_t = g + \beta T_t - \beta T_{t-1} + \epsilon_t$ and the growth model by $g_t = g + \gamma T_t + \epsilon_t$. The growth baseline g was set at 0.01 representing 1% per year baseline growth, the temperature coefficients β and γ were both set at -0.05 representing 5% decrease in growth per degree of warming, and a random noise was drawn from a normal distribution with standard deviation of 0.005, representing growth rate variability unexplained by temperature.

The persistence test consists of regressing growth on temperature after filtering the temperature time series to remove higher frequency oscillations. We use a low-pass Butterworth filter in R (`pass.filt` from `dplR` library) that removes all oscillations with periodicity between 2 and the desired upper boundary of the filter. We perform the regressions of simulated growth on simulated temperature for 4 sets of filters (upper boundary = 3, 5, 10 and 15 years), and an unfiltered case. 15 years is the longest periodicity we filter because the algorithm needs data that spans at least twice the maximum period, so after 30 years data for many countries started to be missing. The unfiltered case, in both the simulations and the main regressions also includes a one-year temperature lag. This is required for generating an unbiased estimate of the levels effect - if temperature affects levels then T_{t-1} determines g_t (i.e. Equation 7). Omitting T_{t-1} will therefore bias estimates of the effect of contemporaneous temperature shocks (T_t) if there is temporal autocorrelation in the timeseries. Lags are not included in regressions using filtered temperature data since these regressions are intended to integrate the effect of persistent temperature excursions. Figure 17 shows the mean value of the estimates after filtering the temperature data and the 95% confidence interval.

One concern is that applying a frequency filter reduces the amplitude of the temperature time series, effectively attenuating unusual temperatures related to some extreme events and therefore mechanically inflating the estimates of the temperature coefficient, an effect that could lead to spurious evidence of "non convergence" if not corrected. Therefore we apply a correction factor to all estimates. Prior to filtering, the time series is detrended and demeaned. We then compute the median ratio of the amplitude between filtered and unfiltered temperature time series to gauge the magnitude of the (multiplicative) bias; and then divide the estimated coefficient by the ratio. Supplementary Figure 6 illustrates the effectiveness of this approach using the simulations also shown in Figure 17.

We retrieved yearly country-level data on economic growth for the 217 countries in the World Bank database (World Bank, 2021) for the period 1960 to 2020. Gridded temperature and precipitation data from the University of Delaware dataset (1900 to 2017; (Matsuura and Willmott, 2018a; Matsuura and Willmott, 2018b)) was aggregated to the country level using 2015 population weighting from the Gridded Population of the World version 4 dataset (Doxsey-Whitfield et al., 2015). Two alternative datasets were used to check for the robustness of the results (See Supplementary Figure 3). The first is the Barro-Ursua economic dataset, covering 43 countries from the late 18th century to 2009 (Barro and Ursua, 2010). The dataset has been constructed with the specific focus of studying periods of macroeconomic crisis during the industrial era. The second is the Maddison Project economic dataset that covers 169 countries during the study period (Bolt and Van Zanden, 2020). The dataset is intended for analysis of the determinants of growth and stagnation in the world economy, reflecting both current international differences in GDP per capita as well as the current knowledge on the historical patterns of growth. It combines multiple approaches to historical time series reconstruction in order to minimize the discrepancies with established historical benchmarks of income or living standards (Bolt and Van Zanden, 2020). Due to the sparsity of temperature and rainfall records pre-1900 and for greater confidence in GDP data, we use only post-1900 data for both datasets. The Supplementary materials list the countries contained in the three datasets.

Temperature, rainfall and economic growth data was demeaned and quadratic trends by country were removed to eliminate both time-invariant country variation and smooth, non-linear, country-specific trends in weather and growth rate. The residuals after demeaning and detrending were used to estimate the temperature effect () on economic growth by performing the following regression for each country and filter: $g_t = \theta T_{f,c} + \pi_f P_{f,c} + \epsilon_t$ where the index f represents the level of filtering applied to the temperature and rainfall data before performing the regressions. We apply a low-pass Butterworth filter of order 4 and periods f=3, 5, 10, 15.

As shown by our simulation (Figure 17, the persistence test consists of identifying whether (θ) is different from zero after filtering higher frequencies. That is, $|\theta_{15}| > 0$ is evidence for the existence of growth effects.

A.2 Comparison with lag models

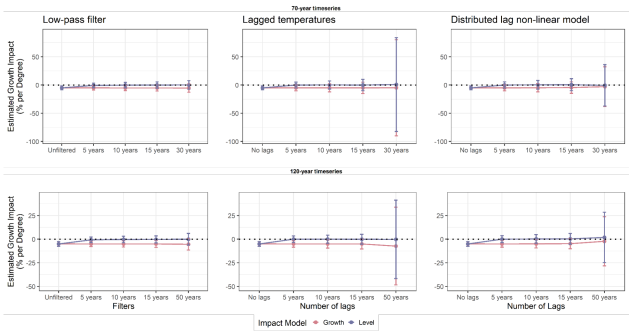


Figure 21: Simulations comparing filtering and distributed lag models. We created a random temperature time-series of 70 years (top) and 120 years (bottom) long and simulated growth and level effects on economic growth as in the original simulation. We then retrieved the temperature coefficients using the three alternative approaches: a low-pass filter (left), a regression with temperature lags (middle) and a regression with an imposed a degree-4 polynomial structure on temperature lags, which, by imposing smoothness on the lag structure, reduces the number of coefficients that need to be estimated (right). In the latter two panels the sum of lagged coefficients are plotted. The low-pass filter becomes more efficient than the distributed lags models for larger number of lags and longer filters.

A.3 Discussion of non-linearities

While in the literature there is evidence of non-linear effects of temperature on growth, this comes from panels of countries where the nonlinearity emerges over the very large cross-sectional variation in country temperatures (i.e. from just above 0°C to almost 30°C). Since we are interested in the within-country effect, where inter-annual temperature variability typically spans 2°C or less, the responses we estimate can be well-fit using a local linear approximation, even if the global response function across all countries is non-linear. Using a simulation, we show in Supplementary Figure 2 that an hypothetical "true" non-linear curvature as estimated by Burke et al [15] could be closely approximated by a linear relationship at a country-level. Importantly, the test for persistence effects using a linear relationship still successfully distinguishes between persistent and non-persistent effects even if the global, cross-country effect is non-linear. In addition, we test for the significance of a quadratic response at the country level and do not find evidence for this effect. Since adding quadratic terms greatly increases the number of coefficients that must be estimated and complicates the interpretation of the findings, we restrict the analysis to locally-linear, country-specific responses.

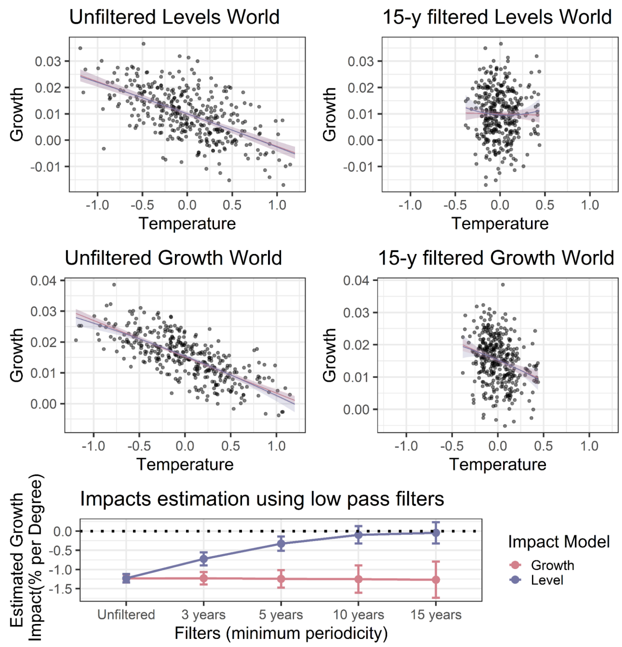


Figure 22: Comparison between a true non-linear effect and a linear regression model. Data for a hypothetical country with a mean temperature of 25 C and a global, cross-country nonlinear effect using the curvature estimated by Burke et al and an inter-annual, within-country time series variability of roughly 2°C as shown in their Extended Figure 1 b-c. Note that because this is a relatively hot country far from the BHM-estimated optimum in the response function, the non-linearity in the response will be larger than that of most other countries that are closer to the optimum. Top row: scatter plot of simulated GDP growth under temperature level effects for the unfiltered (left) and 15-years filtered (right) timeseries. The lines are fitted linear (red) and quadratic (blue) regression models with the shaded area showing the 95% confidence interval. Note that the slopes pass from being negative to be almost horizontal when the temperature time series is filtered. Middle row: scatter plot of simulated GDP growth under temperature growth effects for the unfiltered (left) and 15-years filtered (right) timeseries. The lines are fitted linear (red) and quadratic (blue) regression models with the shaded area showing the 95% confidence interval. Note that the slopes are virtually the same before and after filtering. Bottom: Persistency test using a "misspecified" linear model.

A.4 Impacts of precipitation

While the article focuses on the effects of temperature, we report here the results relative to the effects of precipitation. 67 countries exhibit evidence of growth effects at 90% confidence levels (bottom left panel, in pink). The larger share (60%) are negative growth effects, indicating that variation in precipitation from the climate norm have persistent adverse effects on the economy. In 51 countries, a switch in sign is detected as the 15-year filter is applied, going from positive to negative estimates. In light of Figure 2, this trend can be interpreted as evidence of positive level effects and negative growth, persistent effects.

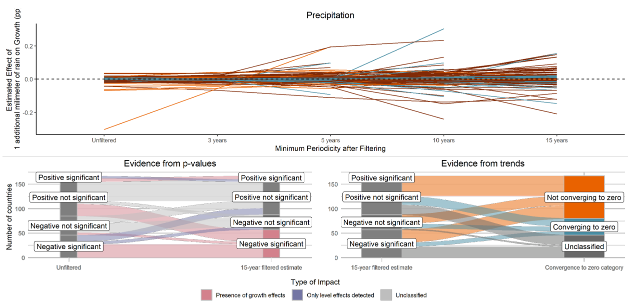


Figure 23: Top panel. Country-level estimates of the effect of precipitation on economic growth. Each line connects the estimated coefficients from regressions at different levels of filtering of the precipitation (and temperature) data. Lines are color coded depending on the trend from the unfiltered to the most filtered estimate: orange when the absolute value of coefficients increases with filtering ("Not converging to zero"); dark orange when the difference between unfiltered and most filtered is significant at 10% ("Intensifying"); blue when the absolute value of coefficients decreases with filtering ("Converging to zero"), and dark blue when the trend is statistically significant at 10% ("Converging to zero"); gray when the most filtered estimate is larger than the unfiltered but with opposite sign. The graph only shows countries with estimates below the 99th percentile for readability. Bottom panel. The left-hand side of the chart displays the number of countries for which there is evidence of growth effects, in pink, and evidence of level effects, in purple. The right-hand side classifies 15-year filtered estimates by the type of trend using the same color code as Panel A.

A.5 Supplementary Figures

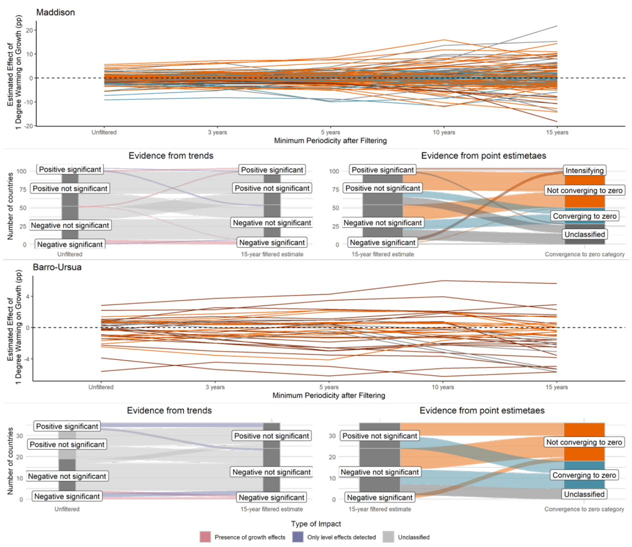


Figure 24: Replication of Figure 3 in the main text g using alternate economic growth datasets. Top: Maddison Project economic dataset [14], Bottom: Barro-Ursua project economic dataset [13].

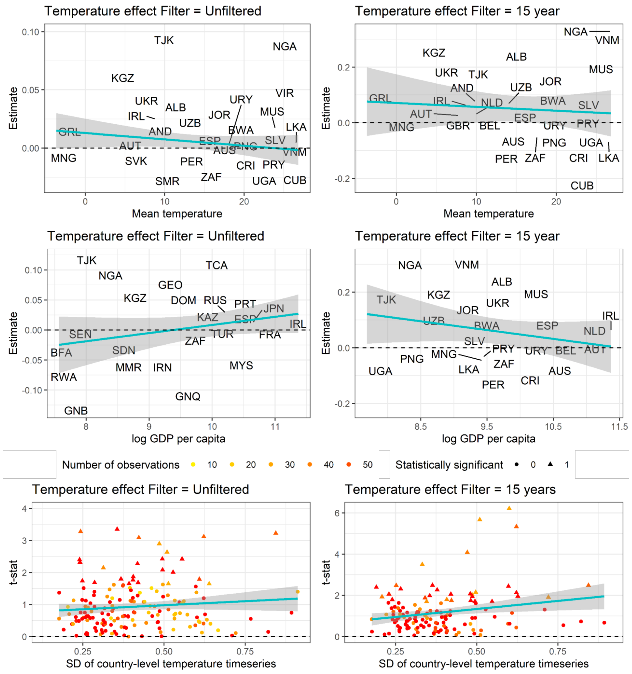


Figure 25: Estimates (only significantly different from zero) across countries mean temperatures (top panel) and log of the GDP per capita in 2019 (middle panel) for unfiltered (left) and 15-year filtered estimates (right). The bottom panel shows that for the 15-year filtered estimates there is a positive relationship between countries that are statistically significant and the standard deviation of the country's yearly temperature, meaning that, on average, larger variance in temperature helps to identify the effect. The blue lines are smoothed linear regression models fitted to the data and the shaded areas show the 95% confidence interval.

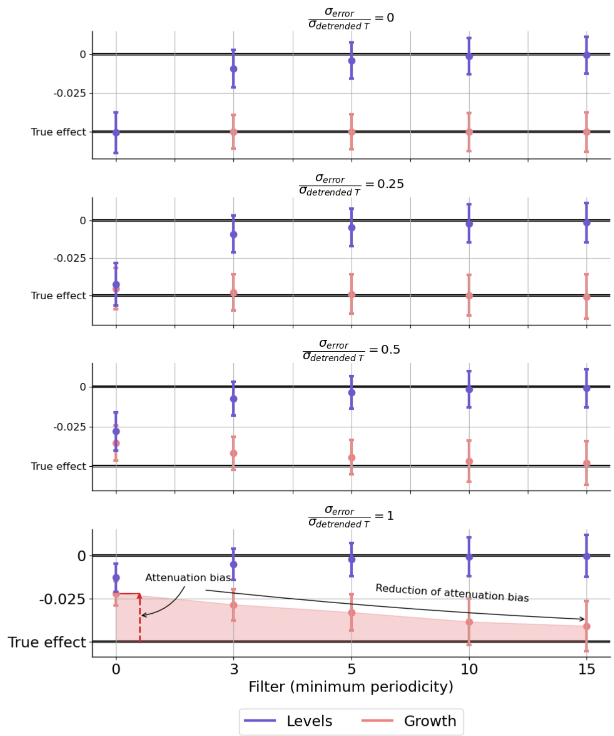


Figure 26: Simulations as described for Figure 2 but adding iid noise of growing magnitude to the temperature time series. True size effect = -0.05% (shown by the black horizontal line in each panel). Note that coefficients in the level model still trend towards zero at longer filters, but impacts in the growth model intensify slightly due to reduced attenuation bias from filtering out noise in the temperature time series. Substantial measurement error in the temperature variable could attenuate the estimated coefficient, biasing it towards zero, and inducing an apparent intensification effect as longer filters gradually filter out noise in the temperature variable, producing larger coefficients closer to the true growth effect. However, measurement error on temperature would need to be very large (i.e of comparable magnitude to inter-annual variation in temperature, bottom panel) in order to explain the intensifying pattern observed in some countries.

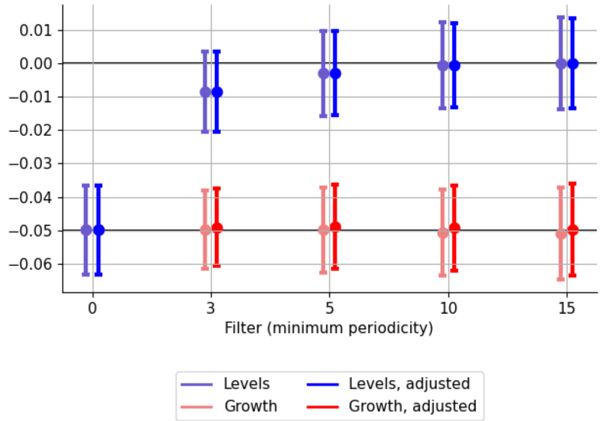


Figure 27: Simulations as described for Figure 2 but comparing adjusted and unadjusted coefficients. True size effect = -0.05%. The filtering of temperature data reduces the amplitude of the climate signal and mechanically inflates the estimated coefficients (blue and orange coefficient). Coefficients are adjusted by a multiplicative factor equal to the median of the ratio of filtered to unfiltered data (green and red coefficients). Longer filters are applied to highlight the bias and bias correction

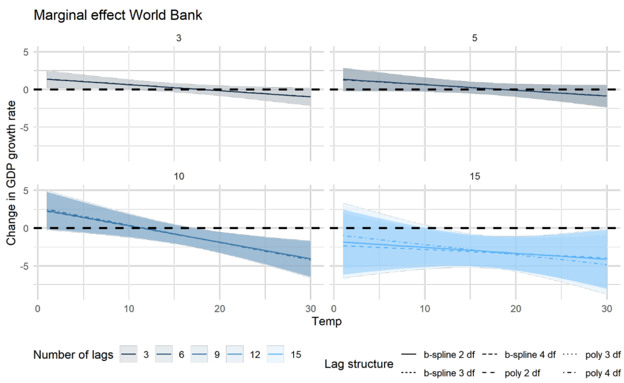


Figure 28: Marginal effect of temperature on GDP growth estimated with distributed lag non-linear models with panel data. GDP growth data comes from the World Bank.

A.6 Supplementary Tables

Table 21: Results of regression model.

	Dependent variable: Estimated coefficient Positive World Bank (-)	Negative World Bank (+)	Positive Barro-Ursua (-)	Negative Barro-Ursua (+)	Positive Maddison (-)	Negative Maddison (+)
Constant (Unfiltered)	-0.013***	0.012***	-0.013***	0.009***	-0.014***	0.009***
	-0.003	-0.004	-0.003	-0.001	-0.003	-0.001
Filter = 3 years	0.002	0.001	-0.002	0.00004	0.003***	-0.0003
	-0.001	-0.002	-0.002	-0.002	-0.001	-0.001
Filter = 5 years	0.001	0.002	-0.005**	0.002	0.003	-0.0005
	-0.003	-0.002	-0.003	-0.003	-0.003	-0.001
Filter = 10 years	-0.002	0.007	-0.003	0.001	0.004	-0.001
	-0.006	-0.004	-0.002	-0.003	-0.003	-0.002
Filter = 15 years	-0.002	0.018***	-0.006*	-0.008*	0.005	-0.002
	-0.005	-0.007	-0.003	-0.004	-0.004	-0.002
Observations	427	342	95	85	259	260
R2	0.002	0.028	0.02	0.058	0.005	0.001
Adjusted R2	-0.007	0.017	-0.023	0.011	-0.011	-0.015
Residual Std. Error	0.186	0.239	0.117	0.108	0.178	0.158

In columns marked with (+) the dependent variable are the positive coefficients obtained estimating equation (4). In columns marked with (-) the dependent variable are the negative coefficients so obtained. World Bank, Barro-Ursua, and Maddison are three different datasets of economic growth used to estimated equation (4). Observations are weighted by the inverse of the standard error from equation (4). Standard errors clustered at the continent level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

REFERENCES

Burke, Marshall and Kyle Emerick (Aug. 2016). “Adaptation to Climate Change: Evidence from US Agriculture”. In: *American Economic Journal: Economic Policy* 8.3, pp. 106–40. DOI: 10.1257/pol.20130025. URL: http://www.aeaweb.org/articles?id=10.1257/pol.20130025.

Mérel, Pierre and Matthew Gammans (2021). "Climate Econometrics: Can the Panel Approach Account for Long-Run Adaptation?" In: *American Journal of Agricultural Economics* 103.4, pp. 1207–1238.

Taraz, Vis (2017). "Adaptation to climate change: Historical evidence from the Indian monsoon". In: *Environment and Development Economics* 22.5, pp. 517–545.

Kolstad, Charles D and Frances C Moore (2020). "Estimating the economic impacts of climate change using weather observations". In: *Review of Environmental Economics and Policy*.

Hsiang, Solomon (Oct. 2016). "Climate Econometrics". In: *Annual Review of Resource Economics* 8.1, pp. 43–75. DOI: 10.1146/annurev-resource-100815-095343.

Diaz, Delavane and Frances Moore (Nov. 2017). "Quantifying the economic risks of climate change". en. In: *Nature Climate Change* 7.11, pp. 774–782. ISSN: 1758-6798. DOI: 10.1038/nclimate3411. URL: https://www.nature.com/articles/nclimate3411 (visited on 08/18/2022).

Tardif, Robert et al. (2019). "Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling". In: *Climate of the Past* 15.4, pp. 1251–1273.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (July 2012). "Temperature Shocks and Economic Growth: Evidence from the Last Half Century". en. In: *American Economic Journal: Macroeconomics* 4.3, pp. 66–95. ISSN: 1945-7707. DOI: 10.1257/mac.4.3.66. URL: https://www.aeaweb.org/articles?id=10.1257/mac.4.3.66 (visited on 11/15/2021).

World Bank (2021). *World Bank open data*. available at: https://data.worldbank.org/.

Matsuura, Kenji and Cort J Willmott (2018a). "Terrestrial air temperature: 1900–2017 gridded monthly time series". In: *Electronic. Department of Geography, University of Delaware, Newark, DE* 19716.

– (2018b). "Terrestrial precipitation: 1900–2017 gridded monthly time series". In: *Electronic. Department of Geography, University of Delaware, Newark, DE* 19716.

Doxsey-Whitfield, Erin et al. (2015). "Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4". In: *Papers in Applied Geography* 1.3, pp. 226–234.

Barro, Robert J and Jose F Ursua (2010). *Barro-Ursúa macroeconomic data*. https://scholar. harvard. edu/barro/publications/barro-ursua-macroeconomic-data.

Bolt, Jutta and Jan Luiten Van Zanden (2020). "Maddison style estimates of the evolution of the world economy. A new 2020 update". In: *Maddison-Project Working Paper WP-15, University of Groningen, Groningen, The Netherlands*.