HEC MONTRÉAL

The Impact of Pollution on Productivity :
Evidence from U.S. Immigration Court Judges

par

Emmanuelle B. Faubert

Sciences de la gestion
(Option Économie Appliquée)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

August 2022

Résumé

Ce mémoire étudie la relation de causalité entre la pollution quotidienne et la productivité des travailleurs. La productivité est mesurée par le nombre de cas résolus par les juges des courts d'immigration des États-Unis et la pollution est mesurée par les niveaux de $Pm_{2.5}$, des particules ultra-fines de taille 2.5. La méthode préférée est un modèle d'estimation avec effets fixes sur des données de panel, incluant plusieurs variables de contrôle temporelles et environmentales, ainsi que plusieurs effets fixes concernant les juges eux-mêmes et les tendances temporelles. Nous trouvons que la pollution a un effet non linéaire sur la productivité. C'est à dire, quand la pollution est faible, une augmentation de $Pm_{2.5}$ mène à une augmentation de la productivité. Cela change et devient négatif après avoir atteint un point pivot. Ces résultats, toutefois, ne sont pas robustes aux valeurs abbérantes et ils contredisent la litérature, qui trouve unanimement un impact linéaire négatif entre la pollution et la productivité. Ceci mène à se questionner sur les suggestion de l'Organisation Mondiale de la Santé.

Mots-clés: pollution; productivité; court d'immigration; environment

Méthodes de recherche: panel; régression quadratique; effets fixes; variables de contrôle; test de signification conjointe

Abstract

This mémoire studies the causal relationship between daily outdoor pollution and worker productivity. Productivity is measured by the number of cases U.S. immigration court judges close daily, and pollution is measured by daily average levels of $Pm_{2.5}$, or particulate matter of size 2.5. The method used is a fixed effect estimation on panel data, with various control variables for time and environmental factors, as well as judge and temporal fixed effects. We find that pollution has a non-linear effect on judges' productivity. That is, when pollution is low, an increase of $Pm_{2.5}$ leads to an increase in productivity, but these effect become negative after reaching a certain level of $Pm_{2.5}$. These results, however, are not robust to outliers and are inconsistent with the previous literature, which only ever found negative linear relationships between pollution and productivity. These results also raise the question of the World Health Organization pollution guidelines.

Keywords: pollution; productivity; Immigration court; environment

Research methods: panel; quadratic regression; fixed effects; control variables; joint-significance test

Contents

List of Tables

List of Figures

List of acronyms

2SLS Two-stage least squares

API Air pollution index

EOIR Executive Office for Immigration Review

FE Fixed effect

HEC Hautes études commerciales

MSc Maîtrise

NO_2 Nitroden dioxide

OLS Ordinary least squares

O_3 Ozone

PM Particulate matter

SO_2 Sulfur dioxide

U.S. United States

WHO World Health Organization

Acknowledgements

I would like to take the time to give my thanks to the many people who have helped me throughout this journey.

I would first like to thank my director Decio Coviello, for the immense support, guidance, and patience throughout the process of writing this memoire. He was always available to share his knowledge and guide me when I was feeling lost. I would also like to thank my family for their unwavering faith and support in me, who always were there for me when motivation was lacking. I would also like to thank all my closest friends who were always there when things got difficult. While writing can be a lonely process, they made me realise that I was not alone.

Introduction

In this ever-changing climate, pollution is a rising concern among the population worldwide. People are becoming more and more conscious of their actions, as they get more informed on how the increasing levels of pollution have negative impacts on the environment and on our health. According to the World Health Orginazation (2021), air pollution is one of the environmental factors that pose the most risk to human health. More specifically, it poses a major risk to cardiovascular health. It was estimated that, in 2016, 4.2 million deaths were caused by air pollution worldwide. The World Health Organization (WHO) publishes guidelines providing updated reports on the effects of pollution on health, as well as thresholds for harmful pollution levels. Amongst the various pollutants present in the air, notably ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and particulate matter (PM), the WHO puts a lot of emphasis on particulate matter because this particular pollutant is deemed to affect people more than any other. More specifically, particulate matter is a combination of solid and liquid particles in suspension in the air. The reason for the danger of particular matter is their incredibly small size. PM_{10}, particular matter with a size of 10 microns or less, can seep inside the lungs and cause damage. $PM_{2.5}$, having particles the size of 2.5 microns or less, are much more dangerous as they not only seep into the lungs, but also go beyond the lungs and enter the bloodstreams. Chronic exposure to $PM_{2.5}$ negatively impacts cardiovascular and respiratory health and is also linked with an increased risk of lung cancer. Particulate matter has negative impacts even when the concentration in the air is low. Therefore, the WHO Global Guidelines recommend reducing exposure as much as possible. The WHO Air quality guideline values are the following: For coarse particulate matter (PM_{10}): $15\mu g/m^3$ annual mean and $45\mu g/m^3$ 24-hour mean. as for fine particulate matter ($PM_{2.5}$): $5\mu g/m^3$ annual mean and $15\mu g/m^3$ 24-hour mean.

In reaction to the increased knowledge of these negative effects, governments worldwide make plans to strive and make a reduction in their emissions of greenhouse gas, like the Paris agreement that was signed in 2016 and had 196 parties included. Although the objective of the agreement regarding the limitation of global warming has not been reached as of 2022, progress has been made in the way of the creation of *low-carbon solutions* and *zero-carbon solutions*, which are becoming more competitive economically, and the establishment of carbon neutrality objectives by local governments and companies (United Nations, 2021). The United Nations predicts that by 2030, these *zero-carbon solutions* could be competitive in economic sectors that make up for around 70% of global greenhouse gas emissions, for example, in the transportation sector.

While it is clear that high levels of pollution can negatively affect people that spend a lot of time outdoors and in highly polluted areas, what about those that spend most of their time indoors, inside climate-controlled environments? $Pm_{2.5}$ are so small that they are known to be able to seep inside buildings, meaning that when $Pm_{2.5}$ levels are high outside, people indoors are still likely to be affected even when in indoor climate controlled environments. Considering the negative impacts of pollution on health, it would be unlikely that such effects would have no impact on the workforce. Indeed, if workers get sick, it would seem likely that efficiency in the workplace would be affected. And in a society where the service sector is a main driver of economic performance, understanding if pollution has a causal relationship with productivity means a better understanding of the effects of pollution on economic performance and growth.

In this context, determining the impact of pollution on individuals' productivity is a significant concern worthy of study, as the situation is unlikely to get reversed in the foreseeable future.

The question that is addressed here is whether outdoors pollution has an impact on worker productivity, more precisely on people working indoors and whose work is of a more intellectual nature. The analysis contributes to the developing understanding of the complex effects pollution has on humans, beyond simply the health status, but also on the efficiency of society. By improving our understanding of the subject, lawmakers are better equipped to make decisions that will positively impact people's living conditions and make regulations that can improve work efficiency by providing better working environments.

The method used is an econometric analysis that focuses on a fixed effect model, followed by a sub-sample analysis and later by an outlier analysis. The population analysed in this paper are 266 immigration court judges spread across the 43 US Federal Immigration Courthouses in the country.

We find that pollution has a significant non-linear impact on productivity, which differs from other studies so far. We find that when $Pm_{2.5}$ levels are closed to the average level of 14.46656 observed in the sample, an increase of $Pm_{2.5}$ levels of one standard deviation leads to an increase of around 1% productivity. This average level of $Pm_{2.5}$ is very close to the threshold of $15\mu g/m^3$ recommended by the WHO. These results, however, are only weakly significant when conducting a joint significant test for the linear and quadratic term.

Our results are not robust to outliers analysis. Our extreme value analysis shows that when extreme values of $Pm_{2.5}$ are removed from the sample, it becomes impossible to isolate the impact of $Pm_{2.5}$ levels from other pollution or weather variables on productivity. When doing a formal outlier analysis, we find that outliers are not the observations that have the most leverage on the model. Depending on the strictness of the analysis (depending on the size of the studentized residuals removed from the sample) we find that either sometimes we are able to isolate the effects of $Pm_{2.5}$ on productivity, while at other values it is not possible.

We also find that when comparing the least polluted city in the sample (Hagatna) to the most polluted city (Los Angeles), the data shows that while $Pm_{2.5}$ levels are 272.85% higher in Los Angeles compared to Hagatna, productivity there is also higher by 34.73%. However, our model predicts an increase in productivity caused by an increase of $Pm_{2.5}$ by only around 1.76%. These results do make sense, because the two cities are very different demographically and geographically, as Hagatna is located on a small island belonging to the United States, while Los Angeles is a major urban centre.

This thesis follows the following structure. The literature review presents articles related to the subject of pollution and productivity or that have influenced the structure of the following study. Chapter 1 explores the data. It aims to give a better understanding of the nature of the data used, including the manipulations executed. Chapter 2 explains the the econometric model used. Chapter 3 focuses on the results and their analysis. The conclusion wraps up the results and discusses possibilities for further studies.

Literature review

As the notion that pollution has a negative impact on health is imprinted into the population, Saberian et al. (2017) looks into how people respond to high pollution alerts and whether people change their behaviour when such alerts, recommending avoiding strenuous outdoors activities, have a real impact on people's behaviour. More precisely, they look into the behaviour of cyclists in Sydney, Australia. Their results indicate a significant reduction of cyclist activity when the alert is sent, but much more in those who cycle for leisure versus those use their bicycle as a method of transportation for work. This shows that, when possible, people will make an effort to avoid extensive pollution exposure. If such results can be extrapolated to the data studied here however, it might be interpreted that high levels of pollution is unlikely to affect judges behaviours on the days where the data was collected (although data regarding the judges' physical activity has not been collected).

In addition, Lee et al. (2014) have looked into the impact of external environmental conditions (weather) on productivity, through a survey where a vast majority of respondents already believed that weather and productivity were positively correlated. However, their experiments conducted in laboratories, on the field and online, conclude that bad weather actually has a positive impact on productivity, mainly due to the reduction of potential alternative activities to occupy the mind. These results lead us to believe that the impact of weather on productivity is due mainly to psychological effects rather than physiological, and is highly influenced by perception. Corroborating this, Coviello et al. (2021) looks into the impact of mood on productivity in a call-centre setting. They use a "mood questionnaire" to measure mood, and measure productivity by counting the number of calls per worker per hour. They find that positive mood has a negative impact on productivity. They also discover that bad weather, more specifically rain, has a negative impact on mood. They also control for pollution and that is an important factor. By linking the results found in these articles mentioned, it would then be unlikely that high levels of pollution would cause a significant change in worker behaviour. Indeed, unless pollution levels are very high, variations in air pollution are pretty much invisible to the naked eye. It is much more likely, then, that pollution would impact productivity by affecting the body directly, for example by affecting people's cognitive ability or by affecting the cardiovascular systems or the lungs, affecting the delivery of the necessary nutrients for the brain to function.

Zhang et al. (2018) looks into this more in detail. They find that pollution has a negative impact on cognitive function, and that this impact becomes greater with age. This is especially true for people who work outdoors, or spend large amounts of time in highly polluted areas. Graff Zivin and Neidell (2012) is such an example, as the authors look into the impact of ozone concentration in the air on the productivity of agricultural workers. Using a Tobit model, they obtain that pollution has a negative impact on productivity

Adhvaryu et al. (2022) applies this to a managerial setting. They look into how environmental shocks affect worker productivity, and if managers are able to mitigate some of those effects. They find a negative linear relationship between pollution and productivity, where one standard deviation increase of $PM_{2.5}$ cause a drop of productivity of 0.8%. They also find that the shock in pollution levels has an immediate impact, and the effect is very heterogeneous in the sample. They observe that people with cardiovascular or pulmonary issues, people that are older and people with more cognitively demanding jobs are affected in a greater magnitude by these shocks. They also note that attentive managers are able to mitigate some of those effects by redistributing work to those less sensitive, especially in settings where tasks are able to be reallocated. In this case, they therefore find that, on a more global scale, shocks in pollution have less impact in settings where work can easily be redistributed amongst the workforce. Regarding our study, this would suggest that we could expect a generally negative impact of an increase of $PM_{2.5}$ on judge productivity, but it is unlikely that those effects could be mitigated on a day-to-day basis, as cases are assigned randomly and it would be difficult to redistribute cases quickly and efficiently as there is already a lot of backlog of cases in the system.

While the impacts in outdoor workers seems more direct and clear cut, it is also important to look into the impact on workers working indoors. Chang et al. (2016) looks into the impact of $PM_{2.5}$ on worker productivity under such conditions. The data analysed is about pear packing factory workers in Northern California from 2001 to 2003. This paper first looks into if high $PM_{2.5}$ have an impact on employees showing up to work. Their results show that even during a two-week period during which pollution alerts sent sent because of large wild fires in the state, pollution levels did not have a significant impact on the probability of people showing up to work, further insinuating that potential productivity changes are unlikely to be of a psychological nature. They later look into the impact on daily productivity. They find that that an increase of $PM_{2.5}$ of 10 leads to a reduced worker productivity of around 6%. They also found a non-linear relationship between $PM_{2.5}$ levels and productivity, finding that these effects start when $PM_{2.5}$ levels reach $15\mu g/m^3$ and increase from that point on.

A lot of recent studies have come out of China regarding this topic because of the high levels of pollution in urban Chinese cities. One of them is by Chang et al. (2019), which looks into the impact pollution has on the daily productivity of call centre workers. They largely focus on the impact of changes in the air pollution level index, the API. The API is an index that ranks air quality daily based on its potential impact on human health. Higher values of API mean higher health risks. One main component of API is particular matter pollution of varying sizes. The smaller the particles, the easier they seep into buildings and into the body, ending up in the bloodstream. This article, however did not have access to this measure as part of their database, having only the API as their pollution measure, which was mostly dominated by levels of PM_{10} during their data collection period. The authors use a linear model using API as the main explanatory variable, temperature as their covariate, time related fixed effects like *day-of-week* and *year-month*, as well as worker related fixed effects. They also included two-way clustering because of the likely autocorrelation of the error term.

Their results suggest that an increase of API levels of 10 leads to a 0.35% decrease in productivity, which they measure by the number of calls per shift. They obtain similar results even when they modify the controls and fixed effects. However, they note that the impact seems to be mostly on the amounts of breaks needed between calls, rather than the time spent on average on each call. These results are interesting, as they imply that pollution would not impact directly the efficiency of single tasks, but rather the ability to navigate quickly between different tasks.

The next main source is a working paper by Kahn and Li (2019), which discusses the impact of pollution and temperature on Chinese public sector workers. More specifically, they look at cases closed by civil and criminal Chinese judges between 2013 and 2018. In this study, they use the time needed to close a case as their productivity measure. They also use $PM_{2.5}$ as their pollution measurement, which had an average level of 54.67 during the data collection period, with a standard deviation of 30.12. This measure, however, is not a daily measure, but but rather an average calculated over the period of time taken to resolve each case. The authors use a log-linear fixed effects model that clusters the standard errors that controls for crime types and complexity, self-protection efforts, as well as time, judge and location related fixed effects.

They conclude that pollution and higher temperatures have a negative impact on judges' productivity, by increasing the amount of time taken to close cases. More specifically, they obtain that a 1% increase in $PM_{2.5}$ leads to an increase of case duration of 19.8%. They also find that pollution and high temperature negatively impact the quality of decisions, by increasing the probability of the case being appealed. However, they find that the impact of pollution is much grater than the one of temperature, and believe this to be because of air conditioning used for climate control in the working environment. From the results of this paper, we therefore predict that pollution and temperature will have a negative impact on the number of cases closed each day by judges.

Another study located in China is by He et al. (2019). They look into the impact of very high $PM_{2.5}$ levels on indoors workers in Chinese industrial cities. Using a variety of models such as fixed effects, OLS and 2SLS models, they find that while day-to-day pollution has no immediate impact on workers productivity, pollution has a significant delayed impact,which can be up to 30 days later. This study suggests that while variations of pollution on a given day may not affect workers immediately, it could affect them later. This means that even if this study may not find significant results, it may not mean that there is indeed no impact whatsoever.

Chapter 1

The Data

1.1 Description of the Database

The research was done using the final database from Heyes and Saberian (2019), named *matched.dta*. The database was collected from the American Economics Journal: Applied Economics website. It combines information about U.S. immigration court asylum cases and daily weather and pollution data collected from January 2000 to August 2004 inclusively. It records 269,744 individual cases closed by a total of 266 different judges. Figure 1.1, taken directly from Heyes and Saberian, shows where all these courthouses are located on the mainland territory. Their data regarding the immigration court judges contains the following information: date of hearing, judge identification, asylum seeker nationality, and category of application (whether the applicant presents their case of their own initiative or whether they have to defend their case initiated by the immigration authorities). In this thesis, however, the main elements used for the present study are the date of hearing and judge identifications. The data could originally be found on the website *asylumlaw.org*, a website that was run by a group of international agencies that aimed to help asylum seekers, but the website is not available anymore. The data regarding the environment includes a large array of information, which was collected from different sources. Data for hourly weather indicators were collected by the National Oceanic

Figure 1.1 – Location of Immigration Courts

Notes: This figure shows the locations of the 43 immigration courthouses across the United-States, excluding Hawaii. Source: Heyes and Saberian (2019)

and Atmospheric Administration. Data for cloud cover was retrieved from the Northeast Regional Climate Center.

Daily pollution data, came from the United States Environmental Protection Agency. The environmental data and asylum data were joined, assigning the data from the closest weather monitoring stations to each courthouse. Figure 1.2 shows in more detail the various environmental data collected. This figure shows that the concentration of $PM_{2.5}$, $14.957\mu g/m^3$ is very close to the daily WHO guideline of $15\mu g/m^3$. A more detailed summary of $PM_{2.5}$ is shown in table 1.1. The results are slightly different because missing values and negative values of $PM_{2.5}$ have been dropped from the sample.

The original article aims to measure the impact of temperature on the decision of asylum court judges in the United States. Using a linear probability model that includes fixed effects and clusters the error term, they conclude that higher temperatures reduce the likelihood of being granted asylum. This leads to the impression that weather does have an

Figure 1.2 – Summary of the Environmental Data

	Mean	SD
Grant indicator	0.164	0.371
Temperature (°F)	57.37	15.721
Heat index (°F)	57.77	16.423
Air pressure (pa)	29.688	0.759
Dew point (°F)	49.372	17.202
Precipitation (mm)	0.003	0.014
Wind speed (km/h)	4.557	3.441
Sky cover (percent)	55.44	0.276
Ozone (ppm)	0.0220	0.0120
CO (ppm)	0.917	0.496
$PM_{2.5}$ (μ/m^3)	14.957	11.569

Notes: These are the environmental summary statistic as taken directly from the original article by Heyes and Saberian (2019)

Table 1.1 – Pm2.5 Detailed Summary

	Percentiles	Smallest		
1%	1.6	0.2		
5%	3.7	0.2		
10%	5	0.2	Observations	92,565
25%	7.8	0.2	Sum of Weight	92,565
50%	11.7		Mean	14.46656
		Largest	St. Dev	11.27727
75%	17.76667	166.3046		
90%	26.3	166.3046	Variance	127.1769
95%	34.6	166.3046	Skewness	3.579826
99%	56.375	166.3046	Kurtosis	30.13224

Notes: Detailed summary statistics of daily Pm2.5 levels measured in $\mu g/m^3$ across the 43 cities where courthouses are located from 2000 to 2004 once negative values of Pm2.5 are removed from the sample

impact on judges. This study uses the same database but aims to answer a different question, being the impact on the productivity of judges. Initially, the goal was to look into the impact of temperature on productivity. The results were inconclusive, as it appeared that temperature did not have any significant impact on the judges' productivity as shown in table 1 found in the appendix. That same table, however, shows a significant coefficient for carbon monoxide. This indicates indicating that there might be a relationship to explore between pollution and productivity.

1.1.1 Immigration Court Judges

Immigration courts are U.S. federal tribunals conducting trials on individuals accused of violating immigration laws by the Department of Homeland Security to decide whether they should be deported from the United States or be allowed to stay in U.S. territory. They are administrated by the Executive Office for Immigration Review (EOIR), which in turn falls within the United States Department of Justice. Cases are delegated by the Attorney General, and decisions are made by Immigration Judges and the Board of Immigration Appeals (National Association of Immigration Judges, 2022). Immigration judges are mandated to independently and impartially resolve complex cases. The United States judiciary system is separated from other legislative and executive branches to make sure cases are dealt with in a just manner (Slavin and Marks, 2015). Immigration courts, however, are not as separated, as they belong to both the executive branch of the government through the EOIR and the department of justice. Judges in this area of the law are considered to have more personal discretion and independence in how they evaluate files (Heyes and Saberian, 2019). Applicants are often vulnerable, as many do not speak English and are unaware of U.S. culture and laws. Therefore, immigration courts judges have a very complex task when it comes to adjudicating these cases (National Association of Immigration Judges, 2021).

As of May 2021, there is a backlog of cases that has reached over 1.3 million pending cases due to limited physical, human, and technological resources (National Association of Immigration Judges, 2021). This is putting a lot of pressure on immigration judges and could impact their efficiency independently of pollution levels.

1.1.2 Construction of the Productivity Measure

This database does not include explicit data regarding productivity, which is the main variable studied in this project. It is possible, however, to transform the data in order to create the variables needed.

The first step is to determine judges' productivity. Since judges do not produce physical products, it can be harder to measure. Coviello et al. (2015) studies Italian labour courts to establish the impact of task-juggling on judges. It mentions that, ideally, the time allocated to a case might be the best measure of productivity. However, due to the nature of the asylum courts, most cases are dealt within one hearing, therefore the time used to close a case cannot be used in our particular context. They also mention in the paper the use of the number of cases closed in a certain amount of time as a measure of productivity. Based on this paper, the measure chosen for the current study is the number of cases closed in a day by each judge.

Therefore, the variable *productivity* was created, which measures the total number of cases each judge closes on a given day during the data collection.

Another manipulation that needs to be done is to eliminate some unnecessary observations. Notably, observations that show levels of pollution below or equal to 0. Indeed, in this day and age, it is impossible to get such levels of pollution, as some of these gases are always present in the air.

1.2 Summary of the Data

Table 1.2 – Productivity Detailed Summary

	Percentiles	Smallest		
1%	1	1		
5%	1	1		
10%	1	1	Observations	92,565
25%	1	1	Sum of Weight	92,565
50%	2		Mean	2.236505
		Largest	St. Dev	1.860389
75%	3	90		
90%	4	98	Variance	3.451048
95%	5	100	Skewness	10.23495
99%	8	117	Kurtosis	436.1894

Notes: Detailed summary statistics of daily productivity, where productivity represents the number of cases closed by each judge on a given day.

Table 1.2 shows a large variation of productivity in the data. Ranging from 1 to 117 cases in a day, a possible cause of that could be the difference in case complexity. However, the database used does not differentiates such cases, which could represent a weakness in the analysis. An article published by Chang et al. (2019) uses the API (air pollution index) as their measure of pollution but mentions that PM (micro-particles) are the main drivers of API. Consequently, this present thesis uses $Pm_{2.5}$ as a proxy for pollution (micro-particles of size 2.5).

A working paper by Kahn and Li (2019) which discusses the impact of pollution and temperature on Chinese judges. Using a fixed effects model that clusters the standard errors, they conclude that pollution and higher temperatures have a negative impact on judges' productivity by increasing the amount of time taken to close cases. However, they find that the impact of pollution is much greater than the one of temperature and believe this to be because of air conditioning used for climate control. From the results of this paper, we predict that pollution will have a negative impact on the number of cases closed daily by judges.

1.3 Strengths and limitations of this dataset

1.3.1 Strengths

The data includes all asylum cases closed across all U.S. Immigration courts during the studied period, so the data is very complete in this aspect. Although geographically bound to certain courthouses due to applicants' locations, cases are randomly assigned to judges, so this should help avoid problems of endogeneity. Furthermore, the environmental data is very extensive, including a large variety of pollution and weather indicators, which should allow us to properly isolate the impact of an increase of $Pm_{2.5}$ on productivity.

1.3.2 Limitations

There is no information regarding the number of opened cases, only about those closed. Although most cases are settled within one hearing, some complicated cases do require multiple hearings, which could result in perceived lower productivity regardless of pollution levels. It would also be great to have data on the number of asylum demands. Such data could be useful to compare the number of cases treated, establishing a temporal trend and comparing it to the backlog of cases currently on stand-by. Analysing this backlog would show how important maintaining efficiency is in a system where people's lives are on stand-by while they wait for their cases to be treated. It would also be beneficial to have more detailed data on each judge, for example their age. That would have been useful in determining if $Pm_{2.5}$ affect productivity differently depending on age and general health status. Data about general economic activity could also be useful, as to establish the relationship between economic activity and pollution levels during the period of data collection, as this is an external factor that could also affect productivity without being directly caused by pollution.

Chapter 2

Econometric Model

As the goal is to look into the impact of pollution on productivity, the model used are fixed effects models, the first one being linear (equation 2.1) and the second (equation 2.2) being quadratic

$$productivity_{it} = \beta_0 + \beta_1 pm2.5_{it} + P_{it}\beta_3 + W_{it}\beta_4 + j_i + dow_t + ym_t + \varepsilon_{it}, \tag{2.1}$$

$$productivity_{it} = \beta_0 + \beta_1 pm2.5_{it} + pm2.5^2_{it}\beta_2 + P_{it}\beta_3 + W_{it}\beta_4 + j_i + dow_t + ym_t + \varepsilon_{it}, \tag{2.2}$$

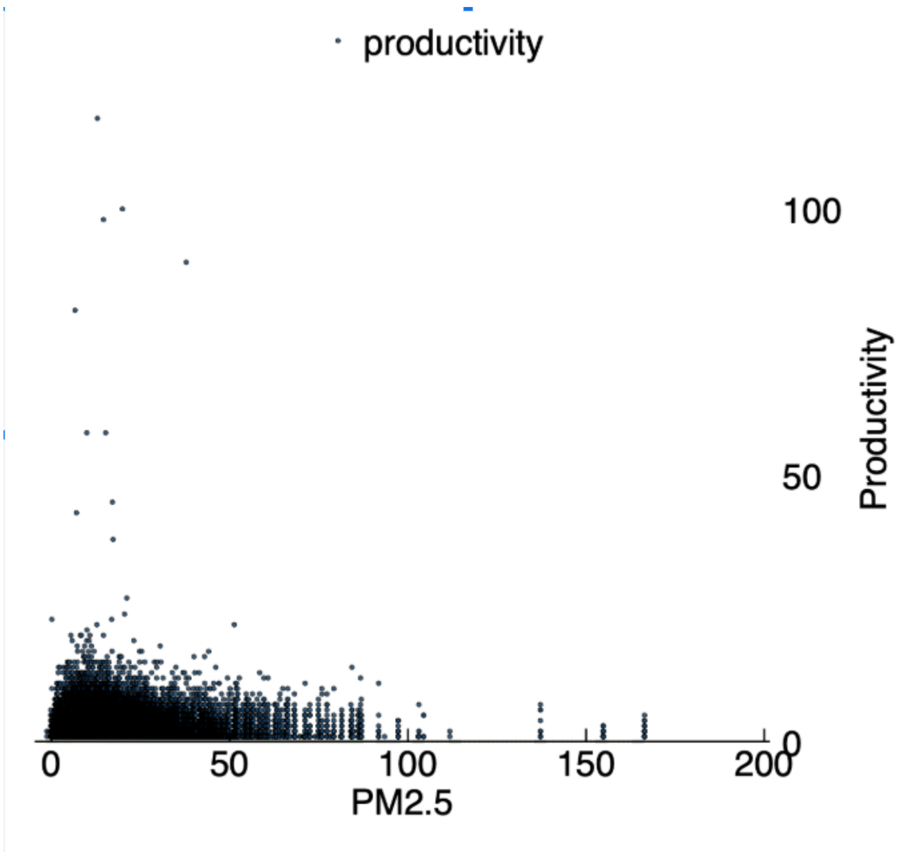
where $PM_{2.5}$ is the variable of interest, P represents a vector of other pollution indicators, W represents a vector of weather controls, j controls for the fixed effect of individual judges, dow represents the temporal effect of which day in the week it is, and ym controls for the temporal effect of the combination of year and month. The last two fixed effects mentioned aim to eliminate any temporal trend that could influence productivity. j is a vector of 266 dummy variables, one for each judge. The goal of adding those controls is to make sure that the variable of interest pm2.5 is not endogenous. Indeed, it is important to add the judge fixed effect because it would be sensible to assume that each individual has different innate productivity, regardless of external factors. Pollution can also affect people differently, depending on the sensitivity of each individual. The weather control includes data on temperature, pressure, precipitations, wind, and sky cover in order to eliminate any potential productivity variation caused by non-pollution-related weather factors. The other pollution indicators included in P are ozone and carbon monoxide. These are included in an attempt to distinguish whether $PM_{2.5}$ pollution affects productivity or simply air pollution in general. We allow for two-way clustering based on what has been done in the literature. We first cluster at the judge level since the error term is likely to show correlation between the same judges, as done by Kahn and Li (2019) and Chang et al. (2019). We also cluster at the city-month level to account for spatio-temporal correlation across cities and between each month. This cluster was also used by Heyes and Saberian (2019).

The inclusion of the quadratic term comes from observing figure 2.1 .

This graph seems to indicate that the relationship between pm2.5 and productivity is not linear, increasing rapidly at the beginning and slowly decreasing later. However, it also shows that most observations are concentrated in the lower end of the spectrum.

To ensure that the right model is selected, both linear and quadratic models were estimated.

Figure 2.1 – Histogram of pollution and productivity



Notes: This figure shows the distribution of productivity and Pm2.5 levels to get a preliminary idea of the relationship between pollution and productivity. It shows that the relationship is likely non-linear, as there is a rapid increase in productivity when Pm2.5 levels are low and increasing, after which productivity slowly decreases as Pm2.5 get more extreme.

Chapter 3

Results Analysis

3.1 Primary Results

This first section presents the main results obtained using the method detailed in the previous chapter. Table 3.1 shows results when the whole dataset is used, besides null and negatives values of $PM_{2.5}$. Column 1 represents the simplest model, which includes basic fixed effects related to individuals, time, and location, but does not include the quadratic term or the other control variables. On the other hand, column 6 represents the preferred model, which includes all the previously mentioned fixed effects adding control variables for other measures of pollution and weather. The goal of executing all six regressions is to ensure that the models do not lose their significance as additional parameters are added. As the table shows, when the quadratic term is not included, adding control variables causes the estimator of $PM_{2.5}$ to lose its' significance, while it is not the case when the quadratic term of $PM_{2.5}$ is included. This confirms the non-linear relationship hinted at by the simple observation made earlier. However, it is not enough to look at each coefficient separately. To be rigorous, we need to verify if β_1 and β_2 are jointly significant. Looking at the joint tests P-values for each quadratic regression, we see that Model 2 is strongly significant, model 6 is weakly significant, and model 4 is not significant.

Now looking more closely at equation 6, we get a $\beta_1 = 0.00336$, and $\beta_2 = -0.0000341$, which are both significant at a 5% level of significance. The marginal impact of $PM_{2.5}$ on productivity (the number of cases closed in a day) is $0.00336 - 0.0000682PM_{2.5}$. This would imply that as pollution levels are low, an increase in pollution has a positive impact on productivity, but as pollution levels increase, the effect gets smaller and ends up becoming negative after reaching a certain pivot point. At the average, an increase of 1 in $PM_{2.5}$ leads to an increase in the daily number of cases closed of 0.002373. Although it is statistically significant, in terms of tangible impact, it is very negligible. It is, therefore, more interesting to look at the impact of a change the size of one standard deviation.

The results of this can be found in table 3.3. Each row represents what happens to productivity when $Pm_{2.5}$ increases by one standard deviation (St.dev = 11.27727), ranging from the smallest value of $Pm_{2.5}$ to close to the maximum level. The column *% Mean Productivity* represents what that productivity change corresponds to when compared to the mean productivity. Therefore, we can see that around the average levels of productivity (row 2), an increase of one standard deviation of $Pm_{2.5}$ levels leads to an increase of around 1% productivity. Figure 3.2 plots these predictions and shows the corresponding confidence intervals. Each point in this figure represents a variation of $Pm_{2.5}$ the size of one standard deviation, and each bracket represents a confidence interval of 95%. It shows that, while the confidence intervals get larger as $Pm_{2.5}$ levels increase, the impact is still statistically significant as the confidence interval never cross 0.

Figure 3.1 shows the plotted regression of models 2 and 6 to contrast the quadratic models containing the fewest and highest amount of control variables. The pivot point is reached when levels of pm2.5 reach 49.222874. This, however, is much higher than the mean of 14.46656 observed in the data, more than one standard deviation higher. It is also much higher than the threshold of $15\mu g/m^3$ daily average recommended by the WHO. This surprising result contradicts the findings of Kahn and Li (2019) and Chang et al. (2019), who both find a simply negative impact on productivity. However, this might be caused by the greatly different levels of pollution observed in the samples. In the paper by Kahn and Li (2019), their sample average is $PM_{2.5} = 54.67$, which is significantly higher than our sample average, as well as being above the pivot point mentioned earlier, which

Table 3.1 – Main Results

VARIABLES	(1) productivity	(2) productivity	(3) productivity	(4) productivity	(5) productivity	(6) productivity
pm25	0.00197**	0.00491***	0.000850	0.00310**	0.000837	0.00336**
	(0.000840)	(0.00147)	(0.000824)	(0.00148)	(0.000856)	(0.00154)
sq_pm25		-4.12e-05***		-3.09e-05**		-3.41e-05**
		(1.48e-05)		(1.43e-05)		(1.47e-05)
Constant	2.208***	2.179***	2.104***	2.089***	-3.490	-3.524
	(0.0123)	(0.0176)	(0.0333)	(0.0350)	(3.165)	(3.163)
Observations	92,560	92,560	92,560	92,560	92,560	92,560
R-squared	0.085	0.085	0.086	0.086	0.087	0.087
Joint test p-value		0.003746		0.087244		0.064914
Judge FE	YES	YES	YES	YES	YES	YES
YearXMonth FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Pollution Control	NO	NO	YES	YES	YES	YES
Weather Control	NO	NO	NO	NO	YES	YES

Notes: The table represents the estimated effects of Pm2.5 on productivity with fixed effects models. Pm25 represents the daily Pm2.5 average, sq_pm25 is the squared value of pm25 and represents the quadratic term. Models 1, 3, and 5 are linear, while models 2, 4, and 6 are quadratic. Judge FE is the fixed effect to control for individual variation in productivity due to personal factors. YearXMonth FE is a temporal fixed effect that controls for yearly and seasonal variations in productivity. Day of Week FE is a fixed effect that controls for productivity variation throughout the week, for example, due to tiredness. Pollution Control is a vector of other pollution indicators, carbon monoxide and ozone, to make sure that Pm2.5 is indeed the cause of the change in productivity. Weather Control is a vector of other environmental necessary to distinguish pollution effects from other environmental effects. This vector includes daily averages of temperature, air pressure, dew point, precipitation levels, wind speed, and cloud coverage. Models 1 and 2 only include Judge, seasonal, and weekly fixed effects, without other environmental controls. Models 3 and 4 have the same fixed effects as previously mentioned but also control for other pollution variables. Models 5 and 6 control all the same things but also control for other weather related effects, making model 6 the less likely to have omitted variables. While model 6 is most likely to have omitted variables, it is only weakly significant when looking at the joint significance test. Clustered standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

could explain why they only find a negative impact.

An interesting point to look into would be to compare the least and most polluted cities in the sample and how pollution affects them differently. In the observed sample, the least polluted city is Hagatna, with a daily average of $PM_{2.5}$ levels of 6.045388, while the most polluted city in the sample is Los Angeles, with a daily average of $PM_{2.5}$ levels of 22.54007, as shown in table 3.2. On average, daily $PM_{2.5}$ levels are higher by 16.494682 in Los Angeles when compared to Hagatna. This is higher than a standard deviation and represents an increase of around 272.85%. The table also shows their respective data for productivity, where we can see that the average daily productivity in Hagatna is 2.171429 cases while it is 2.925559 in Los Angeles. This shows an increase in productivity of 0.75414, which represents an increase of 34.73% productivity when moving from Hagatna to Los Angeles. This, however, is merely a comparison of the observations. How much of this difference is caused by pollution? To answer that, we look at what the model would predict. For this, we use model 6. To calculate this, we use the equation 3.1, where X_1 and X_2 represent the daily average $PM_{2.5}$ levels for Hagatna and Los Angeles, respectively. This equation predicts an increase in productivity of 0.038159, which represents an increase of 1.76% when compared to Hagatna's average productivity. This shows that out of the 34.73% productivity difference between Hagatna and Los Angeles, only 1.76% can be explained by the change in $PM_{2.5}$, which is very little, all things considered.

$$\Delta \widehat{effect} = (\beta_1 X_2 - \beta_2 X_2^2) - (\beta_1 X_1 - \beta_2 X_1^2), \tag{3.1}$$

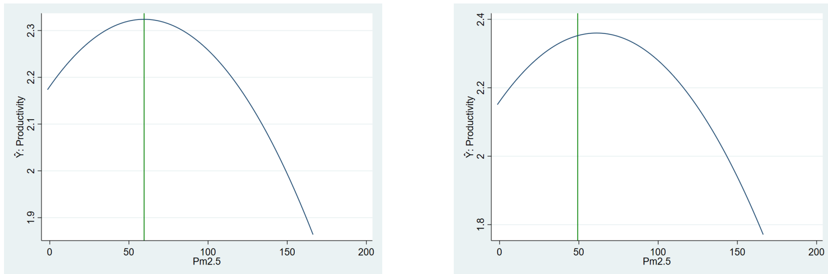
Table 3.2 – Productivity for the Least and Most Polluted Cities

City	Variable	Observations	Mean	St. Dev.	Minimum	Maximum
Hagatna	Pm2.5	35	6.045388	3.974297	1	17.9
	Productivity	35	2.171429	2.345029	1	13
Los Angeles	Pm2.5	13,395	22.54007	16.57502	0.8	166.3
	Productivity	13,395	2.925569	2.098351	1	22

Notes: Table shows summary statistics for Pm2.5 and Productivity in the least and most polluted cities in the sample. It shows that, on average, Los Angeles has 272.85% more Pm2.5 in the air compared to Hagatna and is also 34.73% more productive. Equation 3.1, however, shows that only 1.76% variation in productivity is caused by the increase of Pm2.5.

Figure 3.1 – Plotted Main Results

(a) Regression Plot of model 2 (b) Regression Plot of Model 6



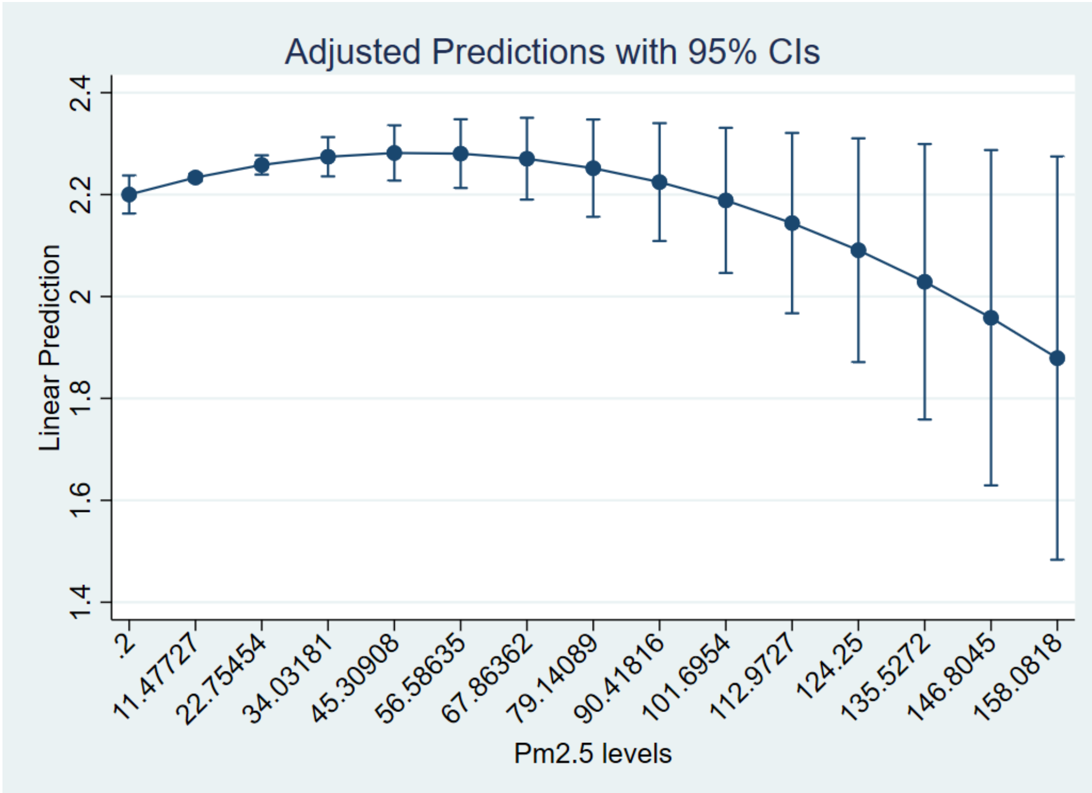
Notes: This figure shows the plotted predicted regressions of model2 and 6 to compare predicted models with and without additional control variables. The pivot point of each model is represented by the green vertical line, after which the marginal effect of increasing pollution leads to a negative impact on productivity. For model 2, the pivot point is 59.587379, while the pivot point is 49.222874 for model 6.

Table 3.3 – Estimated Productivity Size Effects for Model 6

Change in Pm2.5 (St.dev = 11.27727)	ΔProductivity	% Mean Productivity
From 0.2 to 11.47727	0.03340108	1.4934503
From 11.47727 to 22.75454	0.02472762	1.105637
From 22.75454 to 34.03181	0.01605416	0.71782378
From 34.03181 to 45.30908	0.0073807	0.33001053
From 45.30908 to 56.58635	-0.00129276	-0.05780273
From 56.58635 to 67.86362	-0.00996622	-0.44561599
From 67.86362 to 79.14089	-0.01863968	-0.83342925
From 79.14089 to 90.41816	-0.02731314	-1.2212425
From 90.41816 to 101.69543	-0.0359866	-1.609558
From 101.69543 to 112.9727	-0.04466006	-1.996869
From 112.9727 to 124.24997	-0.05333351	-2.3846823
From 124.24997 to 135.52727	-0.06200697	-2.7724955
From 135.52727 to 146.80451	-0.07068143	-3.1603008
From 146.80451 to 158.08178	-0.07935389	-3.5481221

Notes: This table shows the variation in productivity when Pm2.5 levels increase by one standard deviation, or $11.27727 \mu g/m^3$ Column 1 represents the jumps from the smallest value of Pm2.5 to the closest standard deviation interval to the largest value of the sample, being 166.3046. Column 2 represents the variation in productivity resulting from this change, calculated using equation 3.1. Column 3 represents what this variation represents compared to the average productivity. This is done by multiplying the results from equation 3.1, multiplying it by 100, and dividing it by the average productivity, which is 2.236505. Around the average Pm2.5 levels ($14.46656 \mu g/m^3$). A one standard deviation increase of Pm2.5 leads to an increase in productivity of around 1%.

Figure 3.2 – Confidence Interval Analysis



Notes: Figure plots the adjusted predictions of regression 6 with a 95% confidence interval. Elements in the x-axis are values of $Pm_{2.5}$, and each tick represents a jump the size of one standard deviation. The y-axis represents the predicted values of the regression. Each bracket represents a confidence interval of 95%. This graph shows that, while the confidence interval gets larger as levels of $Pm_{2.5}$ increase, the confidence intervals indicate that these predictions are still significant, as the confidence intervals never cross 0.

3.2 Extreme Pollution Analysis

In this second part of the analysis, the sample is further reduced to only include the observations up to the 99^{th} percentile in levels of $PM_{2.5}$. We are then left with the sample shown in table 3.4. In this instance, we can see that around 900 observations were dropped, which leads to the new highest value of $PM_{2.5} = 56.375$ instead of $PM_{2.5} = 166.3046$. This is a significant decrease but does not impact the average very much, which drops by around 0.5. The goal of this is to get an idea of whether the impact of pollution on productivity differs when we exclude the most extreme values of pollution. While this is not a formal method, the goal of this analysis is more exploratory than anything.

Table 3.4 – Pm2.5 Detailed Summary: Subsample

	Percentiles	Smallest		
1%	1.541667	0.2		
5%	3.672917	0.2		
10%	5	0.2	Observations	91,640
25%	7.8	0.2	Sum of Weight	91,640
50%	11.55417		Mean	13.8365
		Largest	St. Dev	9.046607
75%	17.5	56.3		
90%	25.3	56.3	Variance	81.84109
95%	32.5	56.375	Skewness	1.597922
99%	46.5	56.375	Kurtosis	6.213194

Notes: Detailed summary statistics of daily Pm2.5 levels measured in $\mu g/m^3$ across the 43 cities where courthouses are located from 2000 to 2004 when negative values of Pm2.5 and values of Pm2.5 above the $99^t h$ percentile are removed from the sample.

Once this is done, we can execute the same regressions as before, which leads to the results found in table 3.5. In this table, the negative impact is completely eliminated, which solidifies the hypothesis discussed in section 3.1 about the source of the difference between our results and previous research. It also shows that once we start to introduce control variables, the models lose their significance, whether for the regular or the quadratic model. The only significant coefficients are found in column one. However, that does not mean that column one represents the right model since we have concluded in the previous section that excluding control variables was likely to introduce endogeneity through the omitted variable bias. When looking at the joint significance tests, we can see that Model 2, which does have a quadratic coefficient but does not control for weather and pollution factors, becomes the most significant, suggesting that when we are looking at relatively lower levels of pollution, the impact of pollution on productivity, while statistically significant, cannot be distinguished from weather effects.

Table 3.5 – Secondary Results

VARIABLES	(1) productivity	(2) productivity	(3) productivity	(4) productivity	(5) productivity	(6) productivity
pm25	0.00288***	0.00252	0.00159	0.00117	0.00171	0.00156
	(0.000980)	(0.00234)	(0.00103)	(0.00227)	(0.00106)	(0.00248)
sq_pm25		8.52e-06		1.01e-05		3.51e-06
		(5.39e-05)		(5.30e-05)		(5.60e-05)
Constant	2.193***	2.196***	2.106***	2.109***	-3.557	-3.556
	(0.0139)	(0.0205)	(0.0346)	(0.0399)	(3.205)	(3.210)
Observations	91,635	91,635	91,635	91,635	91,635	91,635
R-squared	0.084	0.084	0.085	0.085	0.086	0.086
Joint test p-value		0.0123		0.2960126		0.26904543
Judge FE	YES	YES	YES	YES	YES	YES
YearXMonth FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Pollution Control	NO	NO	YES	YES	YES	YES
Weather Control	NO	NO	NO	NO	YES	YES

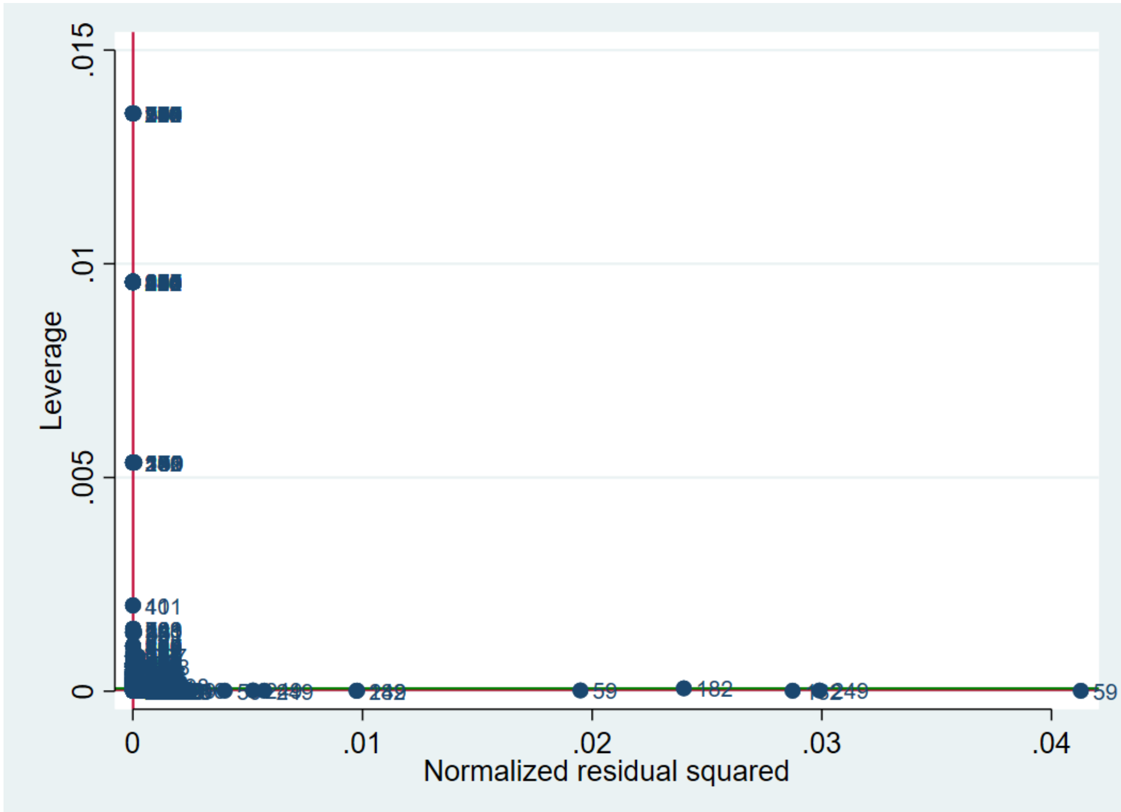
Notes: Table represents the estimated effects of Pm2.5 on productivity using the same fixed effect models and controls as in the primary analysis, but when observations with Pm2.5 levels above the $99^t h$ percentile are excluded. The joint significance test shows that only model 2 is significant, meaning that with this method, it is impossible to distinguish Pm2.5 effects from other environmental effects. Clustered standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

3.3 Outlier Analysis

This last section analyses the presence of potential outliers and their impact on the model. To identify potential outliers, we studentized the residuals by doing a simple regression of productivity on $PM_{2.5}$ and $PM_{2.5}^2$ and predicted the studentized residuals to see which judges (chair) could be outliers (as the proper way to deal with outliers in non-linear panel data is currently the subject of research and discussion). Following the guidelines provided by UCLA's Institute for Digital Research & Education, we look closer at the observations that have studentized residuals above 2, 2.5, and 3, respectively.

Then, looking at the leverage of the observations, we identify observations that deviated highly from the mean. The comparison of outliers and leverage points led to the following graphic.

Figure 3.3 – Outlier and Leverage Analysis



Notes: Figure plotting potential outliers versus the leverage of each observation. Each point represents an observation and shows how much of an outlier they are and how much leverage they have on the model. It shows that the outliers do not have the most leverage. With this, it is unclear whether there are conflicting outliers to remove from the data

Observing figure 3.3, it would seem that the biggest outliers are not the observations with the highest leverage, which makes it impossible to see if there were indeed influential outliers to remove from the data. For this reason, we go back to the analysis of the studentized residuals and see manually what happens when the most important outliers are removed. We started by removing all observations that had a studentized residual ≥ 3, going in increments of 0.5, and finishing with those with studentized residuals ≥ 2.

When looking at the detailed summary of $PM_{2.5}$ when studentized residuals $r \geq 3$ are removed in table 3.6, we can see that, interestingly, the largest levels of $PM_{2.5}$ are not likely to be outliers, and neither are the smallest ones. We can also see that the average and standard deviation barely change. The same type of results are found when we are more stringent and remove studentized residuals $r \geq 2.5$ and $r \geq 2$, as shown respectively in table 3.7 and table 3.8.

Table 3.6 – $PM_{2.5}$ Detailed Summary: excluding $r \geq 3$

	Percentiles	Smallest		
1%	1.577084	0.2		
5%	3.7	0.2		
10%	5	0.2	Observations	91,200
25%	7.8	0.2	Sum of Weight	91,200
50%	11.7		Mean	14.45199
		Largest	St. Dev	11.28134
75%	17.7	166.3046		
90%	26.3	166.3046	Variance	127.2686
95%	34.6	166.3046	Skewness	3.596105
99%	56.3	166.3046	Kurtosis	30.38218

Notes: Detailed summary statistics of daily Pm2.5 levels measured in $\mu g/m^3$ across the 43 cities where courthouses are located from 2000 to 2004 when negative values of Pm2.5 and values of Pm2.5 with studentized residuals above 3 are removed. The table shows that removing these has very little impact on global statistics.

Table 3.9 shows that removing outliers with studentized residuals of 3 and above makes the preferred model 6 now weakly significant when looking at the joint significance test. In this case, model 2 seems to be the most significant, showing that at this point, it would not be possible to properly distinguish between the impact of temperature and pollution on productivity. Also, This could indicate that $PM_{2.5}$ may not be the main driver of the pollution's impact on productivity. This corroborates the results found in table 1 of appendix A. This table is an adaptation of the regression made by Heyes and Saberian (2019) and shows that the pollution element that seems to have a more significant impact would rather be carbon monoxide rather than $PM_{2.5}$, which is rather surprising.

Figure 3.4 Shows the Plotted regressions of Model 2 and 6, which respectively have pivot points of 76.94717 and 69.422886. This shows that when we remove residuals $r \geq 3$, the negative impact on productivity is felt at a much higher point than in the initial analysis, and this time the point is even higher than the average of $PM_{2.5} = 54.67$ found by Kahn and Li, who had found a negative linear effect.

Table 3.7 – $PM_{2.5}$ Detailed Summary: excluding $r \geq 2.5$

	Percentiles	Smallest		
1%	1.6	0.2		
5%	3.7	0.2		
10%	5	0.2	Observations	91,184
25%	7.8	0.2	Sum of Weight	91,184
50%	11.67292		Mean	14.45775
		Largest	St. Dev	11.28975
75%	17.75	166.3046		
90%	26.3	166.3046	Variance	127.4585
95%	34.6	166.3046	Skewness	3.59423
99%	56.3	166.3046	Kurtosis	30.38942

Notes: Detailed summary statistics of daily Pm2.5 levels measured in $\mu g/m^3$ across the 43 cities where courthouses are located from 2000 to 2004 when negative values of Pm2.5 and values of Pm2.5 with studentized residuals above 2.5 are removed. The table shows that removing these has very little impact on global statistics.

Table 3.8 – $PM_{2.5}$ Detailed Summary: excluding $r \geq 2$

	Percentiles	Smallest		
1%	1.6	0.2		
5%	3.6875	0.2		
10%	5	0.2	Observations	88,586
25%	7.8	0.2	Sum of Weight	88,586
50%	11.7		Mean	14.46782
		Largest	St. Dev	11.27394
75%	17.8	166.3046		
90%	26.3	166.3046	Variance	127.1017
95%	34.5	166.3046	Skewness	3.593241
99%	56.3	166.3046	Kurtosis	30.54152

Notes: Detailed summary statistics of daily Pm2.5 levels measured in $\mu g/m^3$ across the 43 cities where courthouses are located from 2000 to 2004 when negative values of Pm2.5 and values of Pm2.5 with studentized residuals above 2 are removed. The table shows that removing these has very little impact on global statistics.

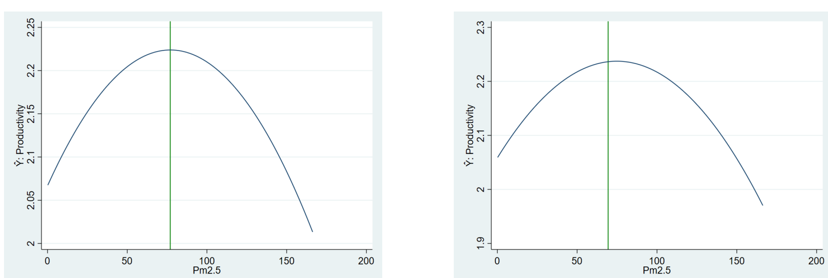
Table 3.9 – Results removing outliers $r \geq 3$

VARIABLES	(1) productivity	(2) productivity	(3) productivity	(4) productivity	(5) productivity	(6) productivity
pm25	0.00219***	0.00408***	0.00139*	0.00279**	0.00130*	0.00279**
	(0.000788)	(0.00125)	(0.000750)	(0.00127)	(0.000770)	(0.00129)
sq_pm25		-2.65e-05*		-1.92e-05		-2.01e-05
		(1.42e-05)		(1.42e-05)		(1.41e-05)
Constant	2.085***	2.067***	2.015***	2.006***	-0.381	-0.401
	(0.0110)	(0.0143)	(0.0230)	(0.0237)	(0.688)	(0.689)
Observations	91,195	91,195	91,195	91,195	91,195	91,195
R-squared	0.104	0.105	0.105	0.105	0.106	0.106
Joint test p-value		0.00200489		0.06199002		0.07872634
Judge FE	YES	YES	YES	YES	YES	YES
YearXMonth FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Pollution Control	NO	NO	YES	YES	YES	YES
Weather Control	NO	NO	NO	NO	YES	YES

Notes: Table represents the estimated effects of Pm2.5 on productivity with the same model specifications as in the primary analysis, but when observations with studentized residuals above 3 are removed from the sample. The joint significance test shows that models 4 and 6 are weakly significant, while model 2 is strongly significant. We are now unable to strongly affirm that it is indeed Pm2.5 levels that affect productivity. Clustered standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Figure 3.4 – Plotted Results Removing studentized residuals ≥ 3

(a) Regression Plot of model 2 (b) Regression Plot of Model 6



Notes: This figure shows the plotted predicted regressions of model2 and 6. Model 2 has a pivot point of 76.94717 while model 6 has a pivot point of 69.422886, much higher than what was found in the primary analysis

Table 3.10 shows that when removing outliers with studentized residuals of 2.5 or greater, model 6 becomes once again the most significant. This is because, although all three quadratic models are now highly significant (when looking at the joint significance test), model 6 has more control variables. In this model, however, the pivot point of the regression is now 74.74902, as shown in figure 3.5, which is much higher than the American and Chinese averages of $Pm_{2.5}$ discussed in the primary analysis.

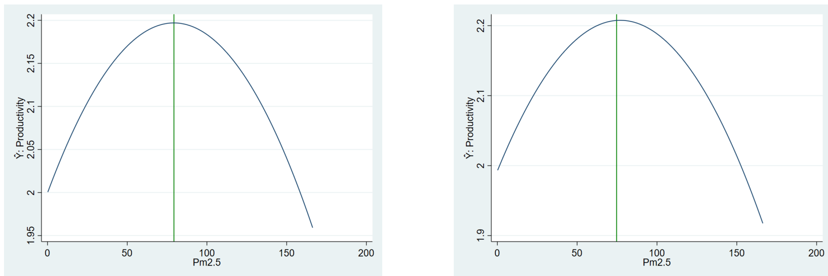
Table 3.10 – Results removing outliers $r \geq 2.5$

VARIABLES	(1) productivity	(2) productivity	(3) productivity	(4) productivity	(5) productivity	(6) productivity
pm25	0.00273***	0.00498***	0.00205***	0.00389***	0.00197**	0.00390***
	(0.000783)	(0.00120)	(0.000749)	(0.00121)	(0.000772)	(0.00124)
sq_pm25		-3.14e-05**		-2.51e-05**		-2.61e-05**
		(1.26e-05)		(1.25e-05)		(1.25e-05)
Constant	2.021***	2.000***	1.957***	1.945***	-0.190	-0.217
	(0.0112)	(0.0138)	(0.0213)	(0.0221)	(0.643)	(0.644)
Observations	90,179	90,179	90,179	90,179	90,179	90,179
R-squared	0.105	0.105	0.106	0.106	0.106	0.106
Joint test p-value		0.00006051		0.00296876		0.00442684
Judge FE	YES	YES	YES	YES	YES	YES
YearXMonth FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Pollution Control	NO	NO	YES	YES	YES	YES
Weather Control	NO	NO	NO	NO	YES	YES

Notes: Table represents the estimated effects of Pm2.5 on productivity with the same model specifications as in the primary analysis, but when observations with studentized residuals above 2.5 are removed from the sample. The joint significance test shows that model 6 is now strongly significant. Clustered standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Figure 3.5 – Plotted Results Removing studentized residuals ≥ 2.5

(a) Regression Plot of model 2 (b) Regression Plot of Model 6

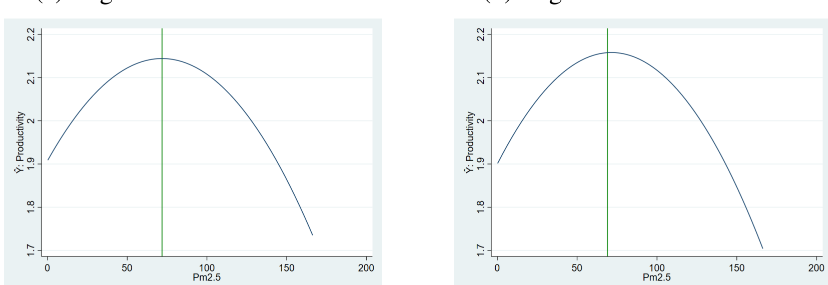


Notes: This figure shows the plotted predicted regressions of model2 and 6. Model 2 has a pivot point of 76.286624 while model 6 has a pivot point of 74.74902, even higher than what was found when excluding studentized residuals above 3

Table 3.11, which are the same regression results but when studentized residuals above 2 are excluded, gives similar results showing that, while these potential outliers might not be very influential on the model, model 6 is now even more statistically significant than before when looking at the joint significance test. However, the whole model shifts to the right compared to the original analysis, with model 6 having a pivot point of 68.944844, which is again much higher than what was discussed in the primary analysis, and goes against the hypothesis discussed about why the results obtained differed from those of Chinese studies. The pivot point, however, is lower than in the other two outlier analysis.

Figure 3.6 – Plotted Results Removing studentized residuals ≥ 2

(a) Regression Plot of model 2 (b) Regression Plot of Model 6



Notes: Figure shows the plotted predicted regressions of models 2 and 6 when studentized residuals above 2 are excluded. Model 2 has a pivot point of 71.866812, while model 6 has a pivot point of 68.929257, which is higher than the results in the primary analysis but lower than when residuals above 3 and 2.5 are excluded.

Table 3.11 – Results removing outliers $r \geq 2$

VARIABLES	(1) productivity	(2) productivity	(3) productivity	(4) productivity	(5) productivity	(6) productivity
pm25	0.00331***	0.00658***	0.00271***	0.00567***	0.00266***	0.00575***
	(0.000725)	(0.00108)	(0.000704)	(0.00108)	(0.000730)	(0.00109)
sq_pm25		-4.58e-05***		-4.06e-05***		-4.17e-05***
		(9.36e-06)		(9.34e-06)		(9.15e-06)
Constant	1.940***	1.908***	1.883***	1.864***	-0.173	-0.216
	(0.0107)	(0.0130)	(0.0181)	(0.0194)	(0.604)	(0.602)
Observations	88,581	88,581	88,581	88,581	88,581	88,581
R-squared	0.104	0.104	0.104	0.105	0.105	0.105
Joint test p-value		0.00000001		0.00000096		0.00000091
Judge FE	YES	YES	YES	YES	YES	YES
YearXMonth FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Pollution Control	NO	NO	YES	YES	YES	YES
Weather Control	NO	NO	NO	NO	YES	YES

Notes: The table represents the estimated effects of Pm2.5 on productivity with the same model specifications as in the primary analysis, but when observations with studentized residuals above 2 are removed from the sample. Joint significance test shows that once again, model 6 is now strongly significant. Clustered standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

According to the Pennsylvania State University's Department of Statistics, any observation with a studentized residual above 3 can be considered an outlier. In this case, the results shown in table 3.9 and figure 3.4 should probably be the results to retain, however, we cannot be one hundred percent of this as the method to analyse outliers in panel data is still a subject of debate and discussion.

3.4 Additional Discussion

Although the positive part of the relationship found between air pollution and productivity are counter-intuitive and difficult to explain, there might still be an explanation for it. Indeed, it could simply be that at low levels of pollution, an increase of air pollution could coincide with an increase in economic activity in the urban regions, which could explain more cases being brought to the court on a given day. Then, as pollution and economic activity continue to grow, this effect could be crowded out when pollution levels are too high. The data used in the context of this paper does not permit to measure out overall economic activity, further research exploring this phenomenon could be very useful in determining the direct impact of pollution on worker productivity.

Conclusion

In this thesis, we analyse the impact of pollution on the productivity of indoor workers, particularly on judges of U.S. immigration courts. A first analysis shows that the marginal impact of an increase in levels of $PM_{2.5}$ leads to an increase in productivity of $0.00336 - 0.0000682PM_{2.5}$, and starts to have a negative impact once we reach pollution levels of $PM_{2.5} = 49.2668$. These results are different from previous research in the way that pollution impacts productivity when low levels of air pollution are observed. This part of the impact could possibly be linked to increased economic activity leading to increased pollution when pollution levels are low. It is very interesting to compare these results to what has been found in previous research, as the positive part of the impact has not been detected before and deserved to be explored further than simple hypothesis.

An extreme value analysis shows that extreme levels of pollution levels are important drivers of the impact of the effect of $Pm_{2.5}$ on productivity. When pollution levels above the $99^t h$ percentile are excluded, the results become inconclusive. Indeed, the non-linear remains significant, but it also becomes impossible to clearly determine if pollution impacts productivity since we cannot control for weather and other polluting elements.

The outlier analysis, gives similar results to the subsample analysis. By removing outliers with studentized residuals of 3 and above, we are unable to distinguish between the impact of pollution and weather. When we are stricter, however, and we remove outliers with studentized residuals above 2 or 2.5, we become once again able to isolate the impact of $Pm_{2.5}$ on productivity.

However, it is very important to note that the sample used in this paper takes data collected from 2000 to 2004. In this case, it is very likely that average levels of pollution have changed since then. It would therefore be very interesting to replicate this study using more recent data. If average levels of pollution have increased since then, it would be interesting to see whether we are able to find the same results found by the Chinese researchers or if the quadratic element of the model is still present.

Although the results found here did not show the immediate negative impact expected of particular matter pollution on worker productivity, it does not mean that it does not have any impact long term. Indeed, as repeated exposure to high levels of pollution negatively impact health, it could very likely cause health problems later on in life, which could in term affect productivity by either slowing down daily productivity or directly affecting the number of days working because of the previously mentioned health problems.

Also, as the turning points in which $Pm_{2.5}$ levels start to have negative impacts on productivity is much higher than the threshold of $15\mu g/m^3$ given in the WHO guidelines, it would be interesting to do further research on whether that threshold is the right one.

Bibliography

Adhvaryu, A., Kala, N., and Nyshadham, A. (2022). Management and shocks to worker productivity. *Journal of Political Economy*, 130(1):1–47.

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

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

Coviello, D., Deserranno, E., Persico, N., and Sapienza, P. (2021). Effect of mood and worker incentives on workplace productivity.

Coviello, D., Ichino, A., and Persico, N. (2015). The Inefficiency of Worker Time Use. *Journal of the European Economic Association*, 13(5):906–947.

Graff Zivin, J. and Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–73.

He, J., Liu, H., and Salvo, A. (2019). Severe air pollution and labor productivity: Evidence from industrial towns in china. *American Economic Journal: Applied Economics*, 11(1):173–201.

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

Kahn, M. E. and Li, P. (2019). The effect of pollution and heat on high skill public sector worker productivity in china. Working Paper 25594, National Bureau of Economic Research.

Lee, J. J., Gino, F., and Staats, B. R. (2014). Rainmakers: Why bad weather means good productivity. *Journal of Applied Psychology*, 99(3):504.

National Association of Immigration Judges (2021). NAIJ Position on Legal Representation in Immigration Court. `https://www.naij-usa.org/images/uploads/newsroom/NAIJ_Position_on_Legal_Representation_-_Final_5.5.2021.pdf`.

National Association of Immigration Judges (2022). About the naij.

Pennsylvania State University (2021). Identifying outliers.

Saberian, S., Heyes, A., and Rivers, N. (2017). Alerts work! air quality warnings and cycling. *Resource and Energy Economics*, 49:165–185.

Slavin, D. N. and Marks, D. L. (2015). *3. You Be the Judge: Who Should Preside Over Immigration Cases, Where, and How?*, pages 89–112. New York University Press.

UCLA (2021). Regression with stata chapter 2 – regression diagnostics.

United Nations (2021). The paris agreement.

World Health Orginazation (2021). Ambiant (outdoor) air pollution.

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

Appendix A – Additional figures

Table 1 – Temperature Results

VARIABLES	(1) productivity	(2) productivity	(3) productivity	(4) productivity	(5) productivity	(6) productivity
temp6t4	0.00127	0.00157	0.00157	0.000402	0.00103	0.000357
	(0.000952)	(0.00255)	(0.00177)	(0.00322)	(0.00177)	(0.00305)
temp6t410				0	0	0
				(0)	(1.49e-06)	(5.69e-07)
press6t4			0.197*	0.197*	0.186*	0.186*
			(0.105)	(0.105)	(0.106)	(0.106)
dew6t4			0.000429	0.000547	0.000434	0.000502
			(0.00162)	(0.00166)	(0.00160)	(0.00164)
prcp6t4			-0.0472	-0.0451	-0.144	-0.143
			(0.387)	(0.387)	(0.392)	(0.392)
wind6t4			-0.000576	-0.000580	0.00519*	0.00518*
			(0.00307)	(0.00307)	(0.00297)	(0.00296)
skycover			-0.00757	-0.00819	-0.00918	-0.00952
			(0.0295)	(0.0296)	(0.0296)	(0.0299)
ozone					0.0704	0.0717
					(0.889)	(0.887)
co					0.124***	0.124***
					(0.0336)	(0.0336)
pm25					0.000793	0.000784
					(0.000855)	(0.000864)
sq_temp		-3.08e-06		1.11e-05		6.37e-06
		(2.69e-05)		(2.77e-05)		(2.70e-05)
Constant	2.164***	2.158***	-3.702	-3.685	-3.500	-3.491
	(0.0539)	(0.0670)	(3.152)	(3.144)	(3.165)	(3.158)
Observations	92,645	92,645	92,645	92,645	92,645	92,645
R-squared	0.085	0.085	0.086	0.086	0.087	0.087
Judge FE	YES	YES	YES	YES	YES	YES
YearXMonth FE	YES	YES	YES	YES	YES	YES
Day of Week FE	YES	YES	YES	YES	YES	YES
Weather Control	NO	NO	YES	YES	YES	YES
Pollution Control	NO	NO	NO	NO	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1