DISSERTATION

PARTICULATE MATTER AND HUMAN VULNERABILITY: IMPACTS ON COGNITION, RESPIRATORY HEALTH, AND PUBLIC SAFETY

Submitted by

Wenjie Chen

Department of Agricultural and Resource Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Jesse Burkhardt

Christopher Goemans
Stephan Kroll
Terrence Iverson

ABSTRACT

PARTICULATE MATTER AND HUMAN VULNERABILITY: IMPACTS ON COGNITION, RESPIRATORY HEALTH, AND PUBLIC SAFETY

This dissertation explores the broad impacts of air pollution on human vulnerability by examining how environmental stressors affect cognitive performance, respiratory health, and public safety. Focusing primarily on fine particulate matter ($PM_{2.5}$), the study conducts three empirical analyses using behavioral, health, and transportation data. Chapter 1 uses chess performance as a proxy for cognitive function and estimates the effect of $PM_{2.5}$ exposure on decision-making using fixed-effects logistic regression and a beta regression model. The results show that higher $PM_{2.5}$ concentrations significantly reduce winning probability, with stronger effects observed among older, higher-income, and male players. Precipitation, relative humidity, rating difference, and age all have positive effects on winning percentage, while the opposite was true for the income variable. Chapter 2 investigates the link between air pollution and asthma prevalence using multinomial logistic regression and U.S. health survey data. Lagged exposure to $PM_{2.5}$ and SO_2 increases the likelihood of currently having asthma, while higher temperature and humidity reduce asthma risks. Women with higher BMI, higher frequency of smoking, poorer general health condition, and younger age have higher asthma recurrence rates. Interaction effects suggest that SO_2 has a weaker impact among older individuals. Chapter 3 analyzes crash-level data from Colorado between 2007 and 2020 to examine the relationship between air pollution, weather, and traffic accidents. Negative binomial regression results indicate that $PM_{2.5}$ and low visibility are associated with more frequent accidents. Accident counts are higher in locations and times with more DUI involvement, female drivers, $PM_{2.5}$ amplifies the effect of over speeding and DUI. The number of crashes is higher on Wednesdays and relatively low at night. Together, these chapters provide novel evidence that air pollution has diverse and significant consequences for cognitive capacity, chronic disease, and public safety, underscoring the urgent need for coordinated environmental and public health policies.

ACKNOWLEDGEMENTS

I would like to offer my deepest gratitude to my advisor, Dr. Jesse Burkhardt, for his professional guidance, suggestions, and encouragement throughout the research process. His inputs and comments greatly improved the quality of this thesis. I also thank committee members Dr. Christopher Goemans, Dr. Stephan Kroll, Department of Agricultural and Resource Economics, and Dr. Terrence Iverson, Department of Economics, Colorado State University, for creating a supportive academic environment. In addition, I acknowledge the use of OpenAI and DeepL as writing assistance aids during composition. It was used to polish language, strengthen clarity and academic tone, and assist with summarizing numbers and document structure. All the material generated with AI assistance was extensively examined, corrected, and finished by me to ensure accuracy and originality. The entire research design, data analysis, and conclusions represented in this thesis are original work. All the data and code used in this thesis can be found in an open-access online Github repository to allow for transparency and reproducibility of findings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Air Quality and Human Behavior on Chess

1. Introduction

Particulate matter (PM) is tiny airborne particles that can be absorbed by the human respiratory system. Among all the types of PM, the one that is less than 2.5 micrometers in diameter is known as $PM_{2.5}$. These particles are so fine that they can be inhaled deep into the lungs, where they become trapped and can have various health outcomes.

Exposure to $PM_{2.5}$ has numerous adverse health impacts. It contributes to an increased susceptibility to cardiovascular and respiratory disease, reduced lung function, and enhanced risk of premature death. $PM_{2.5}$ has the potential to lead to several cardiovascular and neurologic conditions, such as a heightened risk of heart attack, stroke, and other cardiovascular events. Long-term $PM_{2.5}$ exposure has also been found to contribute to the incidence of chronic cardiovascular diseases and hypertension, while short-term exposure can cause acute cardiovascular events in susceptible subjects.

One of the landmark studies from Augsburg, Germany (Pfeiffer et al. 2015), looked at heart disease death rates of over 15,000 between 1995 and 2009, along with weather and pollution information, including $PM_{2.5}$. The authors said that $PM_{2.5}$ has a significant effect on myocardial infarction in the short term, particularly in those with a history of heart disease.

The adverse physical effects of $PM_{2.5}$ also fall on mental health and intellectual ability. Continuous exposure to air pollution has been discovered to impair cognitive task performance in verbal and mathematical tests, especially in older adults. The impact on cognitive abilities can influence daily decision-making and behavior. For instance, the study by Joris Klingen and Jos van Ommeren reveals that air pollution, including $PM_{2.5}$, makes individuals less cautious. A case in point is the decline in the SSE Composite Index, which was observed to recover by 0.18% after a hike in $PM_{2.5}$ concentration.

Building on this line of inquiry, this study investigates how $PM_{2.5}$ pollution affects cognitive performance in the highly structured and mentally demanding context of chess games. By examining chess, a domain that isolates pure intellectual ability, we provide micro-level evidence on the cognitive effects of air pollution.

1.1 Research Question and Hypotheses

This chapter explores how short-term $PM_{2.5}$ exposure affects high-level cognitive performance utilizing online chess games as an empirical setting. Chess provides a standardized, competitive environment that demands sustained attention, strategic thinking, and rapid decision-making. By examining game outcomes and winning rates, this study investigates whether higher pollution levels affect decision quality and cognitive resilience under pressure.

This paper builds on existing research by examining the number of wins in chess games and the winning rate, taking into account factors such as income and age. By controlling for time effects, the study seeks to isolate the impact of weather conditions and air pollution on these cognitive performance outcomes. The findings have important implications for public health policy and environmental regulations, highlighting how elevated pollution levels can impair cognitive function and decision-making. This research underscores the need for continued monitoring and regulation of air quality to protect not only physical health but also cognitive performance in the general population.

This study proposes the following hypotheses:

H1: Higher $PM_{2.5}$ levels harm cognitive performance, resulting in a lower chance of winning chess games. (Variables: $PM_{2.5}$(current), $PM_{2.5}$(20-hour lag))

H2: Other environmental factors, such as temperature, precipitation, humidity, and sea level pressure, also correlate with cognitive performance and strategic decision-making, and are included as control variables to isolate the effect of $PM_{2.5}$. (Control variables: Temperature, Precipitation, Relative Humidity, Sea Level Pressure)

H3: The impact of pollution on cognition varies and is influenced by individual traits like age and income. (Interaction Variables: $PM_{2.5} \times$ Rating Difference, $PM_{2.5} \times$ Age, $PM_{2.5} \times$ Income, $PM_{2.5} \times$ Sex)

1.2 Significance of the Research Question

Understanding how $PM_{2.5}$ affects cognition is crucial not only for public health and environmental policy but also for fields such as cognitive science and behavioral economics. Unlike previous studies that examine overall cognitive test performance or reaction-based performance among esports players, this article contributes theoretically by investigating prolonged, strategic thinking in the context of competitive chess. By holding time variables constant and utilizing detailed EPA $PM_{2.5}$ and NOAA weather data from 2015 to 2018, this research controls for the effects of environmental stressors on cognitive outcomes.

Furthermore, the findings of this study may inform air quality policy by highlighting, alongside physical health consequences, the impact on cognitive productivity and decision-making quality. The study is particularly significant given emerging urbanization, and the cognitive load placed on human beings in modern societies.

Subsequent research can build upon this work by utilizing more recent data to analyze long-term trends. Furthermore, examining other air pollutants and interactions among multiple air pollutants could enhance our understanding of the health impact of air pollution. Machine learning algorithms and sophisticated statistical analysis can enhance the predictive capability and accuracy of the analysis, helping to improve public health interventions.

1.3 Literature Review and Novelty

Several studies have demonstrated that particulate matter (PM) increases stress hormones and blood pressure, thereby elevating individual stress levels. One such study, Risk Attitude and Air Pollution: Evidence from Chess (Klingen & van Ommeren, 2020), examines the short-term effects of respirable PM on decision-making using chess as an experimental setting. The study assumes that PM impairs cognitive function, influencing risk preferences—e.g., increasing health insurance uptake while reducing lottery participation. Chess was chosen due to its short duration, allowing tighter control over time- and location-specific factors, as PM levels are not randomly distributed. A key methodological challenge—non-random variation in PM—was addressed by randomizing chess opponents. The study finds that higher PM levels are associated with reduced risk-taking. A notable insight is the sensitivity analysis: extending the measurement range reduced PM's estimated effect but improved precision, while narrowing it amplified the effect but widened confidence intervals. Another key point is the use of instrumental variables to mitigate attenuation bias stemming from measurement error due to distance from pollution monitors. Compared to this study, our research differs in focus, data, methods, and contributions. While Klingen and van Ommeren analyze how PM_{10} influences chess players' risk preferences (e.g., more draws), our study explores how $PM_{2.5}$ and weather conditions affect win rates and player performance, considering moderating factors such as age and income. Our dataset comprises online chess games from multiple U.S. states (2015–2018), supplemented with pollution data from the EPA and weather data from NOAA, providing broader coverage than the Netherlands-only tournament data used in their study. Methodologically, we applied binary logistic regression to model the likelihood of win/loss outcomes, linear regression with fixed effects to estimate daily win rates, and beta regression to account for heteroskedasticity. In contrast, their analysis primarily relied on fixed-effects regression. Overall, our study offers a more nuanced design that integrates environmental and individual factors, providing greater generalizability.

Health-related research consistently finds an inverse relationship between air pollution and mood. The study Air Pollution and Stock Returns in the US (Bullinger et al., 2011) investigates whether the emotional effects of pollution extend to economic behavior, specifically stock market performance. The authors propose three hypotheses: (1) a correlation between air pollutants and market volatility, (2) a stronger pollution–return relationship near polluted stock exchanges, and (3) the spillover effects of pollution on remote exchanges via local traders' sentiment. They find that stock returns decline on days with "unhealthy" air quality, particularly when pollution levels are high near the exchange. Notably, the influence of New York's pollution on Philadelphia's stock returns increases with the proportion of trades executed by traders based in New York. The study's findings support the notion that air pollution impacts investor mood and risk-taking, suggesting that AQI forecasts could inform trading strategies. In contrast, our study differs significantly in scope, data, methods, and contribution. While Bullinger et al. (2011) and Levy and Yagil (2011) examine how pollution affects market returns via sentiment and risk preferences, my research focuses on how air pollution impacts cognitive function and decision-making, specifically, chess performance and win rates. My dataset comprises online chess games across U.S. states (2015–2018), linked with $PM_{2.5}$ data from the EPA and weather data from NOAA. Unlike financial studies that use OLS regression on stock return data, we apply binary logistic regression, linear fixed-effects models, and beta regression to model chess outcomes while controlling for environmental and individual-level variables. This allows for a more direct assessment of how pollution affects complex cognitive tasks and strategic thinking. Ultimately, our research offers a more precise and more targeted analysis of pollution's impact on human decision-making than the broader sentiment-driven market studies. In a 2020 study, Berman and Ebisu analyzed air quality changes in the U.S. during the COVID-19 pandemic. They found that nitrogen dioxide (NO_2) levels dropped significantly—by about 25%—in many urban areas due to reduced traffic from stay-at-home orders. Fine particulate matter ($PM_{2.5}$) exhibited minor and more variable changes, while ozone (O_3) levels increased in some regions, likely due to complex atmospheric chemistry involving lower NO_2 concentrations. The findings demonstrate that reduced human activity, especially from transportation, can lead to notable improvements in air quality but also reveal the complex and pollutant-specific nature of atmospheric responses, underscoring the need for targeted environmental policies.

Jiang, Huang, and Fisher (2019) conducted a case study in Beijing to investigate how air pollution influences human behavior, particularly in relation to park visitation. Using stated preference techniques and surveys, they found that higher pollution levels significantly reduce individuals' likelihood of visiting urban parks. This effect was observed in both online and in-person surveys. It was more pronounced when pollution levels were relatively low, suggesting that even moderate air quality concerns can deter visits. The study also revealed a gap between people's behavior and their awareness of air pollution's health risks, likely influenced by learning, perception, and other behavioral factors. The authors emphasize the importance of considering lagged pollution exposure (e.g., previous-day AQI) when studying visitation decisions and highlight the strength of using quantile regression as a tool for analysis. Unlike ordinary least squares, quantile regression provides a fuller picture of how pollution affects different parts of the visitor frequency distribution and is more robust to outliers and non-normal errors. In contrast, our research differs substantially in research focus, data, methods, and contribution. While Jiang et al. examine behavioral avoidance of outdoor activity due to pollution, we investigate the direct cognitive impacts of air pollution on chess performance, specifically players' win rates. Our dataset encompasses online chess tournaments from multiple U.S. states (2015–2018), complemented by EPA $PM_{2.5}$ and NOAA weather data. We employ binary logistic and linear regression models, as well as beta regression, to examine how pollution and weather influence decision-making while controlling for factors such as age and income. Unlike Jiang et al., who focus on behavior frequency, our study targets the physiological and cognitive effects of pollution on performance in complex tasks, offering a different lens on how environmental stressors influence human functioning.

Zhang, Chen, and Zhang (2018) investigate the impact of air pollution on cognitive functioning, more specifically verbal and math test scores, by merging CFPS data with air pollution data. Fixed-effects regression analysis, which controls for individual, county, and time-fixed effects, concludes that both long- and short-term exposure to pollution reduces cognitive test scores considerably. The negative impacts are especially pronounced among older individuals and those with lower levels of education. One implication is that demographic variables such as age and gender moderate the cognitive impact of pollution. The study's falsification test also suggests the potential impact of unobserved time-invariant factors such as migration, highlighting the necessity of caution in the interpretation of the results without air quality controls after exposure. This study informs our project in two significant ways: first, it supports the hypothesis that the cognitive effects of pollution are heterogeneous within the population; second, it highlights methodological concerns, such as the need to control for latent confounders. Our project, however, differs significantly in focus, data, and methodology. While Zhang et al. reflect on general cognitive capabilities using standardized tests, we examine the influence of air pollution on the actual decision-making of high complexity by analyzing chess performance and winning rates. Our data include online chess games from different U.S. states (2015–2018), combined with $PM_{2.5}$ data from the EPA and weather information from NOAA. Methodologically, we employ binary logistic regression and linear regression to analyze game outcomes, controlling for demographic and environmental factors. Unlike Zhang et al.'s study on general abilities, ours focuses on higher-order cognition tasks under competition pressure, providing evidence of the impact of pollution on strategic thinking and performance in competitive high-stakes settings.

Chen and Schwartz (2009) investigate the impact of long-term exposure to PM_{10} and ozone on adult cognitive performance, utilizing data from NHANES III (1988–1994). They find that PM_{10} has a negative effect on neurobehavioral test outcomes, with results moderated by race and socioeconomic status. While their study focuses on pollutants similar to ours—albeit PM_{10} and ozone versus $PM_{2.5}$ —it differs in scope and design. Their analysis centers on standardized cognitive tests (e.g., Symbol-Number Substitution, Memory Recall) and uses generalized linear models to control demographic variables such as age, income, and education. Our study, by contrast, investigates how short-term exposure to $PM_{2.5}$ affects high-level cognitive performance through the lens of chess tournament outcomes. We utilize online game data from various U.S.

states (2015–2018), combined with EPA pollution and NOAA weather data, to examine how environmental and individual-level variables influence win rates and decision-making. Our methods include binary logistic and linear regression, incorporating a broader range of controls such as weather and individual traits, though not physiological health variables like obesity or blood pressure—an area for future refinement. Ultimately, while Chen and Schwartz focus on the general cognitive decline due to long-term exposure, our research highlights the real-time impact of air pollution on complex, strategic decision-making, offering a more applied view of how environmental stressors impair performance in cognitively demanding situations.

Marr and Ely (2010) examined the impact of air pollutants—NO_2, SO_2, O_3, and PM_{10}—on marathon performance through regression analysis of pollutant levels and race performance in major U.S. marathons. They find that higher levels of PM_{10} are associated with reduced performance, particularly for female runners. The regression accounts for environmental modifiers by covarying for Wet Bulb Globe Temperature (WBGT), a variable that incorporates temperature, humidity, wind speed, and solar radiation—parameters also relevant to air pollutant dispersion and human physiological response. While this study offers helpful insights into how pollution affects physical performance, our research varies considerably in terms of subject, data, and methodology. We examine the influence of $PM_{2.5}$ on intellectual performance in an intellectually demanding task—competitive chess—using online game data (2015–2018) from a number of U.S. states, merged with EPA pollution data and NOAA weather data. Using binary logistics and linear regression, we quantify the effect of air quality on game outcomes after controlling for individual demographic traits and environmental conditions. In contrast to Marr and Ely's focus on physical endurance and lung function, our work targets the cognitive and decision-making effects of pollution, highlighting the mental performance ramifications of environmental stressors in real-time, high-stress settings.

Jana et al. (2022) investigate the effect of indoor air pollution (IAP) on cognitive health in older adults in India, utilizing data from the 2017–2018 Longitudinal Study of Aging in India (LASI). Their analysis reveals that indoor pollutants negatively impact cognitive test scores among the elderly, as determined by OLS regression. Our study focused on the short-term effects of outdoor $PM_{2.5}$ pollution on cognitive performance in chess, which differs in methodology, data sources, and focus. We use online chess tournament data (2015–2018) from U.S. states, combined with EPA $PM_{2.5}$ and NOAA weather data. Our analysis employs binary logistic regression and linear regression, controlling for individual and environmental factors. Unlike Jana et al., who examine basic cognitive functions in an aging population, our research investigates strategic decision-making in chess. While their study addresses long-term indoor pollution risks, ours explores the immediate effect of outdoor pollution on real-time performance in competitive settings. We also control for individual characteristics (age, income level) to improve precision in assessing the impact of pollution. Our study contributes to air pollution literature by focusing on high-level cognitive tasks, offering implications for cognitive science, environmental economics, and public health policy.

Park and Kim (2022) explore how air pollution impacts risk appetite, finding that higher pollution levels reduce people's willingness to take risks, such as in venture capital or lottery ticket choices. Their study's careful sample adjustments and participant classifications (e.g., single vs. divorced) could inform my research. For example, the risk preferences of employed vs. unemployed chess players might differ and influence game outcomes. Christensen et al. (2022) examine how air pollution interacts with neighborhood socioeconomic status (nSES) to affect cognitive performance in older adults. They find that pollution alone doesn't always cause cognitive decline—low SES often exacerbates the effects. Their use of self-organizing maps (SOM) to visualize pollutant distributions is a useful spatial analysis tool. Our study, unlike theirs, focuses on how air pollution ($PM_{2.5}$) impacts cognitive abilities and win rates in chess, using online chess tournament data (2015–2018) from U.S. states, alongside $PM_{2.5}$ and weather data. We employ binary logistic regression for win-loss predictions and linear regression for performance trends while controlling for factors such as age and environmental variables. While Christensen et al. analyze long-term cognitive decline, we assess the short-term impact of pollution on high-level cognitive tasks, like strategy execution in chess, offering new insights into real-time decision-making and performance in competitive settings.

Park et al. (2022) investigate how air pollution influences risk tolerance, linking higher pollution levels with reduced risk appetite in scenarios like venture capital decisions and lottery ticket choices. We find their methodology valuable, particularly how they reduce sample bias by adjusting monitoring coverage based on population ratios. Their categorization of participants, such as distinguishing divorced individuals, helps mitigate bias from family status. Similarly, our study could classify chess players by employment status to assess how their risk preferences impact performance, given their distinct thinking patterns in tournaments. Christensen et al. (2022) investigate how neighborhood socioeconomic status (nSES) and air pollution jointly affect cognitive performance, finding that low SES exacerbates the cognitive decline caused by pollution. Their use of self-organizing maps (SOM) to visualize pollution dispersion is also a valuable tool for spatial analysis. Unlike Christensen et al. (2022), my study focuses on how short-term $PM_{2.5}$ exposure affects cognitive performance and win rates in chess. Their research explores the long-term cognitive decline due to air pollution and socioeconomic status, using data from the Emory Healthy Aging Study (EHAS) and sophisticated regression models (LASSO, BKMR, and G-computation). In contrast, I use online chess tournament data from multiple U.S. states (2015–2018), combined with EPA $PM_{2.5}$ and NOAA weather data. We analyze how pollution affects strategic thinking and decision-making in high-stakes chess games using binary logistic regression for win-loss predictions and linear regression for performance trends, while controlling for factors such as age and environmental conditions. While Christensen et al. focus on long-term cognitive decline, our research examines the immediate effects of air pollution on competitive performance in cognitively demanding tasks, offering new insights into how air pollution affects decision-making in real-time, high-performance settings like chess.

Künn et al. (2023) analyze the influence of indoor air pollution on chess decision-making in terms of $PM_{2.5}$ exposure. They analyzed data from three German tournaments between 2017 and 2019, rating over 30,000 moves by 121 players using the Stockfish AI engine. The findings indicate that higher indoor $PM_{2.5}$ levels increase the likelihood of committing key errors by 26.3%, particularly during time-sensitive stages of the game. The experiment also controlled for factors external to the experiment, e.g., traffic congestion and background pollution, which strengthens the causal link between indoor air quality and decision-making errors. Their study highlights the importance of indoor air quality to mental work, especially in challenging areas such as chess, finance, and law. Whereas Künn et al. are interested in decision quality on a step-by-step basis, our study extends their work by examining the long-term performance impacts, temperature and humidity factors, and daily winning percentages. We examine the impact of air pollution on not only decision-making but also long-term strategic endurance, providing a more comprehensive picture of its effect on cognitive function. Unlike Künn's study, which is based on indoor pollution measures, our study uses outdoor monitoring equipment to track pollutants and their lagged impacts on cognitive performance, providing evidence of the long-term consequences of exposure.

Mo et al. (2023) examined the effect of air pollution on competitive eSports performance, using data from 2,638 League of Legends Professional League (LPL) matches played in China between 2017 and 2021. They found that heightened pollution unfairly damaged lower-income teams, which made more errors, had faster game terminations, and a larger skill difference. Pollution primarily disrupted complex collaboration instead of reflexive reactions. Their work identifies environmental stressors in team settings and suggests that eSports tournaments be held in less polluted venues to provide an equitable playing field. Our work, however, involves chess, and it entails planning over the long term, intellectual endurance, and logical thinking. Unlike Mo et al., who studied short-term strategy and team performance, our work examines how chronic exposure to air pollution influences individual cognitive resilience in terms of long-term attentional focus and decision-making. Our work employs ongoing hourly and daily rates of pollution, whereas Mo's work only accounted for the pollution on the competition day. Our work examines chess to provide evidence of the effects of pollution on complex decision-making and cognitive stability in the long term.

Hoogeveen (2023) analyzed the impact of air pollution, as represented by $PM_{2.5}$ and the Air Quality Index (AQI), on eSports performance in 669 U.S. Counter-Strike: Global Offensive (CS: GO) LAN tournaments from 2016 to 2019. The study revealed that heightened $PM_{2.5}$ levels led to reduced match duration, decreased accuracy of players, and higher single-team dominance, reflecting cognitive deterioration and motor inaccuracy. Seasoned players were affected less, corroborating the theory that elite players are better equipped to deal with environmental stress.

The study refers to the impact of pollution on reflex-based decision-making and motor function and recommends that eSports competitions be held in cleaner environments to ensure fairness. In contrast, our research focuses on chess, a game that requires extended concentration and strategic thinking. In contrast to Hoogeveen's focus on short-term reflexes, our research examines the effect of air pollution on long-term cognitive function and strategic thinking. Our research also examines other environmental stressors such as temperature and humidity, providing a broader perspective on the impacts of multiple stressors on cognitive function.

Compared with the existing literature, this chapter is novel in several ways: It focuses on short-term pollution exposure rather than long-term chronic effects. It uses real-life performance in an intellectual task (chess) rather than survey-based risk assessments or standardized cognitive tests. It explores win/loss outcomes and daily winning percentages, offering a dynamic measure of decision performance. It incorporates a broad range of individual and environmental controls, including age, income, temperature, and humidity. The study builds on and differs from the work by Künn et al. (2023), who analyzed individual moves in indoor chess tournaments. While their study focuses on decision-by-decision error rates under indoor $PM_{2.5}$ exposure, our paper examines longer-term strategic resilience under outdoor pollution using a larger dataset across U.S. states. The research also extends beyond Esports-related studies (e.g., Mo et al., 2023; Hoogeveen, 2023) by examining not only short-term reactivity but also extended cognitive load under pollution.

2. Data

2.1 Data Sources

In this experiment, we collected more than 1,080,000 games from 71 players, spanning from January 1, 2015, to December 31, 2019, from unrbusiness.az1.qualtrics.com. To understand the basic profile of the players, we collected the age, gender, and income of the subjects. We used numbers 1-9 to indicate income status, with number 9 indicating that the subject had the highest income. The same goes for age. The sex variable is coded as a binary indicator, with female players assigned 0 and male players assigned 1. The study also collected data on the locations of chess players from all over the United States, to understand the local air quality conditions. In addition to $PM_{2.5}$, the main target of the experiment, we also collected data on the maximum temperature, minimum temperature, maximum humidity, minimum humidity, and rainfall for the day of the experiment. Rating differences will also affect the result, which is obvious, with the number indicating the level of rating. Hourly air pollutant ($PM_{2.5}$) data were obtained from the United States Environmental Protection Agency. The zip code $PM_{2.5}$ was determined by referencing the State and County codes to match the experimental area. Other hourly weather data (including temperature, humidity, etc.) were also obtained. Other hourly weather data (including temperature, humidity, etc.) were obtained from the National Centers for Environmental Information website, also by looking up the zip code corresponding to the experimental area.

2.2 Data Processing

Since the questionnaire covered only daily climatic conditions on the day of the match and did not include $PM_{2.5}$ data, we first downloaded hourly $PM_{2.5}$ data packages from the EPA, which contain data for various U.S. regions from 2015 through 2019, and selected the hourly $PM_{2.5}$ data corresponding to the zip code where each game was played. Secondly, hourly data on weather factors other than $PM_{2.5}$ for each zip code were downloaded from the NOAA database. The first step was to merge the $PM_{2.5}$ and weather data by creating a new column that combined time and zip. To facilitate the analysis of the winning rate model, we then averaged each variable over the hourly data to obtain the daily weather data. This one consolidated dataset contains 2,541,167 data points. The second step was to merge the games and players' information through the same player ID. This data set consists of the original questionnaire data for 'games', 'players', and 'weather'. This step is to merge the games and players' information data with the same player ID. Since the original $PM_{2.5}$ data did not include 2019 data, we downloaded the 2019 daily $PM_{2.5}$ data from the EPA database and merged it with the two previous data sets by zip code. The consolidated data file contains a total of 834,887 data points. To investigate how $PM_{2.5}$ and weather conditions impact the results of the chess, we added a new column, win or not (1 means win, and 0 means lose), to the data. All weather data for NA was removed from the original dataset to make the data more comprehensible. The next step was to further merge the weather data with personal information data. After removing blank rows where no player had played the game and rows with no valid information, the merged data contained a total of 260,000 pieces of data. After further data sorting and removing invalid data, including dropping games without player information (ID, rating, age, income). Moreover, the game occurred outside a zip code with $PM_{2.5}$ data. The final hourly data is 153,362, and the daily data is 33,999. We also performed linear regression on the winning rate, excluding 0 and 1 winning rates, with a dataset of 26,286.

2.3 Summary Statistics

Table 1.1 Summary statistics of binomial model on winning a chess game or not

Variable	Obs.	Mean	Max	Min	SD
Win	153362	0.5832	1	0	0.493031
Rating difference	153362	75.16407	1302	-1416	218.4571
Age	153362	52.4989	77	18	8.733846
Income	153362	5.732549	9	1	2.41083
Sex	153362	0.74091	1	0	0.438138
Hourly $PM_{2.5}$ (µg/m^3) with 20-hour lag	153362	7.469064	269.02	0	6.136622
Hourly $PM_{2.5}$ (µg/m^3)	153362	7.557385	291.2	0	7.245596
Hourly Dry Bulb Temperature (°F)	153362	63.1794	113	-22	17.59842
Hourly Precipitation (in)	153362	0.002249	0.86	0	0.016776
Hourly Relative Humidity (%)	153362	64.2071	100	3	19.39275
Hourly Sea Level Pressure (Pa)	153362	30.03541	30.91	29.07	0.179796
Hourly Wind Speed (mph)	153362	7.64959	43	0	5.040547

Note: The table above lists the means, standard deviations, minimum and maximum values, and counts of observations for all the variables used in the analysis of the binomial model for winning or losing a chess game, all data are aggregated based on hourly data.

Table 1.2 Summary statistics of linear model on winning rate in the chess game

Variable	Obs.	Mean	Max	Min	SD
Winning Rate	33999	0.55316	1	0	0.2813
Rating Difference	33999	46.6710	1209	-1589	189.001
Age	33999	54.0123	77	18	12.044
Income	33999	5.84985	9	1	2.4274
Sex	33999	0.72010	1	0	0.4487
Daily $PM_{2.5}$ (µg/m^3) with 20-hour lag	33999	7.99791	169.09	0	5.3307
Daily $PM_{2.5}$ (µg/m^3)	33999	7.9983	193.466	0	5.4017
Daily Dry Bulb Temperature (°F)	33999	59.3325	92.9285	-13.2917	17.566
Daily Precipitation (in)	33999	0.00610	1.69	0	0.0237
Daily Relative Humidity (%)	33999	68.3733	100	6.75	14.636
Daily Sea Level Pressure (Pa)	33999	30.0278	30.9759	29.135	0.1834
Daily Wind Speed (mph)	33999	6.7028	94.79167	0	3.70516

Note: The table above lists the means, standard deviations, minimum and maximum values, and counts of observations for all the variables used in the analysis of the linear model for winning rate in the chess game, all data are aggregated based on daily data.

Table 1.3 Summary statistics of linear model on winning rate (without 0 or 1) in the chess game

Variable	Obs.	Mean	Max	Min	SD
Winning Rate (without 0 or 1)	26286	0.54508	0.97826	0.04	0.1749
Rating Difference	26286	45.9038	1209	-1589	185.16
Age	26286	53.8262	77	18	11.852
Income	26286	5.87069	9	1	2.4641
Sex	26286	0.72000	1	0	0.4490
Daily $PM_{2.5}$ (µg/m^3) with 20-hour lag	26286	8.05932	169.09	0	5.3936
Daily $PM_{2.5}$ (µg/m^3)	26286	8.06166	193.466	0	5.5013
Daily Dry Bulb Temperature (°F)	26286	59.4605	92.9285	-13.292	17.504
Daily Precipitation (in)	26286	0.00623	1.69	0	0.0248
Daily Relative Humidity (%)	26286	68.4760	100	6.75	14.578
Daily Sea Level Pressure (Pa)	26286	30.0277	30.9759	29.135	0.1821
Daily Wind Speed (mph)	26286	6.65913	94.7917	0	3.6759

Note: The table above lists the means, standard deviations, minimum and maximum values, and counts of observations for all the variables used in the analysis of the linear model for winning rate in the chess game but removed all 0 and 1 winning rate in the model, all data are aggregated based on daily data.

Table 1.1 presents summary statistics based on the full hourly-level dataset comprising 153,362 observations. The dependent variable is a binary indicator of whether the player won the chess game. The average win rate is approximately 58.3%, which may initially seem surprising since one might expect it to be around 50% in a two-player game. However, this average reflects the fact that the data aggregates across many players with varying skill levels and numbers of games played. For example, stronger players tend to have consistently higher win rates, which raises the overall mean. Key player-level variables include rating difference (mean = 75.16), age (mean = 52.5 years), income level (scale of 1–9), and sex (74.1% male). Environmental conditions are measured hourly, including $PM_{2.5}$ concentrations (both current and with a 20-hour lag), dry bulb temperature, precipitation, relative humidity, sea level pressure, and wind speed. The data captures a wide variation in both meteorological conditions and air pollution exposure, providing a rich basis for modeling performance outcomes under different environments.

Table 1.2 provides descriptive statistics for a daily-level dataset containing 33,999 observations used in the linear model estimating the player’s winning rate. The mean winning rate is 55.3%, with values ranging from 0 to 1. Compared to the hourly sample, this dataset is aggregated at the day level. Key control variables include rating difference (mean = 46.67), age (mean = 54.0), income level (mean = 5.85), and sex (72.0% male). Environmental variables are averaged daily and include $PM_{2.5}$ concentrations (current and 20-hour lag), temperature, precipitation, humidity, pressure, and wind speed. This dataset enables the identification of correlations between daily ambient conditions and chess performance.

Table 1.3 reports summary statistics for the restricted sample used in the linear model analysis, excluding all observations with a winning rate of exactly 0 or 1. This results in a sample of 26,286 observations where the winning rate ranges from 0.04 to 0.978, with a mean of 54.5%. This trimming helps avoid boundary bias in linear estimation. The mean rating difference is 45.9, with player ages averaging 53.8 years and 72.0% of participants being male. Income remains centered around 5.87. Environmental variables are aggregated to the daily level, including $PM_{2.5}$ concentrations (current and lagged), temperature, precipitation, humidity, pressure, and wind speed. The cleaned sample is better suited for continuous outcome modeling of cognitive performance under varying environmental exposures.

3. Empirical Framework

3.1 Hourly Win or Not Regression Model

In terms of model selection, we start with two aspects: whether the first model is winning or not in the chess game, this model is accurate to the hour, including the hourly $PM_{2.5}$ at the time of the game, as well as the hourly weather factors. In this model, we validated two different scenarios: a binomial model with only weather and pollution factors, and personal information (age, income), as well as variables for scoring the game (rating difference). And the other model with interaction variables based on $PM_{2.5}$ and characteristics. We also added fixed effects on time for heterogeneous issues.

$$\Pr\left(\mathrm{Win}_{ijt} = 1\right) = \mathrm{logit}^{-1}\left(a_0 + a_y + a_m + a_d\right) + \sum_{k=1}^{6} \beta_k X_{jt}^{(k)} + \sum_{l=1}^{4} \gamma_l Z_{ijt}^{(l)} + \sum_{m=1}^{4} \delta_m \left(\mathrm{HPM}_{jt} \cdot Z_{ijt}^{(l)}\right) + \varepsilon_{ijt}$$

$$\Pr(\mathrm{Win}_{\mathrm{ijt}} = 1) = \mathrm{logit}^{-1}(\mathrm{a}_0 + \mathrm{a}_\mathrm{y} + \mathrm{a}_\mathrm{m} + \mathrm{a}_\mathrm{d}) + \beta_1 \mathrm{HPMlag}_{\mathrm{jt}} + \beta_2 \mathrm{HPM}_{\mathrm{jt}} + \beta_3 \mathrm{HDBT}_{\mathrm{jt}} + \beta_4 \mathrm{HP}_{\mathrm{jt}} + \beta_5 \mathrm{HRH}_{\mathrm{jt}} + \beta_6 \mathrm{HSLP}_{\mathrm{jt}} + \gamma_1 \mathrm{RATE}_{\mathrm{ijt}} + \gamma_2 \mathrm{AGE}_{\mathrm{ijt}} + \gamma_3 \mathrm{INC}_{\mathrm{ijt}} + \gamma_4 \mathrm{SEX}_{\mathrm{ijt}} + {} + \delta_1(\mathrm{HPM}_{\mathrm{jt}} \cdot \mathrm{RATE}_{\mathrm{ijt}}) + \delta_2(\mathrm{HPM}_{\mathrm{jt}} \cdot \mathrm{AGE}_{\mathrm{ijt}}) + \delta_3(\mathrm{HPM}_{\mathrm{jt}} \cdot \mathrm{INC}_{\mathrm{ijt}}) + \delta_4(\mathrm{HPM}_{\mathrm{jt}} \cdot \mathrm{SEX}_{\mathrm{ijt}}) + \varepsilon_{\mathrm{ijt}} \tag{1}$$

i: It represents individual players involved in chess games.
j: This index represents a geographic or spatial unit. In this context, it refers to a specific county nationwide where the chess games are played.
t: It denotes the temporal dimension of the analysis. Here, it represents the exact time when the chess games are played.
$X_{jt}^{(k)}$: is a set of environmental variables for location j at time t, where k = 1,…,6. These include lagged $PM_{2.5}$ ($HPMlag_{jt}$), current $PM_{2.5}$ (HPM_{jt}), dry bulb temperature ($HDBT_{jt}$), precipitation (HP_{jt}), relative humidty (HRH_{jt}), and sea level pressure ($HSLP_{jt}$).
$Z_{jt}^{(l)}$: denotes player-specific characteristics for player i at location j and time t, where l = 1,…,4. These include rating difference ($RATE_{ijt}$), age (AGE_{ijt}), income level (INC_{ijt}), and sex (SEX_{ijt}).
The interaction terms ($HPM_{jt} \cdot Z_{ijt}^{(l)}$): capture how the effect of $PM_{2.5}$ varies depending on individual player traits.

3.2 Daily Winning Rate Regression Model

The second model is a study on how daily weather conditions and pollution affect the winning percentage of chess games. The winning rate for the same day was obtained by looking at the win/loss results for that day and then removing the duplicates. For the model selection, we chose a basic linear model, and again, we divided the model into two groups: the first group examined the cases including 0 and 1 winnings, and the second group excluded the cases with 0 and 1 winnings. We also use the Beta regression model for better-suited proportion outcomes like winning rates that range between 0 and 1(excluding 0 and 1 rates). In all the models above, we studied the cases of weather conditions, pollution, and personal information with fixed effects on time and locations. Similarly, we add interaction variables between $PM_{2.5}$ and characteristics to find out their interactive effects on the output results.

$$\mathrm{WIN}_{\mathrm{ijt}} = \mathrm{a}_0 + \mathrm{a}_\mathrm{y} + \mathrm{a}_\mathrm{m} + \mathrm{a}_\mathrm{d} + \beta_1 \mathrm{DPMlag}_{\mathrm{jt}} + \beta_2 \mathrm{DPM}_{\mathrm{jt}} + \beta_3 \mathrm{DDBT}_{\mathrm{jt}} + \beta_4 \mathrm{DP}_{\mathrm{jt}} + \beta_5 \mathrm{DRH}_{\mathrm{jt}} + \beta_6 \mathrm{DSLP}_{\mathrm{jt}} + \gamma_1 \mathrm{RATE}_{\mathrm{ijt}} + \gamma_2 \mathrm{AGE}_{\mathrm{ijt}} + \gamma_3 \mathrm{INC}_{\mathrm{ijt}} + \gamma_4 \mathrm{SEX}_{\mathrm{ijt}} + \delta_1(\mathrm{DPM}_{\mathrm{jt}} \cdot \mathrm{RATE}_{\mathrm{ijt}}) + \delta_2(\mathrm{DPM}_{\mathrm{jt}} \cdot \mathrm{AGE}_{\mathrm{ijt}}) + \delta_3(\mathrm{DPM}_{\mathrm{jt}} \cdot \mathrm{INC}_{\mathrm{ijt}}) + \delta_4(\mathrm{DPM}_{\mathrm{jt}} \cdot \mathrm{SEX}_{\mathrm{ijt}}) + \varepsilon_{\mathrm{ijt}} \tag{2}$$

i: It represents individual players involved in chess games.
j: This index represents a geographic or spatial unit. In this context, it refers to a specific county nationwide where the chess games are played.
t: It denotes the temporal dimension of the analysis. Here, it represents the exact date when the chess games are played.

Variable Definitions:

Win_{ijt} :Win or not by player i in county j at time t,

a_y,a_m,a_d : Year,month,day fixed effects,

$X_{jt}^{(k)}$: Environmental variables at time t and location j,including:

$HPMlag_{jt}$: Hourly 20-hour- lagged $PM_{2.5}$ concentration,

$DPMlag_{jt}$: Daily 20-hour- lagged $PM_{2.5}$ concentration,

HPM_{jt} : Hourly Current $PM_{2.5}$ concentration,

DPM_{jt} :Daily Current $PM_{2.5}$ concentration,

$HDBT_{jt}$, HP_{jt}, HRH_{jt}, $HSLP_{jt}$: Hourly weather controls(Dry bulb temperature, precipitation, relative humidity, sea level pressure)

$DDBT_{jt}$, DPjt, DRH_{jt}, $DSLP_{jt}$: Daily weather controls(Dry bulb temperature, precipitation, relative humidity, sea level pressure)

$Z_{ijt}^{(l)}$: Players' characteristic controls:

$RATE_{ijt}$: Rating difference (Own rating minus Opponent rating)

AGE_{ijt}, INC_{ijt}, SEX_{ijt} : Demographic covariates (Age, income level, gender)

ε_{ijt} : Error term

These two models estimate the probability of winning a chess game based on $PM_{2.5}$ levels, weather conditions, and player characteristics. It directly tests the hypothesis that higher $PM_{2.5}$ impairs cognitive performance by examining its effect on the probability of winning. Weather variables such as temperature and humidity are included to assess their influence on cognitive outcomes. Interaction terms between $PM_{2.5}$ and age, gender, and income test whether the impact of pollution varies across different groups. Significant results in these variables would support the core hypotheses about pollution and cognition.

4. Results

From the analysis of the data collected, we can conclude the following.

4.1 Hourly Regression Model on How Win or Not Frequency Being Affected

Table 1.4 The Effect of $PM_{2.5}$ and Weather Conditions on Chess Game Outcomes

Dependent variable: Probability of winning (=1 if player wins, 0 otherwise)		
Variables	Hourly Win or Not Models	
	FE Logit	FE Logit with Interaction
Intercept	-1.53300	-1.79700*
	(1.05400)	(1.05600)
Air Pollution		
$PM_{2.5}$ (current) ($\mu g/m^3$)	-0.00319***	-0.00324***
	(0.00118)	(0.00118)
$PM_{2.5}$ (20-hour lag) ($\mu g/m^3$)	0.00331**	0.00346***
	(0.00157)	(0.00869)
Weather Controls		
Temperature (°F)	0.00217***	0.00227***
	(0.00046)	(0.00046)
Precipitation (in)	0.19690	0.20270
	(0.35450)	(0.35460)
Relative Humidity (%)	0.00145***	0.00140***
	(0.03482)	(0.00032)
Sea Level Pressure (Pa)	0.05190	0.05306
	(0.03482)	(0.03483)
Player Characteristics		
Rating Difference	0.00570***	0.00583***
	(0.00004)	(0.00006)
Age	0.00227***	0.00091
	(0.00066)	(0.00113)
Income	-0.01456***	-0.00848*
	(0.00259)	(0.00503)
Sex (Female = 0)	-0.07619***	-0.03955
	(0.01382)	(0.02540)
Interaction Variables		
$PM_{2.5}$ × Rating Difference	—	-0.00002**

		(0.00001)
$PM_{2.5}$ × Age	—	-0.00042***
		(0.00012)
$PM_{2.5}$ × Income	—	-0.00087
		(0.00054)
$PM_{2.5}$ × Sex	—	-0.00549**
		(0.00264)
Model Fit		
Log Likelihood	-87,124.05	-87,113.22
Deviance	174,248.09	174,226.45
Observations	153,362	153,362

***p<0.001, **p<0.01, *p<0.05

Notes: In Table 1.4 above, I built 2 models on how hourly weather conditions and pollution impact the number of winnings in the chess game. In the FE Logit model, I added time fixed effects. The FE Logit with Interaction model has interaction variables between $PM_{2.5}$ and characteristics variables including rating difference, age, income.

Table 1.4 presents the results from fixed-effects logit models estimating the probability of winning a chess game as a function of air pollution exposure ($PM_{2.5}$), weather conditions, and individual player characteristics. The primary specification includes interaction terms to examine potential heterogeneity across demographic groups.

The current concentration of $PM_{2.5}$ remains a significant predictor of performance: a one-unit increase in $PM_{2.5}$ ($\mu g/m^3$) is associated with a 0.00324 decrease in the log-odds of winning a game ($p < 0.001$), holding other variables constant. This negative effect persists even after accounting for interaction terms, reinforcing the hypothesis that short-term exposure to air pollution may impair cognitive functioning during gameplay. Interestingly, the 20-hour lag of $PM_{2.5}$ shows the opposite effect. A one-unit increase in $PM_{2.5}$ concentration 20 hours prior is associated with a 0.00346 increase in the log-odds of winning ($p < 0.001$), suggesting that the timing of pollution exposure matters—while immediate exposure may hinder performance, prior exposure may be less harmful or even linked with compensatory mechanisms such as rest, adaptation, or selection bias (e.g., less fatigued players playing at those times).

Among the weather variables, temperature exhibits a consistent and significant positive association with the likelihood of winning. A one-degree Fahrenheit increase is associated with a 0.00227 increase in log odds ($p < 0.001$), potentially due to enhanced comfort or alertness under moderate temperatures. Relative humidity also shows a small but significant positive impact (coefficient = 0.00140, $p < 0.001$), possibly reflecting improved cognitive conditions under more humid environments. In contrast, precipitation and sea-level pressure remain statistically insignificant.

Individual attributes show expected patterns. The rating difference remains the strongest predictor of winning probability, with each additional point in a player's favor increasing the log odds of winning by 0.00583 ($p < 0.001$), confirming the robustness of skill differentials. Age, positively associated with winning in the baseline model (coefficient:0.00227, $p < 0.001$), loses statistical significance after interactions are introduced. Income exhibits a significant negative association (coefficient = -0.00848, $p < 0.05$), implying that players with higher income levels may perform slightly worse. This counterintuitive finding might reflect reduced playing time, higher opportunity costs, or stress from external commitments. Gender also plays a role: female players are more likely to win in the baseline model, but this effect becomes statistically insignificant in the interaction model, suggesting that gender differences may be partly explained by confounding variables.

The interaction terms provide deeper insight into how $PM_{2.5}$ influences players differently:

$PM_{2.5}$ × Age is significantly negative (coefficient = -0.00042, $p < 0.001$), indicating that older players are more vulnerable to pollution's cognitive effects. $PM_{2.5}$ × Rating Difference also shows a small but significant negative effect (-0.00002, $p < 0.01$), suggesting that even skilled players may suffer slightly more under polluted conditions, possibly due to the higher cognitive demands of high-level gameplay. $PM_{2.5}$ × Sex is negative and significant (-0.00549, $p < 0.01$), implying that male players experience more performance degradation under pollution. $PM_{2.5}$ × Income, while negative, is not statistically significant, indicating that income does not clearly buffer or amplify pollution's effect on performance in this model.

4.2 The Effect of Air Quality and The Difference of Individual Subjects on The Winning Rate

Table 1.5 The Effect of $PM_{2.5}$ and Weather Conditions on Chess Winning Rates

Dependent variable: Winning rate (continuous, 0–1 scale)

Variables	**Daily Winning Rate Model**	
	FE OLS	**FE OLS with Interaction**
Intercept	0.49520*	0.52030**
	(0.25470)	(0.25520)
Air Pollution		
$PM_{2.5}$ (current) (µg/m³)	-0.00049	-0.00055
	(0.00071)	(0.00071)
$PM_{2.5}$ (20-hour lag) (µg/m³)	0.00008	0.00076
	(0.00073)	(0.00189)
Weather Controls		
Temperature (°F)	0.00033***	0.00030**
	(0.00012)	(0.00012)
Precipitation (in)	0.11460*	0.10840*
	(0.06327)	(0.06329)
Relative Humidity (%)	-0.00012	-0.0001
	(0.00011)	(0.00011)
Sea Level Pressure (Pa)	0.00226	0.00135
	(0.00842)	(0.00843)
Player Characteristics		
Rating Difference	0.00048***	0.00047***
	(0.00001)	(0.00001)
Age	0.00030**	0.00000
	(0.00013)	(0.00022)
Income	-0.00739***	-0.00580***
	(0.00061)	(0.00117)
Sex (Female = 0)	-0.03545***	-0.01936***
	(0.00330)	(0.00604)

Interaction Variables		
$PM_{2.5}$ × Rating Difference	—	0.00000
		(0.00000)
$PM_{2.5}$ × Age	—	0.00003
		(0.00002)
$PM_{2.5}$ × Income	—	-0.00018
		(0.00013)
$PM_{2.5}$ × Sex	—	-0.00189***
		(0.00063)
Model Fit		
R^2	0.11905	0.11929
Adj. R^2	0.11762	0.11776
Observations	33,999	33,999
***p<0.001, **p<0.01, *p<0.05		

Note: In Table 1.5 above, I built 2 models on how daily weather conditions and pollution impact winning rate in the chess game. In the FE OLS model, I added time fixed effects. The FE OLS with Interaction model has interaction variables between air pollutants and characteristics besides the basic weather, pollution, and personal information variables.

Table 1.5 presents the results from fixed-effects OLS models estimating the daily winning rate as a function of air pollution ($PM_{2.5}$), weather conditions, and individual characteristics. The current $PM_{2.5}$ concentration and its 20-hour lag do not show statistically significant effects in either the baseline or interaction model. Specifically, the coefficients for current $PM_{2.5}$ are –0.00049 and –0.00055 across the two models, indicating that holding other variables constant, a one-unit increase in $PM_{2.5}$ will lead to an average decrease of 0.00050 in the winning rate, and those for lagged $PM_{2.5}$ remain close to zero. This may be due to aggregation effects—daily averaging may smooth out short-term pollution shocks that influence game-by-game performance.

Among weather variables, temperature exhibits a statistically significant positive relationship with winning rate across both models (coefficient = 0.00033, 0.00030, $p < 0.001$ or $p < 0.01$ in baseline OLS and OLS with interactions), indicating that holding other variables constant, one unit increase in temperature will lead to average 0.00031 increase in winning rate, suggesting that players may perform slightly better under warmer conditions, possibly due to improved physical comfort. Precipitation also shows a positive and statistically significant association (coefficient = 0.1146 and 0.1084, $p < 0.05$ in both models), indicating that holding other variables constant, a one-unit increase in precipitation leads to an average increase of 0.1115 in the winning rate. In contrast, relative humidity and sea-level pressure do not show statistically significant associations with performance, differing from results in the hourly model.

In terms of player characteristics, rating difference remains the most robust predictor of winning rate (coefficient = 0.00048 and 0.00047, $p < 0.001$ for both models), indicating that holding other variables constant, one unit increase in rating difference between players and their opponents will lead to average 0.00047 increase in winning rate, consistent with the notion that skill advantage translates directly into higher success probability. Age has a small positive effect in the baseline model (0.00030, $p < 0.01$). Income shows a statistically significant negative effect across both models (coefficient = –0.00739 and –0.00580, $p < 0.001$ in both models), indicating that holding other variables constant, one unit increase in income level will lead to an average 0.0066 decrease in winning rate, again echoing earlier results and potentially reflecting opportunity cost or time allocation differences—higher-income players may spend less time training or competing seriously. Sex (female = 0) is also significantly associated with lower winning rates in both models (coefficient = –0.03545 and –0.01936, $p < 0.001$), suggesting that female players in the sample tend to outperform their male counterparts after controlling for other factors.

Turning to interaction effects, we explore whether individual-level traits moderate the cognitive effect of $PM_{2.5}$. Among all interaction terms, only the interaction between $PM_{2.5}$ and sex is statistically significant (coefficient = –0.00189, $p < 0.001$). This result implies that male players suffer greater performance losses in response to pollution, as their winning rates decline more steeply when exposed to higher $PM_{2.5}$ levels. The underlying reasons may be biological or behavioral—for instance, male players might be more reactive to stressors. The remaining interaction terms— $PM_{2.5}$ × rating difference, $PM_{2.5}$ × age, and $PM_{2.5}$ × income—are statistically insignificant.

Table 1.6 $PM_{2.5}$ and Weather Effects on Chess Winning Rates (Excluding 0% and 100% Outcomes)

Dependent variable: Winning rate (continuous, excluding 0 and 1)		
Variables	**Daily Winning Rate Model Excluding 0% and 100% Rates**	
	FE OLS	**FE OLS with Interactions**
Intercept	0.58990***	0.58940***
	(0.17870)	(0.17910)
Air Pollution		
$PM_{2.5}$ (current) (μg/m³)	-0.00022	-0.00043
	(0.00048)	(0.00048)
$PM_{2.5}$ (20-hour lag) (μg/m³)	-0.00022	-0.00200
	(0.00048)	(0.00135)
Weather Controls		
Temperature (°F)	-0.00049***	-0.00050***
	(0.00009)	(0.00009)
Precipitation (in)	0.11720***	0.11490***
	(0.04205)	(0.04205)
Relative Humidity (%)	0.00002	0.00002
	(0.00007)	(0.00007)
Sea Level Pressure (Pa)	-0.00121	-0.00163
	(0.00591)	(0.00592)
Player Characteristics		
Rating Difference	0.00032***	0.00029***
	(0.00001)	(0.00001)
Age	0.00070***	0.00071***
	(0.00009)	(0.00016)
Income	-0.00514***	-0.00376***
	(0.00043)	(0.00081)
Sex (Female = 0)	-0.01904***	-0.01024**
	(0.00231)	(0.0043)

Interaction Variables		
$PM_{2.5}$ × Rating Difference	—	0.00000***
		(0.00000)
$PM_{2.5}$ × Age	—	-0.00001
		(0.00002)
$PM_{2.5}$ × Income	—	-0.00017*
		(0.00009)
$PM_{2.5}$ × Sex	—	-0.00107**
		(0.00045)
Model Fit		
R^2	0.13721	0.13764
Adj. R^2	0.13639	0.13669
Observations	26,286	26,286

***p<0.001, **p<0.01, *p<0.05

Note: In Table 1.6 above, I built 2 models on how daily weather conditions and pollution impact winning rate in the chess game. In the FE OLS model, I added time fixed effects. The FE OLS with Interaction model has interaction variables between air pollutants and characteristics besides the basic weather, pollution, and personal information variables. 0% and 100% winning rate are excluded.

Table 1.6 presents the fixed-effects OLS model results for the daily adjusted winning rate, excluding extreme values of 0% and 100%. This approach filters out extreme cases to better reflect nuanced variation in performance.

In the baseline OLS model and the OLS with interactions model, $PM_{2.5}$ is negatively correlated with the winning rate (coefficient: -0.00022 and -0.00043), indicating that holding other variables constant, one unit increase in $PM_{2.5}$ leads to an average 0.00033 decrease in winning rate, while $PM_{2.5}$ with 20-hour lag is also negatively correlated with the winning rate (coefficient: -0.00022 and -0.002).

Among weather variables, temperature displays a consistently significant negative association with performance (-0.00049 and –0.00050, $p < 0.001$ for both models), indicating that holding other variables constant, a one unit increase in temperature leads to an average 0.0005 decrease in the winning rate. Precipitation shows a positive and significant effect (coefficient = 0.11720 and 0.11490, $p < 0.001$), indicating that holding other variables constant, one unit increase in precipitation leads to an average 0.116 increase in winning rate, potentially indicating selection effects (e.g., only highly committed players compete on rainy days) or reduced distractions during indoor matches. Humidity and pressure again exhibit no statistically significant associations.

As expected, rating difference is a robust and highly significant predictor of winning rate (coefficient = 0.00032 and 0.00029, $p < 0.001$ for both models), reaffirming its central role as a skill metric. Age still shows a significant and positive relationship with performance ($\approx$ 0.00071, $p < 0.001$), indicating that holding other variables constant, a one-year increase in age leads to an average 0.00071 increase in the winning rate. Income continues to exert a significant negative effect on winning rate (–0.00514 and 0.00376, $p < 0.001$), consistent with earlier findings that higher-income individuals may spend less time practicing or playing competitively. Female players outperform their male counterparts on average (sex coefficient = –0.01024, $p < 0.01$), even after controlling for full interaction terms.

Turning to interaction effects, $PM_{2.5}$ × Rating Difference is statistically significant (coefficient = 0.00000, $p < 0.001$), though its magnitude is nearly zero. This suggests that while skill level may slightly amplify the pollution effect, the moderation is extremely small in size, even if statistically detectable. $PM_{2.5}$ × Age remains statistically insignificant, indicating no meaningful difference in pollution sensitivity across age groups in this adjusted sample. $PM_{2.5}$ × Income shows a small but statistically significant negative interaction (–0.00017, $p < 0.05$), suggesting that the adverse effects of pollution are slightly increased among higher-income individuals. $PM_{2.5}$ × Sex again reveals a statistically significant negative interaction (–0.00107, $p < 0.01$), confirming that male players' performance is more adversely affected by pollution relative to female players. This aligns with both the unadjusted and daily models, highlighting a robust gender-based difference in sensitivity to environmental stressors.

Table 1.7 Beta Regression Analysis of $PM_{2.5}$ Effects on Chess Winning Rates (Excluding 0% and 100% Outcomes)

Dependent variable: Winning rate (continuous, bounded between 0 and 1, excluding endpoints)		
Variables	**Daily Winning Rate Model Excluding 0% and 100% Rate**	
	FE Beta	**FE Beta with Interaction**
Intercept	0.20160	0.20600
	(0.72730)	(0.72880)
Air Pollution		
$PM_{2.5}$ (current) (μg/m^3)	-0.00010*	-0.00163*
	(0.00195)	(0.00197)
$PM_{2.5}$ (20-hour lag) (μg/m^3)	-0.00068	-0.00818*
	(0.00204)	(0.00550)
Weather Controls		
Temperature (°F)	-0.00196***	-0.00199***
	(0.00035)	(0.00035)
Precipitation (in)	0.44360**	0.43650**
	(0.17260)	(0.17250)
Relative Humidity (%)	0.00017	0.00015
	(0.00030)	(0.00030)
Sea Level Pressure (Pa)	-0.00028	-0.00164
	(0.02406)	(0.02408)
Player Characteristics		
Rating Difference	0.00131***	0.00121***
	(0.00002)	(0.00004)
Age	0.00281***	0.00296***
	(0.00037)	(0.00064)
Income	-0.02219***	-0.01673***
	(0.00174)	(0.00330)
Sex (Female = 0)	-0.07227***	-0.04151**
	(0.00939)	(0.01751)
Interaction Variables		
$PM_{2.5}$ × Rating Difference	—	0.00121***
		(0.00004)
$PM_{2.5}$ × Age	—	-0.00005
		(0.00006)
$PM_{2.5}$ × Income	—	-0.00069*
		(0.00036)

PM$_{2.5}$ × Sex	—	-0.00375**
		(0.00183)
Model Fit		
Pseudo R^2	0.13805	0.13840
Log Likelihood	10887.26473	10892.57877
Observations	26,286	26,286
***p<0.001, **p<0.01, *p<0.05		

Table 1.7 reports results from fixed-effects Beta regression models estimating daily winning rates (excluding endpoints 0 and 1), bound between 0 and 1. This modeling approach is ideal for fractional outcomes, accounting for both skewness and bounded nature of the dependent variable.

In the baseline OLS model and the OLS with interactions model, PM$_{2.5}$ is negatively correlated with the winning rate (coefficient: -0.00010 and -0.000163, $p<0.05$ in both models), while PM$_{2.5}$ with 20-hour lag is also negatively correlated with the winning rate (coefficient: -0.00068 and -0.00818, $p<0.05$ in interaction model). It can be observed that in the logit model, capturing hourly wins or losses, PM$_{2.5}$lag exhibits a significant positive correlation with the win rate; however, this positive correlation is weakened in the daily OLS model and even reverses to a negative correlation in the BETA model. The possible reason is that short-term lagged pollution exposure may lead to compensatory cognitive reactions or changes in player behavior, which can be captured more keenly in the logit model but weakened in the average-based model. PM$_{2.5}$ and its lagged effects are highly significant in the hourly fixed-effects logit models but lose significance in the daily OLS and BETA regression models. This discrepancy is likely due to signal dilution from temporal aggregation. While air pollution impairs cognitive performance in the short term—most visible during or shortly after exposure—daily models average over multiple games, potentially masking these acute effects. Moreover, hourly models have higher variance in the outcome variable and greater statistical power due to large sample sizes, enabling the detection of subtle real-time cognitive disruptions.

Among weather controls, temperature continues to have a significant negative effect (-0.00196 and –0.00199, $p < 0.001$ in both models), indicating that holding other variables constant, one unit increase in temperature leads to an average 0.00198 decrease in winning rate, suggesting that higher temperatures may suppress cognitive performance over a full day of play. Precipitation is again a positive and significant predictor of winning rates (coefficient = 0.44360 and 0.43650, $p < 0.01$ in both models), possibly reflecting favorable indoor play conditions or reduced distractions on rainy days. Relative humidity and sea-level pressure remain statistically insignificant.

Turning to individual-level characteristics, the rating difference remains a strong and highly significant predictor (0.00131 and 0.00121, $p < 0.001$), indicating that a one-unit increase in the rating difference between players and their opponents leads to an average 0.00126 increase in winning rate. Age still exhibits a positive and statistically significant coefficient (0.00281 and 0.00296, $p < 0.001$ in both models), suggesting that older players may show more stable or consistent daily performance. Income continues to have a substantial and significant negative effect (–0.02219 and -0.01673, $p < 0.001$), indicating that holding other variables constant, one unit increase in income level leads to an average 0.01991 decrease in winning rate, possibly reflecting reduced training time or higher opportunity cost among higher earners. Sex also shows a statistically significant negative coefficient (–0.04151, $p < 0.01$) in the interaction model, indicating that male players have a slight disadvantage in adjusted daily performance after controlling for other factors.

The interaction terms provide insight into heterogeneous sensitivity to pollution: $PM_{2.5}$ × Rating Difference is positive and statistically significant (0.00121, $p < 0.001$), suggesting that stronger players (those with larger rating advantage) are less negatively affected—or potentially even slightly benefit—from pollution exposure, perhaps due to better adaptability or experience. The sign of the interaction term between $PM_{2.5}$ and the rating difference varies across models: it is negative in the hourly fixed-effects logit model but positive in the daily beta regression model. This discrepancy likely stems from differences in model structure and the interpretation of outcomes. In the hourly model, the negative interaction suggests that air pollution attenuates the performance advantage of higher-rated players, possibly due to its impact on real-time cognitive functions. Conversely, in the daily model, the positive coefficient may indicate that, on average, stronger players are more resilient under pollution exposure, preserving a greater performance gap over weaker players. These two interpretations reflect different layers of cognitive and behavioral dynamics at different temporal resolutions. $PM_{2.5}$ × Age is not significant, implying that age does not meaningfully moderate pollution's impact on daily winning rate, contrasting with the pattern seen in hourly models. $PM_{2.5}$ × Income is negative and weakly significant (–0.00069, $p < 0.05$), indicating that higher-income individuals may be marginally more affected by air pollution. $PM_{2.5}$ × Sex is negative and statistically significant (–0.00375, $p < 0.01$), implying that male players' performance declines more under high $PM_{2.5}$ exposure compared to female players. This pattern reinforces evidence from earlier models and supports the hypothesis of gender-based differential vulnerability to environmental stress.

4.3 Dose Response Function on Daily Winning Rate Regression Model

While regression coefficients summarize the average effect of air pollution and weather variables on chess performance, they may obscure potential nonlinearities and heterogeneities across exposure levels. To address this, we compute dose-response functions based on the regression models. These functions show how the predicted probability of winning varies with changes in $PM_{2.5}$ concentrations and meteorological conditions, holding other covariates constant at their means.

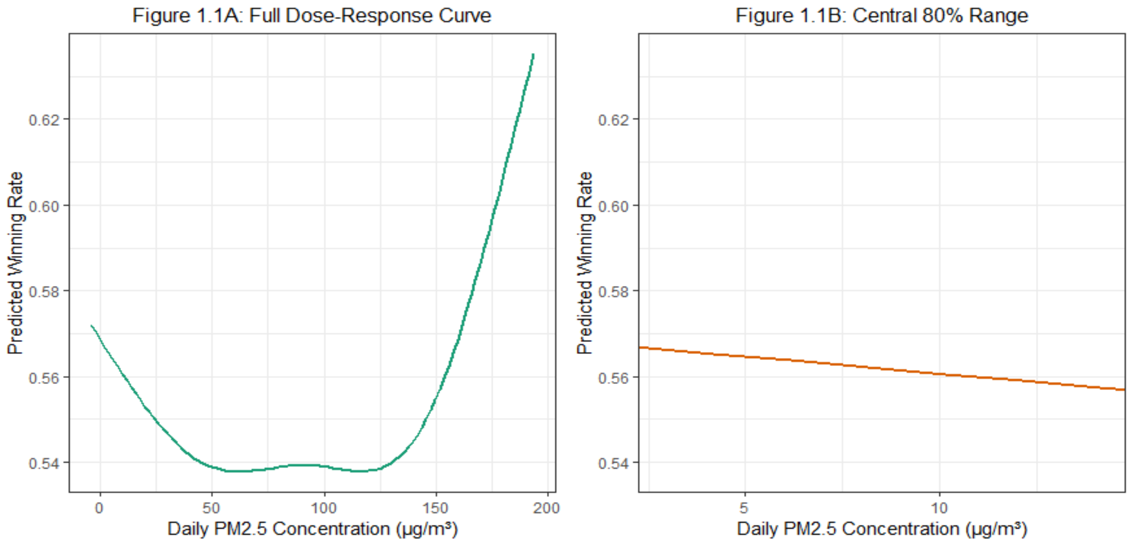

Figure 1.1 Dose-Response Curve between $PM_{2.5}$ and Winning Rate

Figure 1.1 outlines the nonlinear dose-response relationship of the winning rate over $PM_{2.5}$ concentration. On balance, the pattern is in the form of a sharp U-curve: as $PM_{2.5}$ levels increase from low to moderate levels (approximately 0 to 125 µg/m³), the estimated winning rate is declining consistently. After a certain point of approximately 140 µg/m³, the trend reverses, and the winning rate rises exponentially to a saturation point at very high levels of pollution (nearly 200 µg/m³). This pattern may be the cumulative effect of air pollution on a person's performance and behavior, as well as their intellectual capacity. At low levels, higher $PM_{2.5}$ concentrations likely decrease attention, reaction time, and mental health, leading to performance deterioration and, subsequently, lower winning rates. Conversely, performance improvement at extremely high pollution levels can be due to numerous factors. One is sample selection bias—only elite players may play in tournaments under such adverse conditions. Another is behavioral adaptation—players may become more focused and cautious when they are exposed to harsher environments, which can offset the negative impact of pollution.

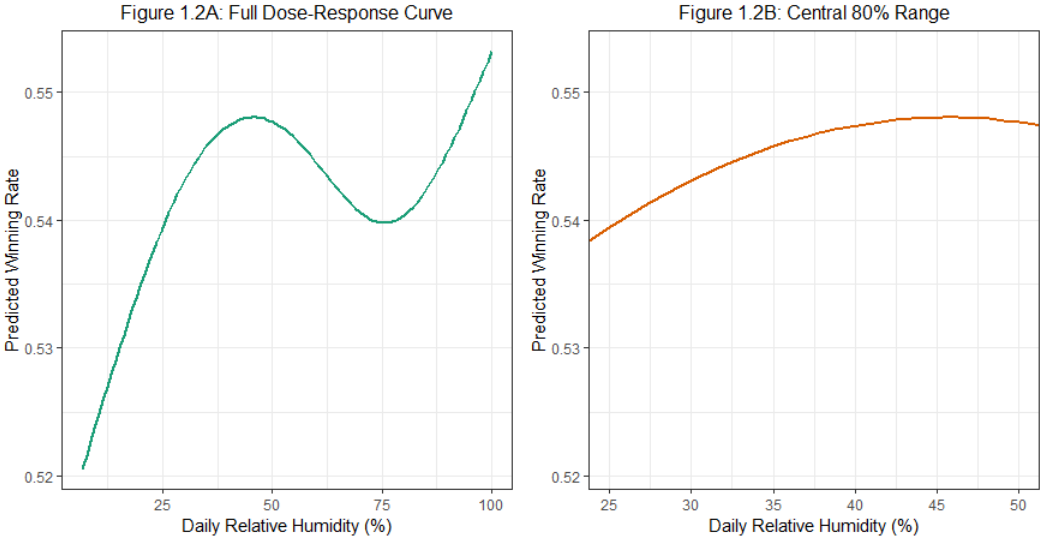

Figure 1.2 Dose-Response Curve between Relative Humidity and Winning Rate

Figure 1.2 presents the plot of the dose-response relationship between relative humidity on a daily basis and the predicted winning rate. The curve increases as relative humidity rises from low (approximately 10%) to high levels (approximately 100%), and the winning rate rises steadily. The slope is steeper at the lower range (10%–50%), with every increase in humidity being associated with significant performance improvement. After approximately 75%, the curve again becomes steeper, but the trend continues to rise. This trend suggests that increased relative humidity may have a positive relationship with overall performance outcomes. The explanation could be that extremely low humidity may lead to physical discomforts—such as dry eyes or respiratory distress—which would, in turn, decrease concentration and mental processes. With increases in humidity to more moderate and comfortable levels, these adverse effects can be negated, allowing for improved performance. Even at higher levels of humidity, while advantages are increasingly lost, there may still be resistance to warning or stress reduction, allowing improved rates of winning to be sustained. The curve ultimately becomes primarily linear and positively indicative of the relative humidity's effect on competitive performance.

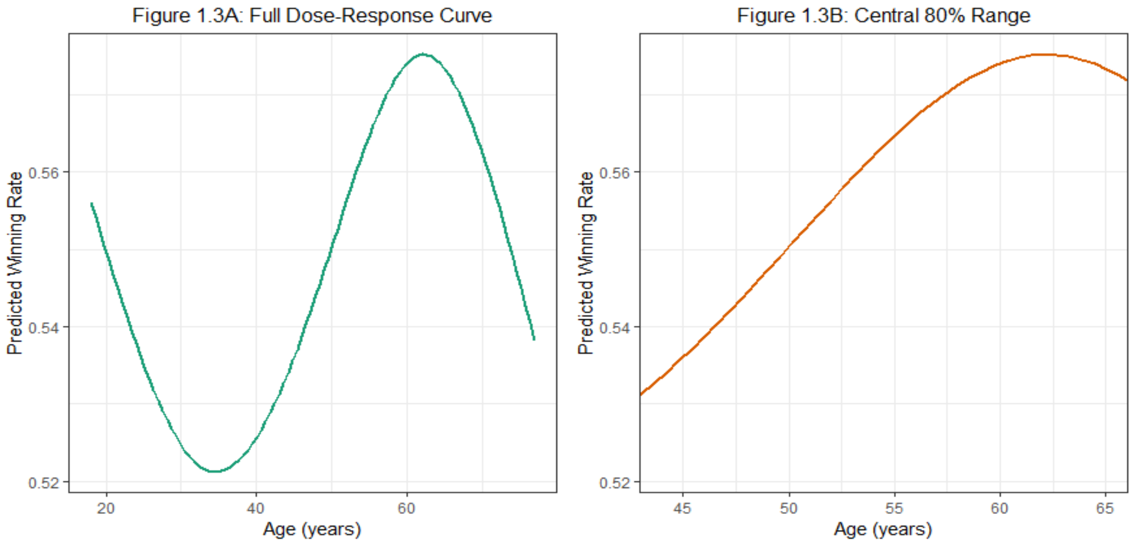

Figure 1.3 Dose-Response Curve between Age and Winning Rate

Figure 1.3 illustrates the dose-response curve of age vs. the winning rate. The relationship is not linear but rather S-shaped and then declining. Initially, as age increases from approximately 20 to 35 years, the predicted winning rate decreases slightly and reaches a local minimum at around age 35. Subsequently, the curve exhibits an increasing trend with the winning rate reaching a peak around the age of 60. Beyond age 60, the curve begins to decline. This tendency suggests that middle-aged and younger individuals may experience declining performance or a lower likelihood of winning due to various reasons, such as poor experience or competing priorities. The continuous increase to the peak age of 60 can be attributed to experience acquired over time, preplanning, or improved decision-making capacity. The minimal fall after the age of 60 can be due to reasons such as aging, a slowed reaction, or a lack of stamina.

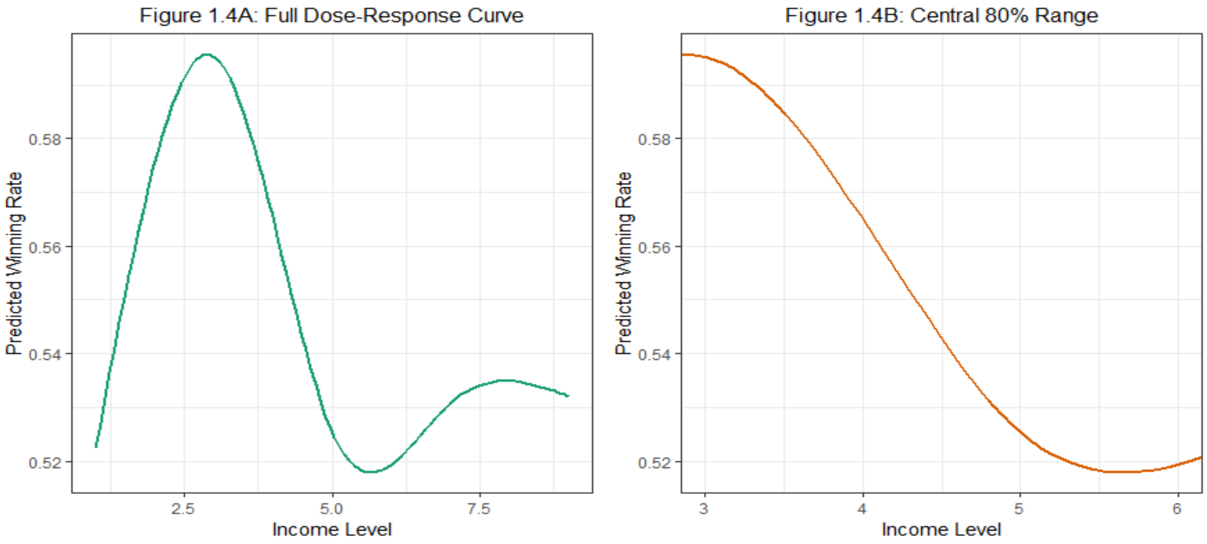

Figure 1.4 Dose-Response Curve between Income and Winning Rate

Figure 1.4 is the dose-response curve of winning rate as a function of income level. The shape of the dose-response curve shows a non-linear pattern. For the income level growing from its minimum up to about level 2.5, the projected winning rate does indeed grow steeply toward its peak. Beyond this peak, the curve shows a significant drop to a trough at an income level of 5. Beyond this trough, the curve increases again, though more gradually, showing a limited recovery in the winning rate at higher income levels. This trend suggests that income levels in moderate ranges are associated with the best-projected performance, perhaps due to equitable levels of resources, motivation, and effort. On the other hand, lower levels may limit one's capacity to access performance-enhancing resources, and higher levels might be reflective of reduced motivation or a deflection away from performance-oriented activities.

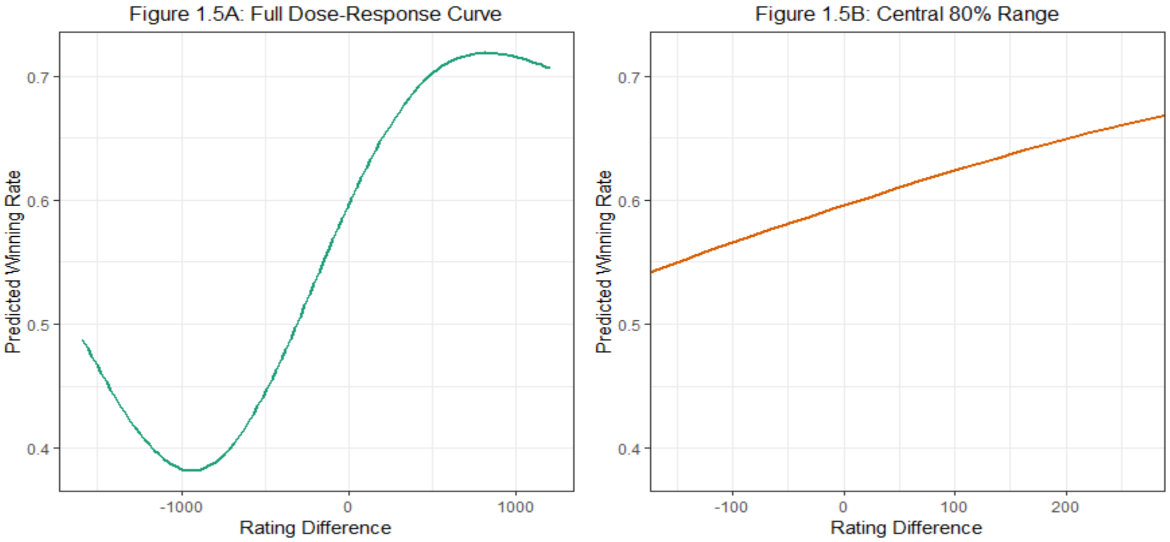

Figure 1.5 Dose-Response Curve between Rating Difference and Winning Rate

Figure 1.5 illustrates the dose-response curve for rating difference and predicted winning rate. The curve illustrates a firm and clear-cut positive relationship: as the rating differences increase from approximately -1000 to +800, the predicted winning rate increases smoothly. When the rating difference is negative (i.e., the player is rated lower than his opponent), the winning rate is small, tending towards a minimum of approximately -1000. However, as the rating difference increases, the predicted winning rate rises dramatically, reflecting a closer contest. As the rating difference becomes positive, i.e., the player has a higher rating than the other, the curve increases further, reaching around a difference in ratings of +800. The curve then declines somewhat after this, suggesting diminishing returns in the predicted winning rate at the highest differential in ratings. This trend represents expected competitive pressures: participants who are significantly behind their rivals (a high rating difference) have reduced odds of victory, whereas those with a higher rating advantage have significantly higher odds of victory. The shape of the curve illustrates a strong, non-linear positive relationship between rating advantage and the probability of victory, with performance improving steeply as the rating advantage increases.

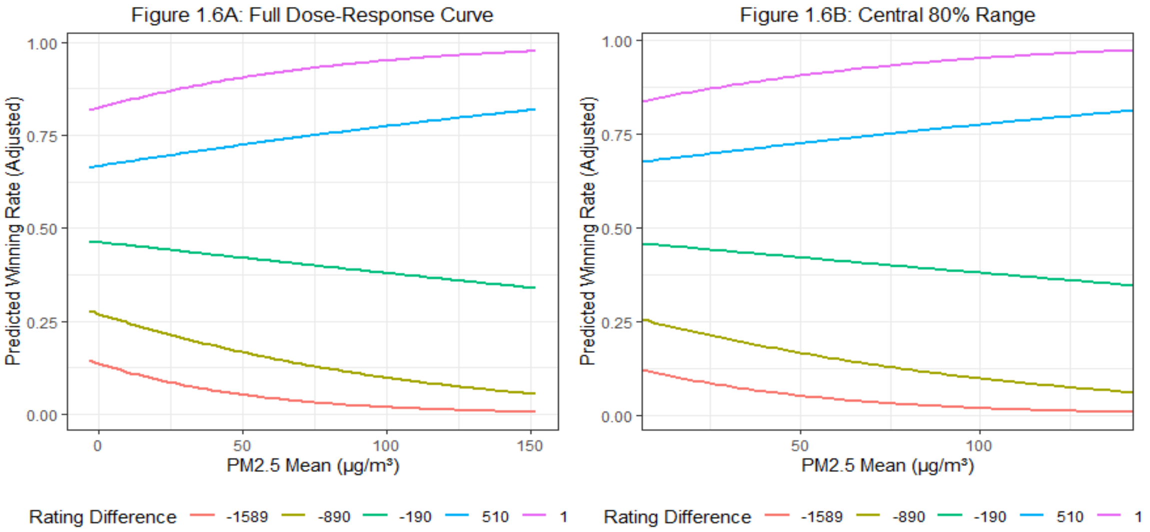

Figure 1.6 Interaction Dose-Response Curves of $PM_{2.5}$ and Rating Difference

Figure 1.6 presents the effect of the $PM_{2.5}$ level on players' estimated win rates, grouped by different levels of Rating Difference (own rating - opponent rating). Each curve represents a different rating difference, illustrating the performance of players with varying relative strengths at different levels of air pollution. The chart indicates a clear pattern: for comparatively weaker players with more minor differences in rating (e.g., -1589, -989, -190), that is, for comparatively weaker players versus stronger players, the theoretical win percentage falls steeply as $PM_{2.5}$ increases. This suggests that air pollution disproportionately disadvantages comparatively weaker players. Conversely, for higher rating difference players (say, 510 and 1209)—opponents playing against weaker players—the curves slope slightly upward, and their winning percentages rise slightly as pollution levels increase. One plausible explanation is that cognitive impairment due to air pollution does exist, but more skilled individuals may possess better cognitive reserve, mental resilience, or adaptability to stress. Therefore, their performance is less prone to deviation under stressful environments, while less capable players are more susceptible to larger performance declines. This differential sensitivity further widens the existing performance gap and virtually magnifies the advantage of the stronger players under adverse conditions. Therefore, the more sloping curves for greater rating differences do not imply that pollution helps better players, but rather illustrate how pollution increases competitive disparities in a way that allows great players to win more substantially when their opponents are less capable.

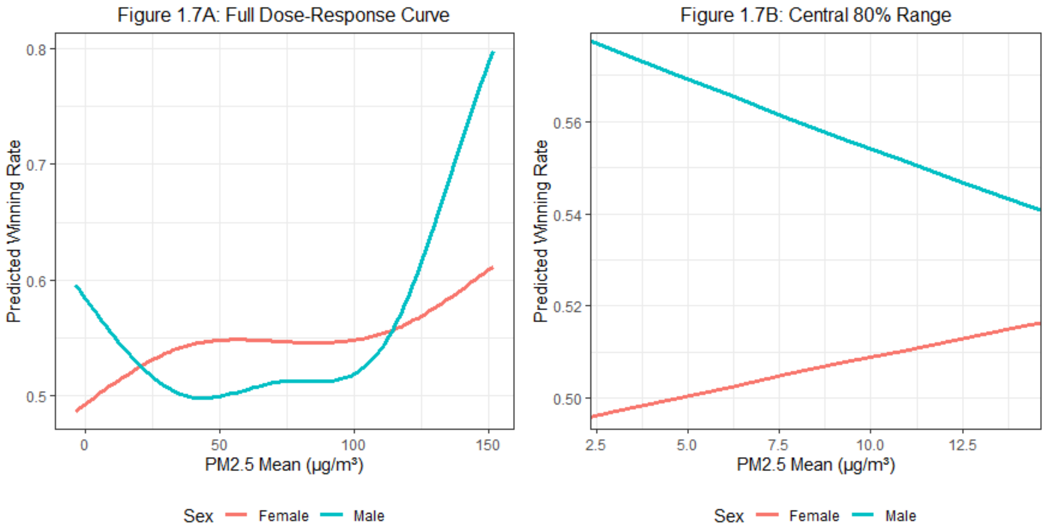

Figure 1.7 Interaction Dose-Response Curves of $PM_{2.5}$ and Sex

Figure 1.7 illustrates the interaction effect of $PM_{2.5}$ concentration and sex on the predicted win rate of players. The graph depicts two different curves, each representing a different sex group (Female = 0, Male = 1), illustrating how the win rate increases with higher air pollution. For female players, the red curve is relatively horizontal, with a very gradual upward trend as $PM_{2.5}$ concentration increases, indicating that their performance is uniform mainly across all degrees of pollution. The male curve (cyan) is U-shaped: with decreasing win rate predicted at low-to-moderate $PM_{2.5}$ concentrations, reaching around 75–100 µg/m³. However, when $PM_{2.5}$ levels are high, the curve rises rapidly, reflecting a recovery or even an enhancement of the predicted win rate under severe pollution conditions. Mechanistically, this pattern may reflect the greater sensitivity of males to moderate environmental stress, with steeper performance declines in such conditions. However, during extreme conditions, possible adaptive or compensatory processes may become predominant, allowing some male players to maintain or even improve performance. Female players would most probably exhibit greater consistency or strength with fewer variations in performance in response to $PM_{2.5}$ exposure.

5. Discussion

5.1 The Hidden Economic Toll of Air Pollution and Cognitive Productivity

The economic consequences of air pollution extend far beyond the cost of health and reduced life expectancy. This research contributes to the growing body of evidence that air pollution also impairs cognitive ability, which in turn affects the quality of decisions made and economic efficiency. Whereas traditional environmental economics has primarily focused on productivity and the health effects of pollution on manual laborers (e.g., construction or farm workers), our findings highlight that demanding cognitive work is also vulnerable to environmental degradation.

Underpinning the greater knowledge are novel investigations. For instance, Heyes et al. (2016) found that poor air quality increased risk aversion and forecasting bias among financial analysts. In the same context, Chang et al. (2019) demonstrated that call center employees experienced lower productivity on smoggy days, even when working indoors. These results suggest that pollution affects decision-making not only physically but also cognitively and emotionally, compromising performance in areas that require mental sharpness, such as finance, law, education, and management.

By demonstrating that even chess players, engrossed in a pure exercise of mental ability, are affected by decreased performance with higher levels of pollution, this article generates empirical micro-level data on the reductionist economic effect of air pollution. These intellectual expenditures, although less visible compared to hospitalizations, are likely to be widespread and persistent, ultimately decreasing levels of efficiency among the workforce, strategic reasoning, and human capital formation in the long term.

Therefore, policymakers must expand the boundaries of cost-benefit analysis of air pollution control to include not only health and mortality impacts but also the cognitive and economic expenses of exposure to pollution. That would more accurately reflect the actual social cost of air pollution and support more assertive investment in clean air policies that protect public health and economic productivity equally.

5.2 Limitation

This chapter has several limitations. First, we estimated individual exposure to air pollution by city-level ambient PM2.5 measures. This is not necessarily a measure of actual inhalation rates among indoor game players. Second, while chess performance is a good proxy for cognitive ability, it is also influenced by psychological, strategic, and situational factors that are independent of cognition. These can add noise to the analysis. Third, the analysis assumes a linear relationship between pollution and performance, which may overlook threshold or non-linear effects. Fourth, the lack of individual demographic and health data limits a richer analysis of differential effects across groups of players. Fifth, emotional responses to pollution, such as stress or fatigue, may also affect performance. Such channels are not explored in the currently available dataset.

6. Conclusion

The paper is divided into two parts. The first part examines the impact of hourly air pollution ($PM_{2.5}$) and meteorological factors (dry bulb temperature, rainfall, sea level pressure, and players' characteristics on the number of rounds won by players. The second part is addressed to the impact of daily $PM_{2.5}$ and meteorological factors, and players' characteristics on the winning rate of the players. The rationale behind different frequencies of data collection stems from the varying impacts that such conditions may have on playing chess. Weather conditions hourly can directly or indirectly impact a player's skills, strategies, or levels of concentration, thereby affecting their performance in a given game. Highly adverse weather conditions may make some strategies difficult to use, thereby reducing the rate of winning. On the other hand, day-to-day weather effects may be a better representation of general performance, as they can shape the relative chance of winning by impacting all players to a certain degree.

Analysis of round wins reveals that $PM_{2.5}$, both current and 20-hour lagged, has negative impacts on chess performance. With an increase in this variable, the log odd of games won fall. Conversely, dry bulb temperature, rain, relative humidity, and sea level pressure are positively correlated with the number of games won. For the personal characteristics of the players, age and the difference in rating between players and their opponents enhance the number of games won, where income is negatively correlated. Female players win more games than male players. These findings suggest that psychological conditions are significantly influenced by weather and air quality, which in turn influence performance. They suggest that environmental control of temperature, humidity, and air pollution can potentially enhance mental health and promote more effective work. This has implications for economic growth as improved mental health can potentially translate into greater productivity and performance in a wide range of activities, not just chess.

The winning rate research reveals the same result as the hourly win or not logit model. $PM_{2.5}$ and income indicate a negative effect on daily winning rate in online chess games, while precipitation, relative humidity, and age all indicate a positive correlation with the winning rate. Females tend to play better than males on average. The difference lies in $PM_{2.5}$ (20-hour lag) and temperature, which have different signs for the two models. The inverse relationship between $PM_{2.5}$ and temperature for both hourly and daily models suggests short-term vs. long-term impacts: at the hourly level, slight pollution or greater temperature can temporarily improve alertness or activity, whereas at the daily level, long-term exposure yields cognitive deterioration and reduced performance.

In general, the interaction term between $PM_{2.5}$ and players' personal information provides some trend of impact. Under the hourly (short-term) model, the $PM_{2.5}$ and difference in strength interaction term is significantly negative, indicating the greater the difference in strength, the more apparent the negative effect of air pollution on the win rate; the $PM_{2.5}$ and age interaction term is significantly negative, indicating older players are more susceptible to short-term pollution exposure and their win rate falls even more significantly. On the other hand, under the long-term (daily) model, both the interaction terms between rating difference and $PM_{2.5}$ are positive and statistically significant, which suggests players with wider strength differences are relatively more able to sustain or even improve their performance in the long run. At the same time, the interaction term of $PM_{2.5}$ and sex is statistically negative, indicating that the long-term pollution effect exacerbates the difference in performance between male and female players, with men being more affected. Gender characteristics have long-term effects.

The study reveals the critical link between air quality, weather, and performance in chess. These findings indicate that environmental factors need to be taken into consideration when researching performance. Future research could explore the control of such variables to maximize mental well-being and efficiency. This study can be applied to other fields where cognitive performance and decision-making are critical.

The implications of this work are far-reaching. By making known and mitigating the detrimental effects of air pollution and adverse weather conditions, one can maximize not only personal performance in tasks but also overall mental well-being and economic productivity. These findings lay the groundwork for new developments in experimentation, which aim to optimize environmental conditions for improved mental health and enhanced performance throughout life and work.

Chapter 2 Assessing the Impact of $PM_{2.5}$, and SO_2 on Respiratory Health and Socioeconomic Factors

1. Introduction

Airborne particulate matter with a diameter of 2.5 micrometers, commonly referred to as $PM_{2.5}$, is a significant concern for both public and environmental health. Due to its fine size, $PM_{2.5}$ will reach deep into the lungs and enter the bloodstream, resulting in serious health effects. This chapter explains the sources, health impacts, and control measures associated with $PM_{2.5}$ and sulfur dioxide (SO_2) in relation to their effects on respiratory and cardiovascular diseases, as well as the broader socioeconomic context. $PM_{2.5}$ is primarily caused by anthropogenic sources like industrial production, transportation, and the combustion of fossil fuels. Wildfires and volcanic eruptions are also natural events that contribute to atmospheric $PM_{2.5}$ concentration. These particles are a mixture of organic and inorganic substances, including sulfates, nitrates, ammonia, sodium chloride, black carbon, and mineral dust. $PM_{2.5}$ particles have a fine size that allows them to remain in the air for a long time and travel long distances from the point of origin. Their minimal diameter also facilitates easy entry into the respiratory system and alveoli, where gas exchange occurs. Upon entering the circulatory system, $PM_{2.5}$ can induce systemic inflammation and oxidative stress, processes associated with various health conditions. Extensive epidemiological evidence has identified a high correlation between exposure to $PM_{2.5}$ and adverse health outcomes. Short-term epidemics or acute exposures to elevated concentrations of $PM_{2.5}$ can exacerbate respiratory illnesses such as asthma and bronchitis, leading to increased hospitalization and emergency room utilization (Bhattarai, et al., 2024). Prolonged exposure causes chronic respiratory diseases, cardiovascular illness, and lung cancer. The mechanisms through which $PM_{2.5}$ exerts its harmful effects involve the initiation of inflammation and oxidative stress, which can lead to damage to airway tissues and cells. $PM_{2.5}$ can also carry harmful compounds like heavy metals and organic chemicals that can increase toxicity. Vulnerable populations like children, the aged, and individuals with pre-existing health conditions are most susceptible to the harmful effects of $PM_{2.5}$.

Sulfur dioxide (SO_2) is also a significant air pollutant, primarily originating from the combustion of fossil fuels, such as coal and oil, and industrial processes. SO_2 is a colorless, irritating gas that is very soluble in water and produces sulfurous acid upon dissolution. SO_2 has been reported to produce a range of respiratory ailments, including cough, sore throat, and labored breathing. SO_2 exposure is particularly perilous for individuals who already suffer from respiratory illnesses, such as asthma and bronchitis. Exposure to SO_2 over long periods was shown to raise the risk of chronic respiratory illness and cardiovascular disease. The adverse health effects of SO_2 are primarily due to the fact that it irritates the respiratory mucous membranes. SO_2 is also likely to react with other constituents of the air to produce fine particulate matter ($PM_{2.5}$), adding to its health effects.

1.1 Research Question and Hypotheses

This chapter addresses the following research question: How do $PM_{2.5}$ and SO_2 concentrations affect asthma prevalence, and how do individual-level and environmental factors moderate these effects?

To answer this question, the study tests three hypotheses:

H1: Higher levels of $PM_{2.5}$ are positively associated with increased asthma prevalence. (Variables: $PM_{2.5}$, $PM_{2.5}$(5-day lag))

H2: Increased concentrations of sulfur dioxide (SO_2) significantly contribute to the prevalence of asthma. (Variables: SO_2, SO_2(5-day lag))

H3: Individual-level characteristics such as age, BMI, and smoking status influence the extent to which $PM_{2.5}$ and SO_2 affect asthma prevalence. (Interaction Variables: $PM_{2.5}$ × Smoking Status, $PM_{2.5}$ × General Health Status, SO_2 × Smoking Status, SO_2 × General Health Status, SO_2 × BMI, SO_2 × Age)

These hypotheses are grounded in a wide range of epidemiological and environmental health literature that links air quality to respiratory disease while emphasizing the importance of heterogeneity in exposure outcomes.

1.2 Significance of the Research Question

To investigate these hypotheses, this chapter employs a rich dataset combining individual health survey responses with high-resolution air pollution and meteorological data. This empirical strategy is designed to disentangle the effects of $PM_{2.5}$ and SO_2 exposure on asthma prevalence while capturing the heterogeneity across population subgroups. This study adapts its analytical focus to capture both the immediate and lagged effects of pollutants as well as their interactions with personal and environmental characteristics. The following section introduces the data sources and variable construction, outlining how these choices enable a robust test of the proposed hypotheses.

It is essential for both public health and environmental policy to understand the intricate relationship between asthma and air pollution. Asthma is a common global issue that imposes significant burdens on individuals and healthcare systems, reducing quality of life, increasing healthcare expenses, and limiting productivity. While numerous studies have confirmed that air pollutants like $PM_{2.5}$ and SO_2 contribute to respiratory and cardiovascular ailments, much remains unknown about how these impacts vary based on individual traits—age, BMI, smoking status—and environmental factors, including temperature and humidity.

This research tries to shine additional light on this issue by examining not only the independent effect of $PM_{2.5}$ and SO_2 on asthma prevalence but also the determinants that moderate effects within individuals and in the environment. By doing so, this methodology enables a more comprehensive understanding of asthma etiology in polluted settings and how susceptibility may vary within population subgroups and under different weather conditions. By incorporating interaction effects, the paper extends beyond average treatment effects and highlights the importance of context in environmental health studies.

Understanding the connection between air pollutants and health issue is important from a policy-making perspective. Both SO_2 and $PM_{2.5}$ are important air pollutants with long-term socioeconomic and health repercussions. Their sources of emission are industrial production, transport, and combustion of fossil fuels, and their co-occurrence in the air is well known to exacerbate respiratory and cardiovascular diseases. Based on that, there is an overriding need for more stringent regulatory actions. This includes the imposition of strict control over emissions, transitioning to cleaner technologies, and expanding public transportation infrastructure.

Furthermore, by identifying which populations are most vulnerable and under what environmental conditions the effects of pollution are most potent, this work can help policymakers allocate resources more effectively. The findings can be used to guide the development of targeted public health interventions, inform climate change adaptation policy, and help manage health inequities in communities exposed to pollution. Ultimately, addressing air pollution and its impact on health requires a multi-stakeholder approach among policymakers, business actors, and citizens to embrace sustainable, equity-based measures.

1.3 Literature Review and Novelty

Epidemiological studies (Sun et al., 2013) have shown that both long- and short-term exposure to particulate matter (especially PM_{10}) can lead to cardiovascular and respiratory diseases, increasing morbidity and mortality (Dockery et al., 1993; Pope et al., 1995; Schwartz, 1999). These health effects are particularly significant for the elderly and children (Chen & Kan, 2008). Sun et al. geographically mapped the PM_{10} level in 2000 using seven monitoring stations in Lanzhou, China, and population counts for estimating the exposure level. They received the daily records of respiratory disease from four hospitals (2001–2005) and temperature, humidity, and other meteorological factors. They used Generalized Additive Models (GAMs) to examine the association of PM_{10}, meteorological factors, and respiratory admissions. Unlike most other studies relying on similar assumptions of uniform pollutants and population distributions, this research used GIS to get rid of such assumptions, which made the exposure estimates better. They highlighted the unnecessary health and economic burdens of PM_{10} in Lanzhou. They called for urgent air pollution control, noting that over 1 million premature deaths in China in 2010 were linked to $PM_{2.5}$ exposure. One of the significant contributions of their work is the reduction of exposure misclassification by incorporating spatial variability through the use of GIS. Generally, PM_{10} pollution has significantly impacted public health and the economy in Lanzhou, and policy interventions are justified. Compared to Assessment of Population Exposure to PM_{10} for Respiratory Disease in Lanzhou (China) and Its Health-Related Economic Costs Based on

GIS, my study differs significantly in research focus, data, methodology, and contributions. Even though Sun et al. (2013) assessed the exposure to and the economic cost of PM_{10} in one Chinese city using GIS and GAMs, my study examined the effects of $PM_{2.5}$ and SO_2 on asthma in multiple states in the United States. They utilized hospital data and meteorological data; mine is based on the CDC BRFSS survey data (2006–2010) and the EPA pollutant data. Methodologically, we utilize logistic regression and multinomial logit specifications, and we control for person-level variables like gender, BMI, smoking, and exercise, which were uncontrolled in Sun et al.

Special Economic Zones (SEZs) are areas designed to attract foreign investment through tax benefits, stimulating manufacturing and exports to promote economic growth. Nair et al. (2010) explored how infrastructure in the vicinity of the Noida SEZ (NSEZ) in Uttar Pradesh, India, affects air quality and respiratory health. Using a cross-sectional study design, they collected measures of lung function (FEV1, FVC, and PEFR) from 161 residents aged 15 to 60 years from three residential areas near the NSEZ and a control area. PM concentrations were measured and correlated with the observed effects. The effects showed significantly elevated rates of impairment of lung function among residents closer to the NSEZ and declining effects as distance increased, indicating diffusion of pollution as an apparent explanation. The study shows a call to action for the environment, with space sampling in the sense of supporting how living close to SEZs impacts lung health. Compared to Respiratory Health Problems Associated with Infrastructural Development Among Residents Living Near Special Economic Zone in India, our research deviates significantly in the subject matter, source of information, technique, and magnitude. While Nair et al. focused on short-term respiratory impacts of infrastructure-emitted pollution in a localized Indian SEZ by spirometry and statistical inference (ANOVA, chi-square), our study examines the effects of $PM_{2.5}$ and SO_2 on asthma prevalence in various U.S. states using the CDC BRFSS (2006–2010) and EPA air quality data. We apply logistic and multinomial logit regression models with controls for individual-level factors such as gender, BMI, smoking, and exercise. Unlike the Nair et al. local approach, our research uses a broader epidemiological scope and regression analysis to yield more generalizable results.

Air pollution poses a significant risk factor for vulnerable groups such as children, as attested to by hospitalization and mortality caused by acute lower respiratory infections (ALRIs) such as bronchitis. He et al. (2023) explored the short-term effects of $PM_{2.5}$, PM_{10}, SO_2, and NO_2 on hospitalizations due to ALRI in children aged between 0 and 14 years in nine Sichuan Province cities in China using time-series generalized additive models (GAMs) and data from the years 2017–2018. Their study examined 192,079 hospitalizations and air quality data from 183 monitoring stations, adjusting for weather, day of the week, and the lagged effects of pollutants. They found that short-term exposure to air pollutants increased ALRI hospitalization risk significantly, with greater effects from NO_2 and SO_2 in children aged 5–14 years and elevated risk during warmer months. While this study applied lag models in a creative way, it suffered from limitations, including limited time and a lack of interaction analysis between the pollutants. Our study differs from the Short-term Effects and Economic Burden of Air Pollutants on Acute Lower Respiratory Tract Infections in Children in Southwest China in focus, data, methodology, and scope. He et al. aimed at ALRI hospitalization in a single Chinese province using short-term time-series GAMs and estimated economic costs. Conversely, our study examines the long-term effects of $PM_{2.5}$ and SO_2 on asthma prevalence in several U.S. states using the CDC BRFSS (2006–2010) and EPA air quality monitoring data. We apply logistic and multinomial logit models, adjusting for socioeconomic determinants like gender, BMI, smoking, and physical activity, to improve robustness. Our broader geographic area and longer time frame enable more generalizable findings and implications regarding long-term trends. Unlike He et al., with their focus on hospitalization cost, our study assists in comprehending pollution's more general health effects and is amenable to evidence-based policy for asthma prevention and environmental management.

Technologies such as biomass fuel production, particularly crop residue combustion (ACRB), are major causes of air pollution in the developing world, but their immediate contribution to respiratory diseases has not been well investigated. Chakrabarti et al. (2019) studied the effect of ACRB on acute respiratory infections (ARIs) from the records of over 250,000 individuals in Haryana (a northern Indian state affected by ACRB) and two unaffected southern states, from September 2013 to February 2014. After fitting generalized linear models (GLMs) while controlling demographic and household factors, they reported that the risk of ARI was three times higher in ACRB locations, ranging from three times to up to 3.6 times higher in children under five. The study estimated ACRB reduction, which would, on average, prevent 149,000 cases of ARI per annum, with potential economic savings of as much as $1.5 billion. Assuming assumptions such as excluding other pollution sources, authors demanded holistic policy responses, i.e., subsidies and legislation, to mitigate ACRB-related health risks. In contrast to the Risk of Acute Respiratory Infection due to Crop Burning in India, our study aims to capture the effects of chronic exposure to $PM_{2.5}$ and SO_2 on asthma prevalence in a number of U.S. states using CDC BRFSS data (2006–2010) and EPA pollution data. We apply logistic and multinomial logit models with controls for socioeconomic factors (gender, BMI, smoking, and exercise) to enhance the stability of the results. Major differences between the studies are: (1) Disease of interest—Chakrabarti et al. studied ARIs, while mine considers asthma as a chronic illness; (2) Source of pollution—they specifically studied ACRB, while ours considers more general ambient pollution; (3) Geographical location—their study considered India as a country, while ours is conducted in multiple states of the United States, which allows for higher generalizability; (4) Study design—they used GLMs to determine the association between fire events and ARIs, while we employed regression models better suited to examine heterogeneous effects and more prolonged exposure. Our research, therefore, provides new insights into chronic respiratory disease and guides policy formulation in response to chronic ambient air pollution.

While many previous studies have estimated the economic burden of particulate matter in developing countries, Quah and Boon (2003) focused solely on Singapore and estimated the economic burden related to health due to PM_{10} exposure using the damage function method. Using their study, they hypothesized that even the lowest levels of PM_{10}—even within international limits—can cause adverse health effects. Using dose-response relationships and benefit transfer methods, they estimated the total cost to health from PM10 exposure in 1999 as approximately \$3.66 billion or 4.4% of GDP for Singapore. This included a \$1.77 billion cost of mortality, a \$1.89 billion cost of morbidity, and an estimated 591 premature deaths. Although hampered by the absence of locally derived dose-response coefficients, the study demonstrates the enormous potential health and economic gains of reducing air pollution and provides a model for future research using localized data. Our study differs from The Economic Cost of Particulate Air Pollution on Health in Singapore in scope, pollutants under investigation, approach, and contribution. While Quah and Boon calculated PM_{10} exposure cost using cost estimators as a function of mortality and morbidity outcomes, the current work investigates the epidemiological relationship of exposure between $PM_{2.5}$ and SO_2 and asthma occurrence in various states of the United States. Employing CDC BRFSS health survey information (2006–2010) and EPA air pollution data, we employ logistic and multinomial logit estimations to calculate the effect of pollution on asthma, holding principal socioeconomic factors such as gender, BMI, smoking status, and exercise constant. Some of the key differences are: (1) Health outcome—our study targets asthma as a specific chronic disease, whereas Quah and Boon estimate health costs in general; (2) Pollutants—we examine $PM_{2.5}$ and SO_2, finer particulates with more serious health implications; (3) Geographic scope—their analysis is limited to Singapore, whereas ours examines multiple U.S. states to have more generalizability; (4) Methodology—we rely more on regression-based causal estimation than on economic cost modeling. Thus, our study provides a more disease-specific, statistically robust, and geographically comprehensive examination of the long-term health impacts of air pollution.

Our study made several new contributions to the research on air pollution and respiratory health, particularly in the context of asthma. Unlike many earlier studies that examine combined or regional data, this research utilizes individual survey data from the CDC’s BRFSS across multiple U.S. states. This method provides a more precise and more personal understanding of how pollution impacts health. This micro-level approach differs from studies such as Sun et al. (2013) and Quah and Boon (2003), which focus on broader population data or economic cost estimates. Our work provides fresh insights into individual risk factors and susceptibility.

Additionally, this study enhances the methods employed by utilizing multinomial logistic regression models. These models capture changes in asthma status over time, rather than simply reporting yes or no outcomes. This allows for a deeper analysis of asthma progression and its link to $PM_{2.5}$ and SO_2 exposure, going beyond previous research that mainly focuses on hospitalization or morbidity statistics.

Another advancement is the use of detailed temporal and spatial controls, which help isolate the effects of pollution more effectively. By controlling for unmeasured differences across time and location, this approach improves how we identify causal relationships compared to studies like Chakrabarti et al. (2019), which employ broader controls and may weaken the findings.

Finally, the dataset covers diverse demographic groups across the entire United States, rather than being limited to specific areas or age groups. This wide range increases the generalizability and relevance of the findings for policy. It contrasts with earlier localized studies (e.g., He et al. 2023; Nair et al., 2010), allowing for a thorough assessment of how individual and environmental factors interact to affect asthma risk.

In summary, this research addresses significant gaps by combining detailed individual-level data, enhanced statistical methods, rigorous causal controls, and comprehensive population coverage. It provides new evidence that can help guide targeted public health interventions and inform the design of environmental policies.

2. Data

2.1 Data Sources

The data for our study comprises two main components. The first component pertains to data on respiratory diseases, specifically asthma, among respondents. This information was sourced from the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) telephone survey. The BRFSS survey gathers data on the chronic health status and health-related risk behaviors of U.S. residents. Currently, BRFSS data is available from all 50 states and the District of Columbia, with approximately 400,000 respondents providing data annually.

The BRFSS (Behavioral Risk Factor Surveillance System) is a random sample survey organized by the Centers for Disease Control and Prevention (CDC) and is not based on hospital visit data. Specifically, the BRFSS uses Random Digit Dialing (RDD) to survey adult residents of the 50 United States and Washington, D.C., to collect information about their chronic health conditions and health-related behaviors, including whether they have been diagnosed with asthma and whether they continue to experience asthma symptoms. It is not based on patient data from hospital visits, but rather on an epidemiological survey of the general population. Therefore, its asthma data reflect the Self-Reported Prevalence (SRP) of asthma in the overall population, rather than being limited to individual cases of healthcare facility visits.

2.2 Data Processing

To facilitate a cross-sectional study, we collected daily data from 2006 to 2010 from the Selected Metropolitan/Metropolitan Area Risk Trends (SMART) BRFSS dataset. This dataset focuses on U.S. metropolitan areas, making it suitable for capturing short-term fluctuations and helpful in studying short-term variations, identifying peaks, and assessing the direct impacts of elevated pollutant levels on asthma.

An important step in processing these sub-datasets is to match the metropolitan and micropolitan statistical area codes (MMSA) with the state and county codes. This matching is crucial for accurately aligning the pollution data with the health data later on. The second component of the data consists of pollutant data, including $PM_{2.5}$ and SO_2. This data was obtained from the United States Environmental Protection Agency (EPA) website and includes county-level pollutant concentrations for each state in the United States. In addition to daily pollutant data, the EPA site also collects data on wind, temperature, barometric pressure, relative humidity, and dew point from 1980 to the present.

To integrate these datasets, we combined $PM_{2.5}$, SO_2, and related meteorological data with the daily BRFSS data for the period from 2006 to 2010, using the state code, county code, and date as matching criteria. For the monthly BRFSS data, we first averaged the daily Air Pollution Index (API) and then merged it with the health data.

During data processing, we first screened the BRFSS data from 2006 to 2010, retaining the variables we wanted to study, including the height, weight, age, and living habits of the respondents, and then merged the data for these five years vertically. Next, we began to collect data on air pollution and weather factors. After collecting daily data for the corresponding years from the EPA website (including $PM_{2.5}$, SO_2, air pressure, temperature, relative humidity, etc.), we also merged them vertically. The next step is to pave the way for merging BRFSS data and meteorological data. First, we created a new variable of date plus county for both sets of data. Since the BRFSS data lacks a county code, it is replaced with an MMSA code. We need to find the county code corresponding to the MMSA code on the government website and replace it one by one. After completing this step, we can merge the two sets of data horizontally according to the standard new variable 'date + county' and delete all the quantities containing NA in the new dataset.

In the original BRFSS data, some respondents used data instead of NA, and we need to remove these data points. For example, when a respondent uses the number 9 to answer his smoking frequency, it means that he is not clear about his smoking habits or does not want to answer. This phenomenon also occurs in variables including height and weight. We also deleted the data containing number 9. The data on height and weight in the BRFSS data are not of a common size. According to the instructions on the data website, we recalculated the height, weight, and BMI, dividing the height and weight in the original data by 100 to obtain data in m and kg. For BMI, we recalculated it according to the official formula for calculating BMI (weight (kg)/height (m)^2).

2.3 Summary Statistics

Table 2.1 Summary statistics of binomial model on having asthma symptoms or not currently

Variable	Obs.	Mean	Max	Min	SD
CASTHMA	98908	0.086	1	0	0.281
ASTHMST	98908	2.784	3	1	0.585
$PM_{2.5}$ (μg/m^3)	98908	11.617	95.252	0	6.525
SO_2 (μg/m^3)	98908	2.985	227.887	0	6.776
Relative humidity (%)	98908	64.681	203.625	-7.917	18.221
Temperature (°F)	98908	57.701	101.75	-19.167	17.616
BMI (m/kg^2)	98908	27.589	183.237	6.489	6.101
Sex (Female = 0, Male = 1)	98908	1.619	2	1	0.485
Age	98908	4.291	6	1	1.476
Smoking status	98908	3.278	4	1	0.981
General status	98908	2.553	5	1	1.115
Exercise status	98908	1.255	2	1	0.449
Wind/ Wind speed (mph)	98908	4.529	51.813	0	2.514

Note: The table above lists the means, standard deviations, minimum and maximum values, and counts of observations for all the variables used in the analysis of the binomial model for having asthma symptoms or not currently, all data are aggregated based on daily data.

Table 2.1 provides an overview of environmental factors ($PM_{2.5}$, relative humidity, temperature) and individual-level factors (like sex, age, BMI, smoking, and health status). The variability, as indicated by the standard deviations, shows the spread of each variable, helping assess how uniform or diverse the data is for each factor. The dependent variable, CASTHMA, is a binary indicator with a mean of 0.086, reflecting that approximately 8.6% of observations report asthma symptoms. The other dependent variable ASTHMST, indicating asthma status with values from 1 to 3, has a mean of 2.784.

$PM_{2.5}$ (µg/m³): The mean of 11.617 µg/m³ represents the average concentration of $PM_{2.5}$ in the air. SO_2 (µg/m³): The mean of 2.985 µg/m³ represents the average sulfur dioxide concentration. Relative humidity (%): The mean of 64.681% indicates the average relative humidity. Temperature (°F): The mean of 57.701°F reflects the average temperature across the dataset. Wind speed (mph): The mean of 4.529 mph indicates an average wind speed. Sex: This is a binary variable (2 = female, 1 = male). BMI (m/kg²): The mean of 27.589 suggests that the average body mass index (BMI) is in the overweight. Age: The mean of 4.291 suggests that age is recorded in categorical groups (Age 18 to 24 = 1, Age 25 to 34 = 2, Age 35 to 44 = 3, Age 45 to 54 = 4, Age 55 to 64 = 5, Age 65 or older =6). Smoking status: This is a categorical variable indicating smoking habits, from 1 to 4. (Current smoker now smokes every day = 1, Current smoker now smokes some days = 2, Former smoker = 3, Never smoked = 4). General status: This variable assesses general health status on a scale from 1 to 5. (Good=1,2,3). Exercise status: This variable likely represents exercise habits with a range from 0 to 1 (Exercise = 1, No exercise = 2).

3. Empirical Framework

3.1 Logistic Regression Model

$$\mathrm{logit}\big(\Pr(\mathrm{CASTHMAijt} = 1)\big) = a_0 + a_y + a_m + a_d + a_j + \sum_{k=1}^{6} \beta_k X_{jt}^{(k)} + \sum_{l=1}^{6} \gamma_l Z_{ijt}^{(l)} + \sum_{m=1}^{2} \delta_m (\mathrm{PMAve}_{jt} \cdot Z_{ijt}^{(k)}) + \sum_{n=1}^{5} \eta_n (\mathrm{SOAve}_{jt} \cdot Z_{ijt}^{(k)}) + \varepsilon_{ijt}$$

$$\log\left(\frac{P(\mathrm{CASTHMA}_{ijt}=1)}{1-P(\mathrm{CASTHMA}_{ijt}=1)}\right) = \alpha_0 + \alpha_y + \alpha_m + \alpha_d + a_j + \beta_1 \cdot \mathrm{PMAve}_{jt} + \beta_2 \cdot \mathrm{SOAve}_{jt} + \beta_3 \cdot \mathrm{RHAve}_{jt} + \beta_4 \cdot \mathrm{TEMPAve}_{jt} + \beta_5 \cdot \mathrm{PMlag}_{jt} + \beta_6 \cdot \mathrm{SOlag}_{jt} + \gamma_1 \cdot \mathrm{BMI}_{ijt} + \gamma_2 \cdot \mathrm{AGE}_{ijt} + \gamma_3 \cdot \mathrm{SMOKER}_{ijt} + \gamma_4 \cdot \mathrm{GENHLTH}_{ijt} + \gamma_5 \cdot \mathrm{SEX}_{ijt} + \gamma_6 \cdot \mathrm{EXERANY}_{ijt} + \delta_1(\mathrm{PMAve}_{jt} \cdot$$

$$\mathrm{SMOKER3}_{ijt}) + \delta_2(\mathrm{PMAve}_{jt} \cdot \mathrm{GENHLTH}_{ijt}) + \eta_1(\mathrm{SOAve}_{jt} \cdot \mathrm{SMOKER}_{ijt}) + \eta_2(\mathrm{SOAve}_{jt} \cdot \mathrm{GENHLTH}_{ijt}) + \eta_3(\mathrm{SOAve}_{jt} \cdot \mathrm{AGE}_{ijt}) + \eta_4(\mathrm{SOAve}_{jt} \cdot \mathrm{SEX}_{ijt}) + \eta_5(\mathrm{SOAve}_{jt} \cdot \mathrm{BMI}_{ijt}) + \varepsilon_{ijt} \quad (1)$$

i: It represents individual players in survey.
j: This index represents a geographic or spatial unit. In this context, it refers to a specific county nationwide where the survey is conducted.
t: It denotes the temporal dimension of the analysis. Here, it represents the exact date when the survey is taken.
$X_{jt}^{(k)}$: is a set of environmental variables for location j at time t, where k=1,…,6. These include ($PMAve_{jt}$, $SOAve_{jt}$, $RHAve_{jt}$, $TEMPAve_{jt}$, $PMlag_{jt}$, $SOlag_{jt}$)
$Z_{ijt}^{(l)}$: denotes player-specific characteristics for player i at location j and time t, where l=1,…,6. These include (BMI_{ijt}, AGE_{ijt}, $SMOKER_{ijt}$, $GENHLTH_{ijt}$, SEX_{ijt}, $EXERANY_{ijt}$)
The interaction terms ($PMAve_{jt} \cdot Z_{ijt}^{(k)}$) ($SOAve_{jt} \cdot Z_{ijt}^{(k)}$): capture how the effect of $PM_{2.5}$ and SO_2 vary depending on individual.

Variable Definitions:

$CASTHMA_{ijt}$: If being attacked or not

$PMAve_{jt}$: average concentration of $PM_{2.5}$ particles

$SOAve_{jt}$: average concentration of sulfur dioxide

$RHAve_{jt}$: average relative humidity

$TEMPAve_{jt}$: average temperature

$PMlag_{jt}$: 5 days accumulated $PM_{2.5}$ level before the day of survey

$SOlag_{it}$: 5 days accumulated SO_2 level before the day of survey

BMI_{ijt}: Body mass index

SEX_{ijt}: Gender

AGE_{ijt}: Age

$SMOKER_{ijt}$: four-level variable indicating smoker status

$GENHLTH_{ijt}$: general health status

$EXERANY_{ijt}$: whether the individual has exercised in the past 30 days

ε_{ijt}: Error term.

3.2 Fit Multinomial Logistic Regression Model

$$\log\left(\frac{\mathrm{P(ASTHMST_{ijt}=k)}}{\mathrm{P(ASTHMST_{ijt}=1)}}\right) = a_0 + a_y + a_m + a_d + a_j + \sum_{k=1}^{6} \beta_k X_{jt}^{(k)} + \sum_{l=1}^{6} \gamma_l Z_{ijt}^{(l)} + \sum_{m=1}^{2} \delta_m (\mathrm{PMAve}_{jt} \cdot Z_{ijt}^{(k)}) + \sum_{n=1}^{5} \eta_n (\mathrm{SOAve}_{jt} \cdot Z_{ijt}^{(k)}) + \varepsilon_{ijt}$$

Where:

k=2,3 comparing to baseline k=1 (currently have asthma)

$$\log\left(\frac{\mathrm{P(ASTHMST_{ijt}=2)}}{\mathrm{P(ASTHMST_{ijt}=1)}}\right) = \alpha_0 + \alpha_y + \alpha_m + \alpha_d + a_j + \beta_1 \cdot \mathrm{PMAve}_{jt} + \beta_2 \cdot \mathrm{SOAve}_{jt} + \beta_3 \cdot \mathrm{RHAve}_{jt} + \beta_4 \cdot \mathrm{TEMPAve}_{jt} + \beta_5 \cdot \mathrm{PMlag}_{jt} + \beta_6 \cdot \mathrm{SOlag}_{jt} + \gamma_1 \cdot \mathrm{BMI}_{ijt} + \gamma_2 \cdot \mathrm{AGE}_{ijt} + \gamma_3 \cdot \mathrm{SMOKER}_{ijt} + \gamma_4 \cdot \mathrm{GENHLTH}_{ijt} + \gamma_5 \cdot \mathrm{SEX}_{ijt} + \gamma_6 \cdot \mathrm{EXERANY}_{ijt} + \delta_1(\mathrm{PMAve}_{jt} \cdot \mathrm{SMOKER3}_{ijt}) + \delta_2(\mathrm{PMAve}_{jt} \cdot \mathrm{GENHLTH}_{ijt}) + \eta_1(\mathrm{SOAve}_{jt} \cdot \mathrm{SMOKER}_{ijt}) + \eta_2(\mathrm{SOAve}_{jt} \cdot \mathrm{GENHLTH}_{ijt}) + \eta_3(\mathrm{SOAve}_{jt} \cdot \mathrm{AGE}_{ijt}) + \eta_4(\mathrm{SOAve}_{jt} \cdot \mathrm{SEX}_{ijt}) + \eta_5(\mathrm{SOAve}_{jt} \cdot \mathrm{BMI}_{ijt}) + \varepsilon_{ijt} \quad (2)$$

$$\log\left(\frac{\mathrm{P(ASTHMST_{ijt}=3)}}{\mathrm{P(ASTHMST_{ijt}=1)}}\right) = \alpha_0 + \alpha_y + \alpha_m + \alpha_d + a_j + \beta_1 \cdot \mathrm{PMAve}_{jt} + \beta_2 \cdot \mathrm{SOAve}_{jt} + \beta_3 \cdot \mathrm{RHAve}_{jt} + \beta_4 \cdot \mathrm{TEMPAve}_{jt} + \beta_5 \cdot \mathrm{PMlag}_{jt} + \beta_6 \cdot \mathrm{SOlag}_{jt} + \gamma_1 \cdot \mathrm{BMI}_{ijt} + \gamma_2 \cdot \mathrm{AGE}_{ijt} + \gamma_3 \cdot \mathrm{SMOKER}_{ijt} + \gamma_4 \cdot \mathrm{GENHLTH}_{ijt} + \gamma_5 \cdot \mathrm{SEX}_{ijt} + \gamma_6 \cdot \mathrm{EXERANY}_{ijt} + \delta_1(\mathrm{PMAve}_{jt} \cdot \mathrm{SMOKER3}_{ijt}) + \delta_2(\mathrm{PMAve}_{jt} \cdot \mathrm{GENHLTH}_{ijt}) + \eta_1(\mathrm{SOAve}_{jt} \cdot \mathrm{SMOKER}_{ijt}) + \eta_2(\mathrm{SOAve}_{jt} \cdot \mathrm{GENHLTH}_{ijt}) + \eta_3(\mathrm{SOAve}_{jt} \cdot \mathrm{AGE}_{ijt}) + \eta_4(\mathrm{SOAve}_{jt} \cdot \mathrm{SEX}_{ijt}) + \eta_5(\mathrm{SOAve}_{jt} \cdot \mathrm{BMI}_{ijt}) + \varepsilon_{ijt} \quad (3)$$

i: It represents individual players involved in surveys.

j: This index represents a geographic or spatial unit. In this context, it refers to a specific county nationwide where the survey is conducted.

t: It denotes the temporal dimension of the analysis. Here, it represents the exact date when the survey is taken.

$X_{jt}^{(k)}$: is a set of environmental variables for location j at time t, where k=1,…,6. These include (PMAve_{jt}, SOAve_{jt}, RHAve_{jt}, $\mathrm{TEMPAve}_{jt}$, PMlag_{jt}, SOlag_{jt})

$Z_{ijt}^{(l)}$: denotes player-specific characteristics for player i at location j and time t, where l=1,…,6. These include (BMI_{ijt}, AGE_{ijt}, SMOKER_{ijt}, $\mathrm{GENHLTH}_{ijt}$, SEX_{ijt}, $\mathrm{EXERANY}_{ijt}$)

The interaction terms ($\mathrm{PMAve}_{jt} \cdot Z_{ijt}^{(k)}$) ($\mathrm{SOAve}_{jt} \cdot Z_{ijt}^{(k)}$): capture how the effect of $PM_{2.5}$ and SO_2 vary depending on individual.

Variable Definitions:

ASTHMST: which categorizes individuals into those who currently have asthma (1), those who had asthma but do not have it anymore (2), and those who have never been diagnosed with asthma (3).

The logistic regression model estimates the probability of asthma occurrence based on air pollution levels ($PM_{2.5}$ and SO_2), weather conditions (temperature and humidity), and individual characteristics. It tests the hypothesis that higher concentrations of $PM_{2.5}$ and SO_2 increase asthma prevalence. The model also includes interaction terms to examine how personal factors such as age, BMI, and smoking modify the effects of pollution on asthma risk. Additionally, temperature and relative humidity are incorporated to assess their independent influence on asthma rates. Significant findings in these variables and interactions would support the hypotheses regarding pollution, environment, and individual susceptibility.

4. Results

4.1 Logistic Regression Model on How Asthma Being Affected

Table 2.2 Effects of Air Pollution on Asthma Symptoms Prevalence

Dependent variable: Asthma symptoms (=1 if present, 0 otherwise)		
Variables	**Asthma Prevalence Models**	
	(1)FE Logit	**(2) FE Logit with Interaction**
Intercept	-4.49738***	-4.58322***
	(0.62737)	(0.63422)
Pollution Exposure		
$PM_{2.5}$ (μg/m³)	0.00087*	0.00299*
	(0.00198)	(0.00787)
SO_2 (μg/m³)	0.00108	0.01749
	(0.00212)	(0.01187)
$PM_{2.5}$ (5-day lag) (μg/m³)	0.00394**	0.00398**
	(0.00191)	(0.00191)
SO_2 (5-day lag) (μg/m³)	0.00131	0.00112
	(0.00207)	(0.00209)
Weather Controls		
Relative Humidity (%)	-0.00140*	-0.00140*
	(0.00088)	(0.00088)
Temperature (°F)	0.00113	0.00117
	(0.00128)	(0.00128)
Health Indicators		

BMI	0.03484***	0.03541***
	(0.00164)	(0.00180)
Age	-0.11397***	-0.10213***
	(0.00809)	(0.00886)
Sex (Male = 1, Female = 2)	0.55319***	0.56066***
	(0.02583)	(0.02813)
Behavioral Factors		
Smoking Status (Frequently = 1, Never = 4)	-0.07463***	-0.09477***
	(0.01134)	(0.02316)
General Health Status(Good = 1,2,3)	0.38991***	0.41129***
	(0.01105)	(0.02140)
Exercise Status (Exercise = 1, No exercise = 2)	0.00641	0.00719***
	(0.02548)	(0.02548)
Interaction Variables		
$PM_{2.5}$ × Smoking Status	—	0.00141
		(0.00172)
$PM_{2.5}$ × General Health Status	—	-0.00225
		(0.00155)
SO_2 × Sex	—	-0.00239
		(0.00327)
SO_2 × Smoking Status	—	0.00132
		(0.00156)
SO_2 × General Health Status	—	0.00153
		(0.00158)
SO_2 × BMI	—	-0.00020
		(0.00024)
SO_2 × Age	—	-0.00363***
		(0.00109)
Model Fit		
Log Likelihood	-27,597.05	-27,589.73
Deviance	55,194.10	55,179.46
Observations	98,908	98,908
***p<0.001, **p<0.01, *p<0.05		

Note: The above table shows the models, namely Logistic model, and with fixed effect on location. The data are all from the original relevant daily data.

Table 2.2 presents the results from fixed-effects logit models estimating the probability of reporting asthma symptoms as a function of air pollution exposure, meteorological conditions, individual health and behavioral characteristics, and a set of interaction terms. The dependent variable is binary, taking value 1 if the individual reports asthma symptoms and 0 otherwise.

In both models, current exposure to $PM_{2.5}$ and SO_2 show positive associations with asthma symptoms (coefficient of $PM_{2.5}$ = 0.00087 and 0.00299, $p<0.05$ in both models) (coefficient of SO_2 = 0.00108 and 0.01749), indicating that one unit increase in $PM_{2.5}$ and SO_2 will lead to average 0.00193 and 0.00929 increase in log-odds of the probability of reporting asthma symptoms. $PM_{2.5}$ concentration with a 5-day lag is significantly and positively associated with asthma risk (coefficient = 0.00394 and 0.00398, $p < 0.01$ in both models). This result is robust across both specifications, suggesting that short-term delayed exposure may be more relevant for respiratory responses than immediate exposure. In contrast, the 5-day lag of SO_2 remains insignificant in both models. A coefficient of 0.00131 and 0.00112 indicates that, holding other variables constant, a one-unit increase in SO_2-lag will lead to an average 0.00122 increase in the log-odds of reporting asthma symptoms.

Turning to weather conditions, relative humidity exhibits a small but statistically significant negative effect (coefficient = -0.00140, $p < 0.05$ in both models), indicating that holding other variables constant, one unit increase in relative humidity will lead to 0.00140 log-odds of reporting asthma symptoms, implying that increased humidity is associated with a slight decrease in asthma symptom prevalence, possibly due to reduced airborne pollutant penetration or respiratory irritation. Temperature shows a positive but statistically insignificant association, indicating no strong independent effect on asthma symptoms within this model.

Among health indicators, BMI is positively and significantly related to asthma risk (coefficient = 0.03484 and 0.03541, $p < 0.001$), indicating that holding other variables constant, one unit increase in BMI leads to an average 0.03513 increase in log-odds of reporting asthma symptoms, suggesting that higher body mass is associated with greater susceptibility to asthma symptoms, possibly due to inflammatory or mechanical respiratory burdens. Age has a strong and negative effect (coefficient = -0.11397 and -0.10213, both $p < 0.001$), indicating that younger individuals are more likely to report asthma symptoms, which may reflect age-related immune differences or variations in self-reporting behavior. Sex also plays an important role: females are significantly more likely to report asthma symptoms than males (coefficient $\approx$ 0.56, $p < 0.001$), even after controlling for individual fixed effects, health status, and environmental exposures. Among behavioral variables, smoking status is inversely related to asthma symptoms (coefficient = -0.07463 and -0.09477, $p < 0.001$ in both models), indicating that holding other variables constant, a one-unit increase in smoking status leads to an average 0.08470 increase in the log-odds of reporting asthma symptoms. This is consistent with our intuition that people who smoke are more likely to suffer from asthma attacks. Similarly, worse general health status is strongly and positively associated with asthma prevalence (coefficient = 0.39–0.41, $p < 0.001$ in both models), consistent with the idea that individuals in poor health are more vulnerable to asthma attacks. Exercise status, interestingly, is not significant in the baseline model, but becomes significant in the interaction model (coefficient = 0.00719, $p < 0.001$), suggesting a possible association between physical inactivity and asthma risk.

Model 2 incorporates several interaction terms to explore heterogeneity in pollution effects across demographic and health dimensions. While most interactions are not statistically significant, their directions offer insights: During these interactions, the only statistically significant interaction is SO_2 × Age (coefficient = -0.00363, $p < 0.001$), suggesting that the impact of SO_2 exposure on asthma declines with age. This could reflect adaptation or reduced outdoor exposure among older individuals.

4.2 Fit Multinomial Logistic Regression Model on How Asthma Being Affected

Table 2.3 Multinomial Logistic Regression of Air Pollution Effects on Respiratory Health Outcomes

Dependent variable: Respiratory health status (Reference category = 1)		
Variables	**Asthma Dynamic Models**	
	Category 2	**Category 3**
Intercept	1.62082***	4.52419***
	(0.23184)	(0.14183)
Pollution Exposure		
$PM_{2.5}$ (μg/m³)	-0.00458	0.00195
	(0.01312)	(0.00796)
SO_2 (μg/m³)	-0.01948	-0.01863
	(0.01789)	(0.01178)
$PM_{2.5}$ (5-day lag)(μg/m³)	-0.00357***	-0.00123*
	(0.00308)	(0.00186)
SO_2 (5-day lag)(μg/m³)	-0.00572*	-0.00434**
	(0.00321)	(0.00185)
Weather Controls		
Relative Humidity (%)	0.00047	-0.00003
	(0.00101)	(0.00064)
Temperature (°F)	0.00307**	0.00135*
	(0.00109)	(0.00066)
Health Indicators		
BMI	-0.01515***	-0.03674***
	(0.00304)	(0.00178)
Age	0.05401***	0.09953***
	(0.01387)	(0.00874)
Sex (Male = 1, Female = 2)	-0.58007***	-0.54822***
	(0.04345)	(0.02814)
Behavioral Factors		
Smoking Status (Frequently =1, Never =4)	0.09223**	0.10997***
	(0.03942)	(0.02340)
General Health Status(Good = 1,2,3)	-0.34875***	-0.39840***
	(0.03625)	(0.02156)
Exercise Status (Exercise = 1, No exercise = 2)	-0.13626**	0.01649
	(0.04467)	(0.02545)
Interaction Variables		
$PM_{2.5}$ × Smoking Status	-0.00106	-0.00183
	(0.00295)	(0.00175)
$PM_{2.5}$ × General Health Status	0.00256	0.00179
	(0.00266)	(0.00158)

SO_2 × Sex	0.00403	0.00144
	(0.00479)	(0.00314)
SO_2 × Smoking Status	-0.00081	-0.00142
	(0.00235)	(0.00151)
SO_2 × General Health Status	0.00151	-0.00257
	(0.00238)	(0.00152)
SO_2 × BMI	0.00026	0.00016
	(0.00032)	(0.00023)
SO_2 × Age	0.00148	0.00411***
	(0.00163)	(0.00107)
Model Fit		
Residual Deviance	89,627.96	
Observations	98,908	
***p<0.001, **p<0.01, *p<0.05		

Note: This is the fit multinomial regression model result showing how asthma being affected by various factors. (2) and (3) show the comparison between current asthma symptoms (1), previous asthma but no current symptoms (2), and never diagnosed with asthma (3).

Table 2.3 presents the estimation results of the fixed-effects multinomial logit model, which is used to analyze the influencing factors of an individual's current asthma status. The dependent variable is a three-category variable:

Category 1 (reference group): currently having asthma symptoms.

Category 2: having asthma in the past but currently asymptomatic.

Category 3: never diagnosed with asthma.

The coefficient explains the trend of change relative to the current asthma (category 1). For example, a positive coefficient for Category 2 indicates that an increase in this variable reduces the individual's asthma symptoms; a negative coefficient for Category 3 means that an increase in this variable increases the likelihood that the individual will develop asthma.

The coefficients of $PM_{2.5}$ current value for both categories (-0.00458 and 0.00195) are not significant, indicating that after controlling for other variables, current exposure to $PM_{2.5}$ has no strong correlation with the distinction between asthma status (1 vs. 2 or 3). However, from the coefficient, it can be found that when $PM_{2.5}$ increases by one unit, the log-odds probability of going from asthma-free to asthma-symptomatic increases by 0.00458. However, although it is only a small probability, the current data suggest that $PM_{2.5}$ has a negative impact on the transition from no asthma to new asthma, although the data results are not statistically significant.

The 5-day lag of $PM_{2.5}$ is significantly negative for Category 2 (-0.00357*) and Category 3 (-0.00123*) ($p < 0.05$ in both models), which means that for every unit increase in $PM_{2.5}$-lag, people with asthma but no symptoms have a 0.00357 increase in log-odds probability of inducing asthma symptoms, and a 0.00123 increase in log-odds probability of going from no history of asthma to newly developing asthma.

The 5-day lag of SO_2 is significantly negative for Category 2 (-0.00572*) and Category 3 (-0.00434**) ($p < 0.05$ and $p < 0.01$), further indicating that long-term SO_2 exposure is positively associated with both new-onset asthma and asthma relapse.

The temperature variable is positively correlated with Category 2 (0.00307**) and Category 3 (0.00135*) ($p < 0.01$ and $p < 0.05$), which means that holding other variables constant, one unit increase in temperature leads to 0.00307 log-odds decrease in probability of asthma recurrence and 0.00135 log-odds decrease in probability of new asthma, which may indicate that high-temperature environment is helpful for both the relief and prevention of asthma. Relative humidity is not significant for both categories, indicating that the effect of humidity on the transition of asthma status is unclear.

BMI is significantly negatively correlated with Category 2 (-0.01515***) and Category 3 (-0.03674***), indicating that holding other variables constant, each unit increase in BMI leads to a 0.01515 increase in log-odds of the probability of recurrence of asthma symptoms, and lead to a 0.03674 increase in log-odds of the probability of new onset of asthma. This proves that people with high BMI are more likely to be attacked by asthma and are more likely to acquire asthma. The age variable is positive for Category 2 (0.05401***) and positive for Category 3 (0.09953***), indicating that holding other variables constant, each unit increase in age will lead to a 0.05401 decrease in log-odds of the probability of recurrence of asthma symptoms, and will lead to a 0.09953 decrease in the probability of new onset of asthma. This may suggest that older individuals are less likely to acquire asthma. The gender variable is significantly negative for both categories (-0.58007*** and -0.54822***), indicating that female asthmatics are more susceptible to asthma attacks and more likely to develop asthma than males. Smoking status is significantly positive with Category 2 (0.09223**) and Category 3 (0.10997***), indicating that for every unit increase in smoking frequency, the log-odds probability of recurrence of asthma symptoms and new asthma will increase by 0.09223 and 0.10997, respectively. In other words, smoking is a risk factor for persistent asthma. The health status variable is significantly negative for both categories (-0.34875*** and -0.39840***), indicating that the better the overall health status, the less likely it is to be attacked by asthma and the less likely it is to develop asthma. The exercise status variable is significantly negative for Category 2 (-0.13626**), indicating that asthma patients who lack exercise are more likely to experience repeated asthma attacks. Holding other variables constant, a one-unit decrease in exercise status will lead to a 0.13626 increase in the log-odds of having asthma symptoms.

Interaction terms are used to explore whether pollution has different effects on different groups of people (by behavior, gender, age, etc.): SO_2: Age variable is significantly positive for Category 3 (0.00411***), meaning that older people are less likely to acquire new asthma in SO_2 exposure, which may be related to their concern for their health. Older people are less likely to go out in highly polluted weather, and the proportion of older adults newly diagnosed with asthma will be smaller. Other interaction terms were not significant, indicating that the interactions between pollution and most individual characteristics in the existing models were not strong enough to explain the differences in asthma status effectively.

4.3 Dose Response Function on Logit Regression Model

As in Chapter 1, we apply dose-response functions to explore potential nonlinear associations between air pollution exposure ($PM_{2.5}$ and SO_2) and asthma prevalence. This approach allows us to visualize how varying pollution levels influence the probability of current or past asthma, and to identify threshold that linear models may overlook.

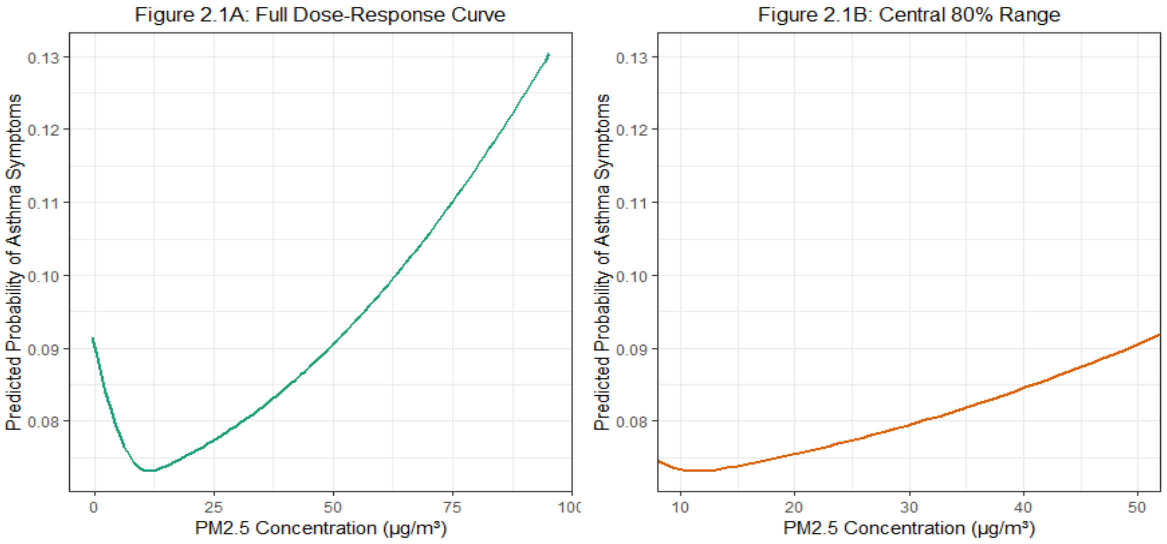
Figure 2.1A: Full Dose-Response Curve

Figure 2.1B: Central 80% Range

Figure 2.1 Dose-Response Curve between $PM_{2.5}$ and Probability of Asthma

Figure 2.1 illustrates the dose-response curve of $PM_{2.5}$ levels and predicted probability of asthma. The shape of the observation curve is U-shaped. The predicted probability of asthma decreases at very low levels (0–10 μg/m³) prior to attaining a trough at approximately 10 μg/m³. After the initial decreasing trend, the curve rises, indicating a steady and increasing rate of increase in asthma probability with increasing levels of $PM_{2.5}$, especially after 40 μg/m³. The trend shows that high levels of $PM_{2.5}$ are definitely associated with an increase in the probability of asthma development. This linear relationship is consistent with recent epidemiological evidence indicating that fine particulate matter can penetrate deeply into the pulmonary system, generating inflammation and respiratory discomfort.

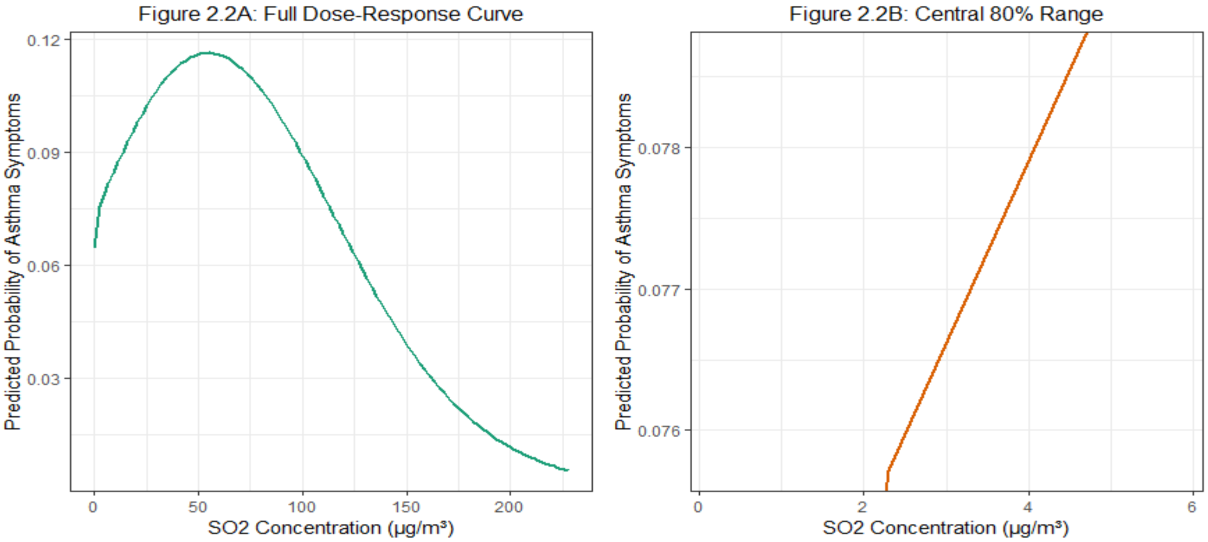

Figure 2.2 Dose-Response Curve between SO_2 and Probability of Asthma

Figure 2.2 illustrates the dose-response curve of SO_2 concentration and the estimated probability of asthma. The curve takes the form of an inverted U-shape, indicating that asthma probability initially increases with higher SO_2 concentrations, with the peak reached at around 60 ppb. After this point, the curve declines, indicating a decrease in predicted asthma risk with higher concentrations. This impact may indicate the nonlinear effects of SO_2 on respiratory health. This upward trend aligns with the longstanding irritating characteristic of sulfur dioxide, which can exacerbate asthma and increase respiratory inflammation, particularly in vulnerable individuals. The decrease in higher concentrations, although unexpected, could be explained by factors such as decreased exposure times at extremely high pollution levels (e.g., school closings and reduced outdoor activity).

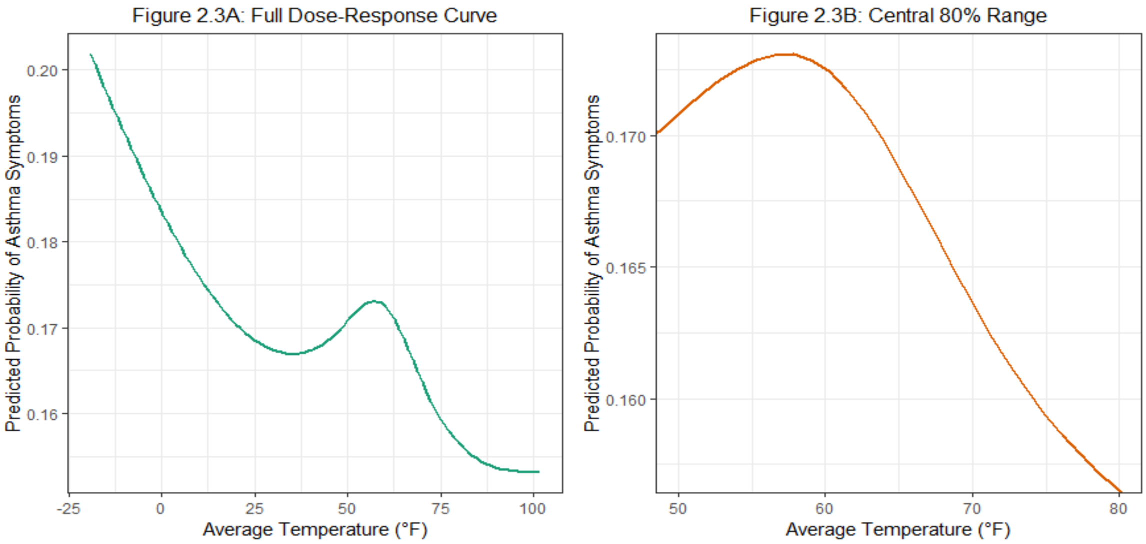

Figure 2.3 Dose-Response Curve between Temperature and Probability of Asthma

Figure 2.3 shows how average temperature affects the likelihood of developing asthma. The line curves downward. The likelihood of asthma drops steeply from cold temperature (at about -25°F) to around 30°F, is steady from 30°F to 50°F, and then rises slightly, reaching a peak at around 60°F. From here, the line curves downward again, reaching its lowest likelihood of asthma at around 100°F. This intricate pattern suggests that extremely cold weather is strongly associated with a higher risk of asthma, likely due to the potential for airway tightening and increased sensitivity from the cold. The intermediate temperatures appear less dangerous, maybe offering a steady and comfortable level. A minor rise of nearly 60°F could be due to seasonal allergens or pollution influencing the temperature.

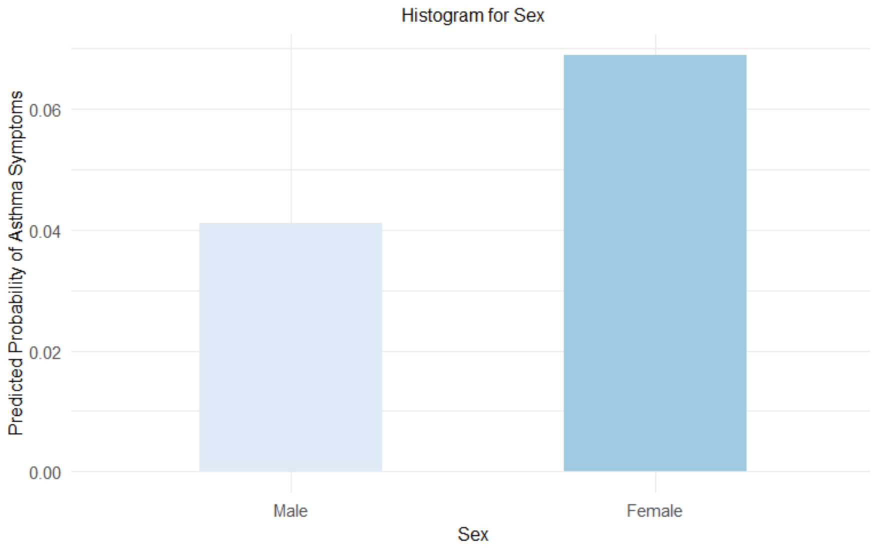

Figure 2.4 Histogram between Sex and Probability of Asthma

Figure 2.4 illustrates the sex-stratified model-predicted probability of asthma, indicating that females have a higher probability than males. This suggests a potential sex difference in the risk of asthma, which is consistent with broader literature indicating greater asthma prevalence or severity in females, possibly due to hormonal or physiological factors.

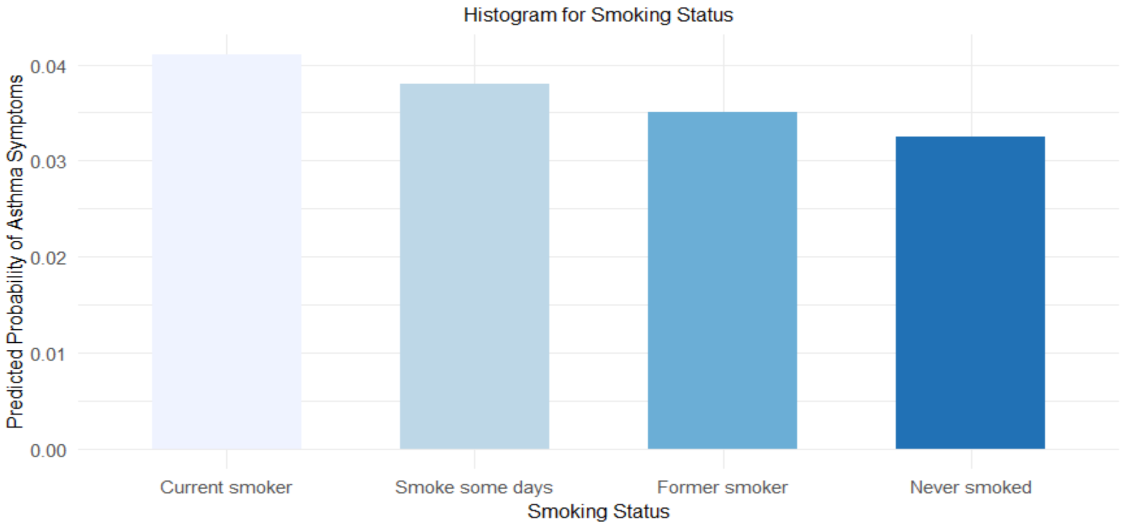

Figure 2.5 Histogram between Smoking Status and Probability of Asthma

Figure 2.5 displays the predicted probability of asthma by smoking status, and there is a steep gradient where the highest probability of asthma occurs in current smokers ("Smoke some days"), followed by former smokers, and finally those who never smoked. The trend is consistent with known results correlating smoking—past and present—to asthma risk based on airway damage and inflammation. The enhanced risk in former smokers compared to never-smokers suggests an ongoing respiratory effect despite smoking cessation.

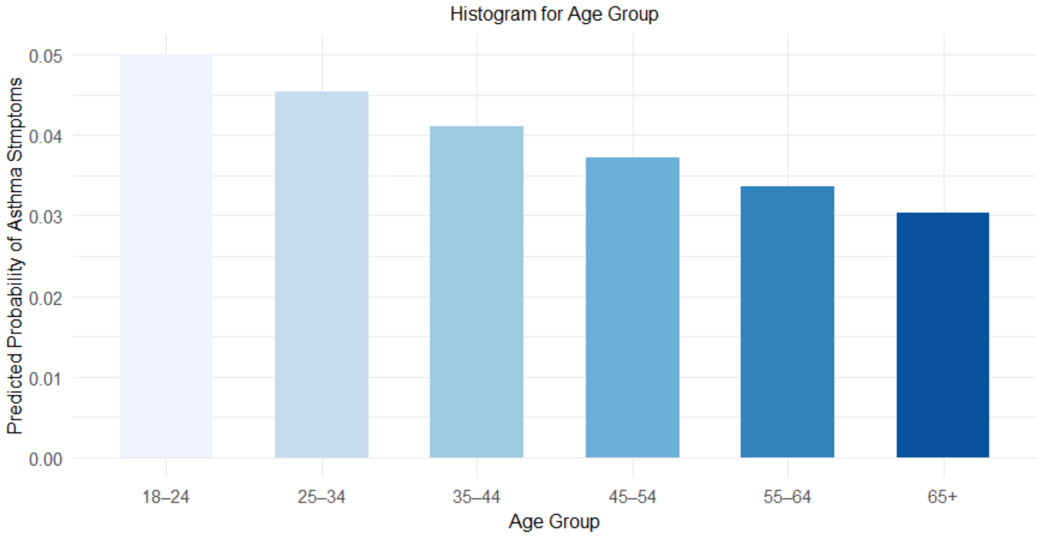

Figure 2.6 Histogram between Age and Probability of Asthma

Figure 2.6 illustrates the relationship between age groups and the estimated probability of experiencing asthma symptoms. From the findings, there is a declining trend: those of younger age, particularly the 18-24 years group, have the highest estimated probability of experiencing asthma symptoms in response to pollution, and the oldest group (65+) has the lowest. This trend suggests that younger people are more susceptible to the respiratory effects of air pollution, perhaps due to greater outdoor exposure, heightened immune sensitivity, or bias in reporting. The elderly may be less susceptible or have compensatory physiological adaptations, so their observed effect is diminished.

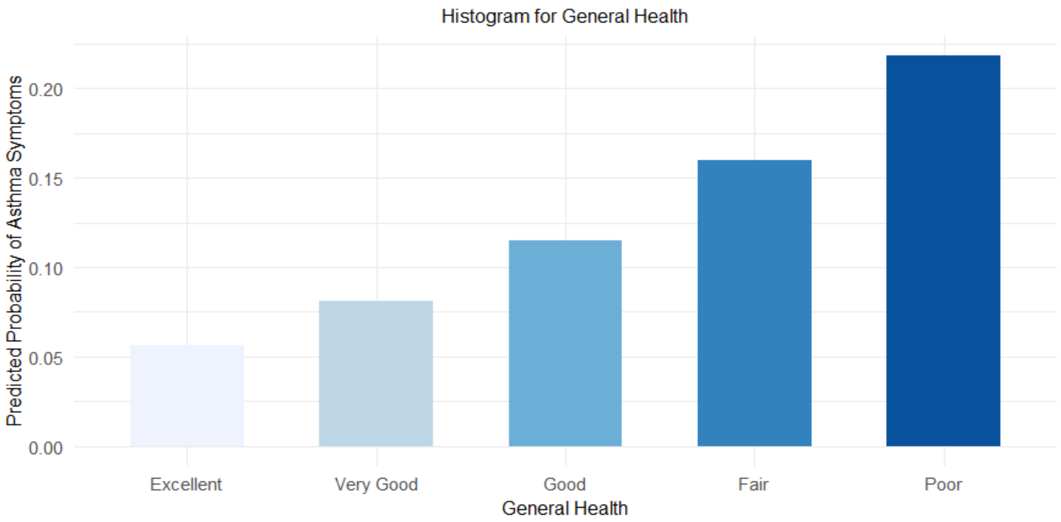

Figure 2.7 Histogram between General Health and Probability of Asthma

Figure 2.7 illustrates the trend of the association between self-reported general state of health and the estimated likelihood of asthma, showing a simple inverse trend. The highest likelihood of asthma would be encountered among the "Poor" health group, followed by the "Fair," "Good," and "Very Good/Excellent" health groups. This gradient suggests that declining overall health is strongly associated with an increased risk of asthma, likely due to the compounded effects of comorbidities, impaired respiratory status, or systemic inflammation associated with poorer health status.

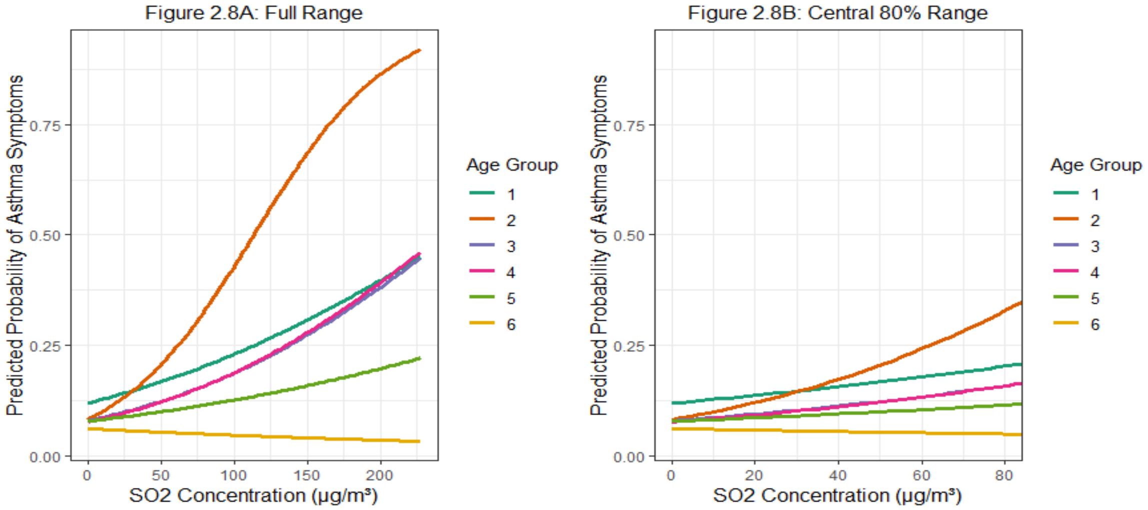

Figure 2.8 Interaction Dose-Response Curves of SO_2 and Age

Figure 2.8 is the dose-response curve of SO_2 level against predicted asthma probability, stratified by age groups. There are two groups with dramatically different patterns: Age Group 2 (25-34 years, orange line) exhibits the most significant gradient of increase in asthma probability as SO_2 levels increase. The relationship is strongly positive and nonlinear, with the predicted risk increasing sharply above 100 $\mu g/m^3$. This suggests that individuals in this age range are more susceptible to air pollution, possibly due to increased mobility, work-related exposure, or biological factors. Age Group 5 (55–64 years, green curve) shows a less sharp and more horizontal response. While there is a slight increase in asthma risk associated with higher SO_2 levels, the overall predicted risk remains comparatively low. This may reflect variations in lifestyle, reduced outdoor activity, or greater health strength in this old but not elderly group.

5. Discussion

5.1 Logistic Regression Model

The mean $PM_{2.5}$ concentration is also immediately linked to respiratory health. As the concentration of $PM_{2.5}$ increases, the prevalence of asthma increases. This may be due to the interaction of $PM_{2.5}$ with other elements in the air, creating new chemicals. For example, $PM_{2.5}$ combines with other volatile organic compounds to create ozone (O_3). Ozone is harmful to the human body when present in high concentrations. Toxic to the respiratory system, yet at low levels, it may be employed as a disinfectant and anti-inflammatory, and is of value in relieving symptoms of asthma. Although the SO_2 findings were not statistically significant, they also showed a tendency to increase the frequency of asthma symptom recurrence. Relative humidity tended to reduce the frequency of asthma symptom recurrence. Gender seems to have a significant impact on respiratory health, with women having higher rates than men. This is linked to the special system for women. Sex hormones are one of the important determinants of asthma development. All three anthropometry measures were significantly associated with respiratory health, including BMI, which suggests that body composition plays an important role in respiratory function and the risk of respiratory disease. The younger risk of respiratory illness in advanced age groups is contrary to what we usually observe in experiments. Our reasoning in this case is as follows: Firstly, compared to younger children and adults, the immune system is more complete and less susceptible to new asthma. As people grow old, they can be more conscious about dealing with and living with long-term illnesses. They can be more conscious about not coming in contact with asthma triggers and consuming medicines and other therapies beforehand. The increase in exercise volume will also significantly impact the overall prevention of asthma. Both the variables, either smoking or overall health, showed significant associations with respiratory health, pointing out the ill effects of smoking on respiratory function and the influence of overall health on respiratory outcomes. Older asthma patients are less susceptible to the effects of SO_2, which may be related to their lifestyle habits.

5.2 Fit Multinomial Logistic Regression Model

Comparisons between Categories 1 and 2 reveal various factors influencing the alleviation or exacerbation of asthma. Comparisons between Categories 1 and 3 predict the potential risk of new onset of asthma or respiratory symptoms in people previously thought not to have asthma, as well as factors influencing the diagnosis and presentation of asthma symptoms.

Category 1 and 2

Suppose the coefficient on a variable is positive in comparison between Category 1 (individuals with asthma) and Category 2 (individuals who had asthma in the past but do not have asthma now). In that case, it indicates that an increase in the variable is associated with an increase in the likelihood that an individual will transition from having asthma in the past but not currently (Category 2) to having asthma (Category 1). In other words, the larger the variable, the greater the likelihood of asthma remission or improvement. Conversely, a negative coefficient means that an increase in the variable is associated with a decrease in the likelihood that an individual will change from having persistent asthma (category 1) to having had asthma in the past (category 2). In this case, an increase in the variable may be associated with a decreased likelihood of asthma remission or improvement, suggesting that the variable may exacerbate asthma symptoms or impede recovery from asthma. Through the data, it was shown that both $PM_{2.5}$-lag and SO_2-lag contribute to the exacerbation of asthma symptoms; smoking frequency and excessive BMI are also potential aggravators of asthma; asthma becomes more severe with age.

Category 1 and 3

Suppose the coefficients on the variables are positive in the comparison between Category 1 (individuals who are suffering from asthma) and Category 3 (individuals without a history of asthma diagnosis). In that case, it suggests that higher coefficients are associated with an increased likelihood of reporting asthma symptoms in individuals without a history of asthma diagnosis (Category 3). In other words, higher variables may be associated with an increased risk of new-onset asthma or respiratory symptoms in individuals previously thought to be free of asthma. Conversely, a negative coefficient would imply that higher values of the variable are associated with a decreased likelihood of reporting asthma symptoms in people with no history of asthma diagnosis (Category 3). In this case, higher variables are associated with a lower likelihood of new onset of asthma or respiratory symptoms in people previously thought not to have asthma. The results show that the average concentrations of $PM_{2.5}$-lag and SO_2-lag have a positive contribution to the new-onset prevalence of asthma. Among weather factors, increases in relative humidity and temperature will have a negative effect on the new-onset asthma. Poorer general health status and higher frequency of smoking all contributed to more asthma cases, consistent with previous model results.

5.3 Health and Economic Costs of Air Pollution

Our study confirms a strong relationship between $PM_{2.5}$ and SO_2 levels (specifically, $PM_{2.5}$-lag and SO_2-lag) and respiratory disease status. Furthermore, rising pollution levels have a broad impact on asthma and other disease incidence. This finding is consistent with other similar findings in the environmental economics literature, which suggest that air pollution poses not only a public health concern but also inflicts immense economic damage, affecting labor productivity, medical expenditures, and socio-economic development (Chay & Greenstone, 2003; Graff et al., 2012). The research employed variation across counties in the US in air pollution levels during the 1981-1982 recession and found that with each 1% decrease in total suspended particulate matter (TSPs) concentration, the infant mortality rate decreased by approximately 0.35%.

The health costs of air pollution are incurred in two ways: higher expenditures on health and lower labor productivity. In areas with extreme air pollution, respiratory illnesses such as chronic bronchitis and asthma increase, leading to higher demand for hospital stays and increased health expenditure (Currie & Walker, 2011). Currie's work confirmed that declining air pollution was associated with increased newborn weight and overall improved health, leading to the conclusion that traffic pollution control is beneficial for infant health. Additionally, air pollution is exacerbating cardiovascular disease and reducing life expectancy (Chen et al., 2013). Chen's study analyzed the differences in air pollution resulting from China's Huai River Policy. It concluded that individuals residing in the northern part of China have a life expectancy about 5.5 years shorter than those in the southern part, primarily due to higher levels of particulate matter resulting from winter heating, which is mostly attributed to increased cardiorespiratory disease. Chay and Greenstone (2003) experimented with pollution changes that occurred during the 1981-1982 US recession and found that reductions in pollution severely reduced infant mortality, and that pollution control not only improves health but also reduces the economic costs of disease and death. Estimates showed that for every 1% reduction in TSPs, infant mortality was reduced by approximately 0.5%. The implication was that approximately 1,300 infant deaths in 1972 were avoided as a result of the act. Apart from healthcare expenditures, air pollution negatively impacts the economy through reduced labor productivity and a lower labor supply. Graff Zivin and Neidell (2012) examine the relationship between farm labor productivity and air pollution, finding that labor productivity falls by about 6% for each 10 $\mu g/m^3$ increase in $PM_{2.5}$ concentration.

5.4 Limitation

Our research on asthma prevalence might still have several limitations. The first is the information loss associated with binary dependent variables: Using binary variables (whether or not asthma attacks occur) simplifies the complexity of health conditions and fails to reflect the severity, frequency, or duration of attacks, thereby limiting in-depth understanding of health impacts. The other limitation may be the short lag setting of time series: Although lagged contamination variables are introduced, they are limited to short-term lags (5-day lag) and cannot fully capture long-term exposure effects. Measurement errors of self-reported health variables could also be a potential problem. Variables such as individual health, smoking status, and exercise status are derived from survey or self-reported data, which may introduce subjective biases and affect the accuracy of the estimates.

6. Conclusion

In this chapter, we examine the association between exposure to air pollution and the risk of asthma attacks, utilizing comprehensive individual-level health information and environmental data. The findings, using a logistic regression model, indicate that $PM_{2.5}$lag and SO_2lag concentrations have a statistically significant association with elevated asthma risk. In addition, the pollution effect is not uniform but also varies by personal characteristics, such as age, smoking status, and self-reported health. The presence of interactive terms suggests that vulnerable groups—most notably smokers and those with poor general health—may experience more severe impacts. However, these results are not statistically significant, possibly due to the sample size.

Several tests of robustness, including geographic and time-fixed effects, cluster-robust standard errors, and tests for multicollinearity, are used to determine the reliability of the findings. These results suggest the need for specific public health interventions, such as issuing pollution alerts or limiting outdoor activity among those most vulnerable. Further, the broader economic burden of asthma is such that pollution control has the potential for high long-term returns. Future studies must consider the longitudinal effects of exposure and examine causal identification methods to enhance the evidence base for public health policy interventions. Collectively, the findings highlight the contribution of air pollution to public health and underscore the need to identify and protect vulnerable individuals—such as young adults and smokers—at both medical and policy levels. Concrete steps may include targeted public warnings, personalized health screenings, or incorporating environmental risk metrics into health service planning and delivery.

Chapter 3 Air Pollution and Driving Behavior: A Study on the Impact of PM and Other Pollutants on Traffic Accident Risk

1. Introduction

1.1 Research Question and Hypotheses

The deleterious effects of air pollution on public health are extensively documented, with suspended particulate matter (PM) increasingly recognized for its significant role in the development of respiratory and cardiovascular diseases. However, the impact of air pollution on road safety remains an issue that necessitates urgent and comprehensive research. Recent studies have initiated an exploration of the relationship between air pollution levels and traffic accidents, suggesting a potential correlation between air quality and elevated accident rates. This article delves into this issue within the context of Colorado's diverse landscape and climate variability, aiming to elucidate the intricate relationship between PM, various meteorological parameters, and road accidents.

While a growing body of research has documented the health consequences of air pollution, its influence on traffic safety remains comparatively underexplored. Existing studies suggest that air pollution may impair driving performance through mechanisms such as reduced visibility and cognitive fatigue. However, there is limited empirical evidence on how pollution interacts with environmental factors and human behavior to influence traffic crash outcomes, particularly in diverse geographic and climatic settings such as Colorado.

This chapter seeks to investigate the multifaceted relationship between environmental conditions, human behavior, and road safety. Specifically, it addresses the following research questions:

H1: Do concentrations of particulate matter ($PM_{2.5}$) and meteorological conditions (e.g., temperature, humidity, wind speed) significantly affect the number of daily and hourly traffic accidents? (Variables: $PM_{2.5}$, Visibility, Temperature, Relative Humidity)

H2: How do environmental factors moderate the relationship between risky driving behaviors (e.g., DUI, speeding) and traffic crash incidence? (Interaction Variables: $PM_{2.5}$ × OS Dummy (One Over Speed), $PM_{2.5}$ × DUI Dummy (One DUI), Visibility × OS Dummy (One Over Speed), Visibility × DUI Dummy (One DUI), Temperature × OS Dummy (One Over Speed), Temperature × DUI Dummy (One DUI), Relative Humidity × Age Dummy (One<25), Relative Humidity × DUI Dummy (One DUI))

H3: Is there a systematic variation in crash frequency across different times of day and days of the week? (Variables: Week Dummy(Monday, Tuesday, Wednesday, Thursday, Saturday, Sunday), Time Dummy(Morning Peak, Day, Evening Peak, Night))

To accomplish this, the study utilized a robust dataset including detailed records of daily incidents and $PM_{2.5}$ concentrations from 2007 to 2020, as well as hourly incident data with $PM_{2.5}$ concentrations in the second part. This extensive dataset facilitates a thorough investigation into how fluctuations in air quality interact with various weather conditions. The research endeavors to identify and quantify the complex interactions between these environmental factors and traffic accident rates, utilizing advanced statistical analyses and modeling techniques. The primary objective of this study is to furnish a comprehensive understanding of how particulate matter and weather conditions contribute to traffic accidents. This research aspires to advance traffic safety and mitigate the adverse impacts of air pollution in Colorado and other regions confronted with similar environmental challenges.

1.2 Significance of the Research Question

Understanding the link between air pollution and traffic safety is crucial for public health and transportation policy. The health effects of particulate matter (PM) are well-documented, particularly in relation to respiratory and cardiovascular diseases. However, its impact on road safety is less studied. This gap is significant because pollution can lead to poor visibility, mental fatigue, and breathing problems, which in turn may increase the chances of driving mistakes and accidents. This study addresses that gap by examining the environmental, behavioral, and timing factors associated with crash risk. It examines how air pollutants ($PM_{2.5}$) directly affect traffic accidents as well as how they interact with weather conditions and risky driving behaviors like DUI and speeding. This approach provides a more precise and more complete view of how external factors exacerbate human errors on the road. The policy implications are extensive. If pollution and severe weather are proven to increase accident rates, this would support targeted traffic safety measures and pollution alerts, especially during hazardous conditions. Additionally, these findings could help shape traffic rules, urban planning (such as creating better air quality zones), and even behavioral programs, like discouraging unnecessary travel during high-pollution periods or tightening penalties for risky driving in adverse conditions. Colorado is a particularly useful case for this analysis due to its mix of mountainous terrain, seasonal weather changes, and high pollution levels in urban areas. The results of this study not only benefit local policies but may also provide valuable insights for other states and countries facing similar environmental issues. Ultimately, this research contributes to a growing body of work that connects environmental economics, transportation safety, and public health. It highlights the need for policy responses that consider the intertwined nature of pollution-related effects, thereby broadening the discussion to include immediate behavioral impacts, such as traffic accidents.

1.3 Literature Review and Novelty

The paper (Sager, L. 2019) estimates the change in the number of traffic accidents in the UK between 2009 and 2014 as a result of changes in air pollution. The authors concluded that pollutants initially lead to a decrease in drivers' cognitive abilities, which in turn increases their aggressiveness and decreases their patience, thereby significantly increasing the incidence of traffic accidents. By measuring $PM_{2.5}$ concentrations and the incidence of traffic accidents in 153 areas, the data results concluded that for every 1 μg/m3 increase in average $PM_{2.5}$, there is a 0.4% increase in the incidence of accidents. Regarding the endogeneity problem, the authors indirectly observed the change of air pollutant concentration in the first phase through the phenomenon of atmospheric temperature reversal, which is a natural manifestation and thus reduces the occurrence of the endogeneity problem. In the first stage of the experiment, the authors used the intensity of the nighttime temperature inversion measure as an instrumental variable to obtain the daily $PM_{2.5}$ concentration. The results showed an average increase in $PM_{2.5}$ concentration of 14% for each level of increase in nighttime temperature difference. In the second phase of the experiment, the natural logarithm of the number of vehicles involved in accidents at a certain time of the day in a certain region was used as the subject of the study, and the hypothesis that the accident rate increases by 0.4% for every 1 μg/m3 increase in $PM_{2.5}$ was further verified through the effect of $PM_{2.5}$ on the number of vehicles involved in accidents. Robust tests are also conducted to ensure the validity of the results. In addition to the primary findings, the study provides a detailed analysis of the atmospheric temperature data, air pollution data, and additional weather data used in the research. The document also discusses the potential impact of air pollution on worker productivity, cognitive performance, and health, drawing upon the findings. The role of reducing air pollution levels in reducing the cost of accident prevention is suggested. Our studies differ in several ways. It focuses on Colorado, USA, and uses data from 2007 to 2020. This gives a broader and region-specific perspective. We also examine how these factors collectively influence driver behavior and accidents. Sager focused only on cognitive decline and aggression. We investigate whether air pollution directly influences behaviors like speeding or sudden braking. While Sager uses instrumental variables, we control for day-of-the-week and other relevant factors. This allows for a more direct understanding of the impact of air pollution on traffic accidents.

Another article (American Economic Journal: Applied Economics Vol. 8, NO. 2, April 2016) focuses on the effects of Daylight Saving Time (DST) on different aspects of human life, including sleep patterns, vehicle accidents, and related health issues. Authors employ the regression discontinuity method to analyze the effects of DST. The study examines the effects of DST on sleep deprivation, fatal vehicle crashes, and the incidence of acute myocardial infarction. Vehicle fatality data are drawn from the Fatality Analysis Reporting System (FARS), a data repository built and maintained by the National Highway Traffic and Safety Administration. Authors also take advantage of vehicle miles traveled (VMT) data provided by Caltrans' Performance Measurement System (PeMS) to better track the effects that VMT adjustments may have on the results. Authors use a regression discontinuity design to measure discontinuities that occur around the time of the policy transition, in an effort to estimate the direct impact of the policy. Authors comment on the potential confounding parameters, ranging from variation in local weather conditions that occur during the end of summer and the transition to pre-winter or winter daylight saving, as well as the different geographic features that may have been beneficial to one highway or another. They also employ a fixed-effects model to take advantage of the variation in DST allocations. The effect of Daylight-Saving Time (DST) on fatal motor vehicle crashes is illustrated using a regression discontinuity (RD) design. This policy research design is well-suited for measuring the immediate effect of a discrete policy transition like the shift from Standard Time (ST) to DST in the US. Using differences in RD designs, the short-term spike in frequency allows for a consistent estimate of the effect of a policy change like DST on fatal crash risk by exploiting the discontinuous change. To control for other factors that affect fatal crash risk that might have trended besides DST to contribute to the spike in crashes around the policy transition, it is necessary for the treated and untreated number of fatal car crashes to vary continuously with the date around the RD offset or transition point. As a whole, this document presents a view that is broad in its literature review and research methods, as it encompasses the entirety of human life, both present and future, as it pertains to daylight saving time. The limitations of knowledge in this area are quite tragic from a scientific perspective. While a study in the American Economic Journal: Applied Economics (2016) examines the consequences of Daylight Saving Time (DST) on fatal automobile crashes, sleep loss, and diseases using a regression discontinuity method, our research differs. Instead of comparing the effect of a policy intervention like DST, our research examines the long-term impact of air pollution, specifically $PM_{2.5}$, on automobile accidents in Colorado. By emphasizing the extent of pollution and its link to traffic crashes over a period (2007-2020), we hope to highlight the gradual, cumulative impact of air pollution on road safety, rather than the immediate, short-term effect of a single policy change. Compared to the DST study, which examines only fatal vehicle crashes, our study examines a broader range of traffic crashes, including additional variables such as driver conduct, weather conditions, and the kinds of pollutants. Furthermore, while the DST study employs regression discontinuity and fixed-effects estimates, we incorporate day-of-week fixed effects, weather, and potential lags to examine the impact of pollution. These differences in scope, organization, and detail make our study feasible for gathering a better insight into how environmental factors, such as air pollution, affect road safety over a span of years.

The paper (Deng, Yongheng and He, Jia and Shen, Xixi and Li, Bingqing, Does Air Pollution Cause More Car Accidents? Evidence from Auto Insurance Claims) discusses the impact of air pollution on road safety and, consequently, the impact on insurance claims related to traffic accidents. The authors investigate the relationship between the Air Quality Index (AQI) and various weather conditions, with the number of car accidents and the severity of car accidents through statistical analysis and regression modeling. The study's results indicate a positive correlation between air pollution and the number of accidents, while a negative correlation exists between air pollution and the frequency of insurance claims. Another conclusion is that air pollution can influence the occurrence of risky behaviors, and drivers may be more cautious when driving in heavily polluted weather, thus reducing the severity of accidents. It also suggests that drivers may react differently to air pollution depending on demographic factors and seasonal changes. The main econometric framework involves the use of fixed effects modeling and instrumental variables (IV) methods to address potential endogeneity. Fixed effects models are used to estimate the impact of air pollution on traffic crashes, while IV methods are used to mitigate potential heterogeneity and endogeneity problems. The paper also discusses the use of weather control variables, atmospheric temperature inversion, and wind variables as IVs to ensure model validity and exogeneity. In addition, the paper presents detailed results of the estimated model, including the impact of air pollution on the number of accidents and the number of claims, as well as the interaction between air pollution and risk preference measures. Whereas the study by Deng et al. (2021) focuses on car accidents and air pollution through insurance claims, mine differs in the sense that I examine the immediate impact of $PM_{2.5}$ on the frequency of occurrence of traffic accidents in Colorado over an extended period of time (2007-2020). Unlike handling insurance claims, we utilize real crash data, including the frequency and the severity of accidents. In comparison to their use of AQI, our study directly measures $PM_{2.5}$ concentrations and conditions in relation to weather factors like temperature and humidity, aiming to better capture the subtle effects of pollution on road safety. Additionally, while Deng et al. employ fixed effects and IV approaches, we use day-of-the-week fixed effects, condition on lag effects of pollution, and analyze different types of accidents, rather than just insurance claim-related accidents.

Another research paper (Joo, S., Oh, C., Lee, S., & Lee, G., 2017) investigates the impact of traffic accidents on air quality near roads, explicitly focusing on the dispersion of air pollutants such as nitrogen oxides (NO_x) and particulate matter ($PM_{2.5}$). The study utilized the Motor Vehicle Emissions Simulator (MOVES) to calculate highway emissions during both accident and non-accident traffic scenarios. Additionally, the California Puff Model (CALPUFF) was employed to analyze the dispersal of air pollutants resulting from accidents. Regression analysis was used to identify factors influencing the severity of vehicle emissions and discuss strategies for improving public health related to transportation. The analytical framework includes data collection, emission estimation, dispersion analysis, and the identification of factors impacting air pollution severity. The study identified several factors that affect the severity of air pollution exposure, including traffic volume, weather conditions during accidents, accident severity, NOx emissions, and pre-accident travel speeds. The research emphasized the importance of maintaining appropriate distances between highways and densely populated residential areas to reduce respiratory illnesses caused by air pollution. Additionally, the study highlighted the significance of implementing traffic management strategies to decrease vehicle emissions. The vehicle emissions model estimated highway emissions for both accident and non-accident traffic conditions, taking into account factors like vehicle type, fuel type, and others to estimate various air pollutants. The air dispersion model simulated continuous pollutant releases from sources, considering meteorological conditions such as temperature, wind speed, and wind direction to calculate downwind ambient concentrations of air pollutants emitted by sources like industrial facilities and vehicular traffic. Specifically designed for long-range air pollution transport applications, especially for distances exceeding 50 km, the model was used to identify areas affected by air pollution and to compare estimated emissions and affected areas under accident and non-accident traffic conditions. While the study by Joo et al. (2017) focuses on how traffic accidents affect air quality by examining pollutant dispersion and the severity of emissions, my research differs. Rather than focusing on the impact of accidents on air pollution, we examine the direct effect of air pollution, specifically $PM_{2.5}$, on traffic accidents in Colorado over a more extended period (2007-2020). Our research links pollution levels to the frequency and intensity of crashes, rather than emissions and dispersion. We apply crash data, as opposed to using Joo et al.'s utilization of emission and dispersion models like MOVES and CALPUFF, and incorporating weather variables as a means of assessing driver behavior and accident risk influenced by pollution. These differences in scope, methodology, and data allow our study to provide a more timely understanding of the relationship between environmental factors and road safety.

One paper (Wang, T., Wang, Y., & Cui, N., 2022) investigates the association between air pollution and driver traffic violations in Wuhan, China. Using traffic violation records and air quality data, the study examines how air pollution affects various types of traffic violations and its impact on drivers' mental state and safety behavior. Using a diverse cross-sectional dataset comprising traffic violation reports and air pollution measurements, including data on violations, driver profiles, vehicle attributes, and pollution levels, the authors employed statistical techniques such as regression modeling and validity assessment to analyze the effects of $PM_{2.5}$ levels on traffic violations under different weather conditions. The study used Wuhan automobile driving data from January to December 2018, using data from three main sources: traffic violation data from the Wuhan Traffic Management Bureau, air pollution data from the Wuhan Environmental Protection Bureau, and additional weather data from the Wuhan Meteorological Bureau, with a focus on the effect of air pollution (especially $PM_{2.5}$) on the severity of traffic violations. The analysis employed ordered logistic regression models to estimate the effect of air pollution on the severity of traffic violations. The study aims to determine the potential impact of air pollution on traffic violation severity and provides robustness checks to ensure the accuracy of the results. In addition, the study reveals the impact of air pollution on individual mental health. The results of the analysis showed that increased daily concentrations of $PM_{2.5}$ were positively associated with the severity of traffic violations associated with inexperience but negatively associated with the severity of traffic violations associated with overconfidence. In addition, gender was found to influence the effect of air pollution on the severity of traffic violations, with females being more likely to commit serious traffic violations associated with inexperience when faced with poor air quality. In conclusion, this paper draws insights from extensive traffic records and air quality data from Wuhan, China. It provides a comprehensive analysis of how air pollution affects traffic violations, which enhances our understanding of the interactions between air pollution and driver safety performance, as well as the effects of air pollution on drivers' psychological states. While Wang et al. (2022) aims at the influence of air pollution, i.e., $PM_{2.5}$, on traffic violation and driving patterns in Wuhan, our study is broader in the sense that it investigates the direct effect of $PM_{2.5}$ levels on traffic accidents in Colorado over a more extended period of time (2007-2020). While Wang et al. focus on offenses caused by drivers' overconfidence or inexperience, our work focuses on real traffic accidents and their severity, as well as on weather conditions and different types of crashes. Our model also has day-of-week fixed effects and lagged air pollution effect controls, adding a temporal component to the analysis that is not addressed in Wang et al. Our examination of actual accidents, rather than violations, is a simpler measure of how pollution affects traffic safety in an area over time.

This chapter adds to the existing research on how environmental factors affect road safety by presenting several important innovations in research focus, methods, and analysis. Previous studies, such as Sager (2019), have shown that higher $PM_{2.5}$ levels are associated with lower cognitive function and increased accident rates in the UK. However, these studies mainly focused on cognitive aspects and used instrumental variables to address specific biases. Our research builds on this work by utilizing U.S. microdata from Colorado, collected between 2007 and 2020, which allows for a longer-term and geographically diverse analysis. Additionally, while Sager examined only the effects of $PM_{2.5}$, our study includes extra weather controls, such as temperature, humidity, and wind speed. This approach gives a richer understanding of how air quality and weather conditions interact to influence crash outcomes.

Moreover, Wang et al. (2022) examine how air pollution affects traffic violations in Wuhan, with a focus on driver overconfidence or inexperience. In contrast, our study analyzes traffic crash data instead of violation reports and covers a longer timeframe, which provides a stronger basis for causal conclusions. We use time and location fixed effects, as well as lagged exposure controls, to enhance the analysis. This method not only examines immediate effects but also captures the delayed impacts of pollution, leading to a better understanding of behavior over time.

Finally, this study employs a detailed temporal structure, utilizing hourly and daily data, and controls for fixed effects in both time and space, thereby enhancing the reliability of causal inferences. By integrating specific environmental variables and utilizing models that examine interaction effects between risky driving behaviors and pollution levels, this research extends beyond simple relationships. It offers more profound insights into how pollution conditions influence traffic risk. These innovations establish a nuanced framework that can inform policy, particularly in areas with variable environmental conditions, such as Colorado.

2. Data

The data for this study were meticulously gathered from the Colorado Department of Transportation (CDOT) and the United States Environmental Protection Agency (EPA) to comprehensively analyze the relationship between air pollution, meteorological conditions, and road safety in Colorado from 2007 to 2020. The meticulous collection, integration, and preprocessing of data from the CDOT and EPA provided a robust dataset for analyzing the interplay between air quality, meteorological conditions, and traffic accidents in Colorado. This comprehensive approach enabled a detailed examination of the factors influencing road safety, with the potential to inform the development of targeted policies and interventions aimed at enhancing both air quality and traffic safety.

2.1 Data Sources

The primary data source for traffic crashes was the Colorado Department of Transportation's official records. This database provided detailed information on each crash, including the exact time (month, day, and hour), location (highway and specific site), and prevailing weather conditions at the time of the incident. Additional details included vehicle type, direction of travel, speed, driver demographics (age and gender), maximum speed limits at the crash site, presence of hazardous materials, and whether the driver was under the influence of alcohol.

Complementing this, environmental data was sourced from the United States Environmental Protection Agency, encompassing daily records of $PM_{2.5}$ levels, visibility, temperature, relative humidity, and other meteorological parameters that potentially affect driving safety.

2.2 Data Preprocessing

To facilitate a robust analysis, data from the Colorado Transportation Department and the EPA were meticulously integrated through a series of critical steps. First, the datasets were filtered to retain only the variables pertinent to the analysis, ensuring a focused investigation into the interplay between environmental factors and traffic incidents, including vehicle speed, drivers' age, sex, and whether they had drunk or not before driving. Next, crash data from the Colorado Transportation Department was merged with environmental data from the EPA, including daily $PM_{2.5}$ levels, daily temperatures, visibility, and relative humidity. This was achieved by utilizing time-specific (date and hour) and location-specific (state and county codes) identifiers to precisely align traffic incidents with corresponding environmental conditions. In the previous step, since the Colorado Transportation Department data only included county names, we converted all county names into corresponding county codes before merging them with the daily weather data.

For sex information, we transferred females into 0 and Males into 1. If DUI information shows Y (yes), we changed it into 1, otherwise, 0. For the overseeding information, first of all, we compared the speed of the vehicle with the speed limit information of the road section, if the speed of the vehicle in question exceeds the speed limit, it will be labeled as 1, otherwise, it labels as 0.

Another problem in the data that needs to be solved is that there are usually two vehicles involved in a crash, and in order to be able to visualize the effects of weather and air pollution on a crash, we need to integrate the information from the two vehicles involved in the crash. The first is the gender of the driver, if both drivers are female, then create a new 'SEX' variable and label it as 2, if there is only one female driver, then label 'SEX' as 1, if both drivers are male, then label 'SEX' as 0. The same as AGE, if both drivers are under 25 years old, then set it as 2, if only one of them is under 25, set it as 1, otherwise, set it as 0. Similarly, for speeding or not, if both vehicles in the accident are speeding, then create a new variable 'Over_Speed' and label it as 2, if there is only one vehicle speeding, then label it as 2. If only one vehicle exceeds the speed limit, it is labeled as 1, otherwise it is 0. For drunk driving, a new variable 'DUI' is created, and if both drivers have been drinking, it is labeled as 2, if only one of them has been drinking, it is labeled as 1, and if neither of them has been drinking, it is labeled 0.

The data was then aggregated by counting the number of accidents occurring at the same location on the same day, and categorical variables such as alcohol influence were converted into numerical formats to facilitate statistical modeling. To maintain the robustness of the subsequent analysis, data cleaning involved the removal of entries with missing values (NA) across all variables, ensuring the integrity of the dataset.

And in the second part of the study, we replaced the daily data in the first part with hourly crash data, while collecting hourly weather data, including temperature, wind speed, and relative humidity. The data processing is similar to the first part, the difference is that we added two variables Day Of Week and Time Of Day, the former counts what day of the week (weekday or weekend) the car accident happened, the latter counts the specific time period when the car accident happened (whether it is the peak commuting time), we divided each different time period of the day according to the hour in the car accident information. Also, the count of the number of crashes shifted from each county per day to each county per hour. This part of the study is more helpful for me to analyze the immediate effect of time on the number of crashes due to the more detailed division of time, and the study focuses on analyzing the distribution of crashes during different time periods (e.g., morning rush hour, evening rush hour) while including the effects of driver-specific behaviors and weather conditions.

2.3 Summary Statistics

Table 3.1 Summary statistics of daily number of car crashes

Variable (Daily)	**Obs.**	**Mean**	**Max**	**Min**	**SD**
Number of Crash (per County per Day)	31787	19.091	77	1	13.013
Mean $PM_{2.5}$ ($\mu g/m^3$)	31787	6.160	98.249	0	4.910
Mean Visibility (statute miles)	31787	9.516	50	0.92	2.415
Mean Temperature (°F)	31787	48.735	87.417	-7.5278	17.406
Mean Relative Humidity (%)	31787	50.318	100	4.7	18.637
Mean Wind Speed (mph)	31787	3.758	23.341	0.533	1.727
Age Dummy (One<25)	31787	0.378	1	0	0.221
Age Dummy (Both<25)	31787	0.0746	1	0	0.118
Sex Dummy (One Female)	31787	0.470	1	0	0.225
Sex Dummy (Both Female)	31787	0.209	1	0	0.184
OS Dummy (One Over Speed)	31787	0.063	1	0	0.116
OS Dummy (Both Over Speed)	31787	0.006	1	0	0.036

DUI Dummy (One DUI)	31787	0.032	1	0	0.085
DUI Dummy (Both DUI)	31787	0.967	1	0	0.085

Note: The table above lists the means, standard deviations, minimum and maximum values, and counts of observations for all the variables used in the analysis of the linear model for car crashes number happened in Colorado per day, all data are aggregated based on daily data.

Table 3.2 Summary statistics of the hourly number of car crashes

Variable	**Obs.**	**Mean**	**Max**	**Min**	**SD**
Number of Crash (per County per Hour)	128481	2.227	15	1	1.576
Mean $PM_{2.5}$ ($\mu g/m^3$)	128481	5.879	88.85	0	5.320
Mean Visibility (statute miles)	128481	9.112	10	0	2.264
Mean Temperature (°F)	128481	54.869	101	-13	20.785
Mean Relative Humidity (%)	128481	45.414	100	4	24.715
Mean Wind Speed (mph)	128481	3.998	27.8	0	2.595
Age Dummy (One<25)	128481	0.380	1	0	0.423
Age Dummy (Both<25)	128481	0.071	1	0	0.226
Sex Dummy (One Female)	128481	0.467	1	0	0.434
Sex Dummy (Both Female)	128481	0.212	1	0	0.355
OS Dummy (One Over Speed)	128481	0.058	1	0	0.208
OS Dummy (Both Over Speed)	128481	0.005	1	0	0.063
DUI Dummy (One DUI)	128481	0.029	1	0	0.153
DUI Dummy (Both DUI)	128481	0.971	1	0	0.154

Note: The table above lists the means, standard deviations, minimum and maximum values, and counts of observations for all the variables used in the analysis of the linear model for car crashes happened in Colorado per hour, all data are aggregated based on hourly data.

Table 3.1 presents summary statistics for the daily number of car crashes aggregated by the county in Colorado, covering 31,787 observations. The average daily crash count per county is approximately 19.1, ranging from 1 to 77 crashes. Environmental variables include mean daily $PM_{2.5}$ concentration (mean = 6.16 $\mu g/m^3$), visibility (mean = 9.52 miles), temperature (mean = 48.74°F), relative humidity (mean = 50.3%), and wind speed (mean = 3.76 mph). Several demographic and behavioral dummy variables capture driver characteristics, including whether one or both drivers are under 25 years old, gender combinations (one or both female), speeding behavior (one or both drivers over speeding), and DUI status (one or both drivers driving under the influence). These variables provide a detailed context for analyzing factors associated with daily crash counts.

Table 3.2 summarizes the hourly-level data on car crashes aggregated by county, with a total of 128,481 observations. The average number of crashes per county per hour is 2.23, with a maximum of 15 crashes in one hour. Environmental conditions recorded hourly include $PM_{2.5}$ concentration (mean = 5.88 µg/m^3), visibility (mean = 9.11 miles), temperature (mean = 54.87°F), relative humidity (mean = 45.41%), and wind speed (mean = 4.00 mph). Dummy variables indicate driver age groups (one or both drivers under 25), gender combinations, speeding behavior, and DUI status. This large, high-frequency dataset facilitates a detailed examination of temporal variations in crash frequency and their associations with environmental and driver-related factors.

3. Empirical Framework

The initial choice for the analysis was a linear regression model due to its simplicity and ease of interpretation. However, the linear model's applicability was limited by the presence of heteroskedasticity and linearity in the data, which violates the assumptions underpinning linear regression. Specifically, the variance of the residuals was not constant, rendering the linear model unsuitable for accurately capturing the relationships in the data.

To address the limitations of the linear model, a Poisson regression model was employed. The Poisson model is particularly appropriate for counting data and ensures that the predicted values remain non-negative, which aligns well with the nature of traffic accident counts. The Poisson regression model can be described by the following equation:

While the Poisson model provided a structured approach to analyzing the data, it was found to be inadequate due to the presence of over-dispersion. Over-dispersion occurs when the variance of the data significantly exceeds the mean, violating the assumptions of the Poisson model, which assumes that the mean and variance are equal.

To effectively manage the over-dispersion issue, the analysis was transformed using a negative binomial regression model. The negative binomial model extends the Poisson model by introducing an additional parameter to account for the extra variance, thus accommodating over-dispersion in the data. The model can be expressed as:

The negative binomial regression model was chosen due to its flexibility in handling data where the variance exceeds the mean, making it a robust choice for this analysis. This model enabled changes in the explanatory variables to appropriately affect the variance of the response variable, thereby providing a better fit for the data and yielding more reliable insights into the relationship between air pollution, meteorological conditions, and traffic accidents.

The selection of the negative binomial regression model represented a critical step in addressing the methodological challenges posed by the data's characteristics. By accommodating over-dispersion, the negative binomial model provided a more accurate and reliable framework for analyzing the complex interactions between environmental factors and traffic accident rates. This approach ensured that the study's findings were robust and reflective of the actual underlying patterns in the data, thereby contributing valuable insights to inform policy and intervention strategies aimed at improving both air quality and road safety.

3.1 Daily Car Crash Regression Model

$$\mathrm{Crash}_{jt} = a_0 + a_j + \sum_{k=1}^{5} \beta_k X_{jt}^{(k)} + \sum_{l=1}^{8} \gamma_l Z_{ijt}^{(l)} + \sum_{m=1}^{6} \delta_m \left(X_{jt}^{(k)} \cdot Z_{ijt}^{(l)} \right) + \varepsilon_{ijt}$$

$$\begin{aligned}
&\mathrm{Crash}_{jt} = \alpha_0 + \alpha_j + \beta_1 \mathrm{Mean\ PM}_{2.5_{jt}} + \beta_2 \mathrm{Mean\ Visibility}_{jt} + \beta_3 \mathrm{Mean\ Temp}_{jt} + \\
&\beta_4 \mathrm{Mean\ Relative\ Humidity}_{jt} + \beta_5 \mathrm{Wind\ Speed}_{jt} + \gamma_1 \mathrm{MeanAgeDummy}_{ijt}^{\mathrm{One}<25} + \\
&\gamma_2 \mathrm{MeanAgeDummy}_{ijt}^{\mathrm{Both}<25} + \gamma_3 \mathrm{MeanOverSpeed}_{ijt}^{\mathrm{One\ OS}} + \gamma_4 \mathrm{MeanOverSpeed}_{ijt}^{\mathrm{Both\ OS}} + \\
&\gamma_5 \mathrm{MeanDUI}_{ijt}^{\mathrm{One\ DUI}} + \gamma_6 \mathrm{MeanDUI}_{ijt}^{\mathrm{Both\ DUI}} + \gamma_7 \mathrm{MeanSexDummy}_{ijt}^{\mathrm{Mixed}} +
\end{aligned}$$

$$\gamma_8 \mathrm{MeanSexDummy}_{\mathrm{ijt}}^{\mathrm{Both\ Female}} + \delta_1 \left(\mathrm{Mean\ PM}_{2.5_{\mathrm{jt}}} \cdot \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}}\right) + \delta_2 \left(\mathrm{Mean\ PM}_{2.5_{\mathrm{jt}}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) \delta_3\left(\mathrm{Mean\ Visibility}_{\mathrm{jt}} \cdot \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}}\right) + \delta_4\left(\mathrm{Mean\ Visibility}_{\mathrm{jt}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) + \delta_5\left(\mathrm{Mean\ Temp}_{\mathrm{jt}} \cdot \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}}\right) + \delta_6\left(\mathrm{Mean\ Temp}_{\mathrm{jt}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) + \delta_7\left(\mathrm{Mean\ Relative\ Humidity}_{\mathrm{jt}} \cdot \mathrm{MeanAgeDummy}_{\mathrm{ijt}}^{\mathrm{One}<25}\right) + \delta_8\left(\mathrm{Mean\ Relative\ Humidity}_{\mathrm{jt}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) + \varepsilon_{\mathrm{ijt}} \quad (1)$$

i: It represents individual drivers involved in the crash.
j: This index represents a geographic or spatial unit. In this context, it refers to a specific county in Colorado where the crashes occurred.
t: It denotes the temporal dimension of the analysis. Here, it represents the exact date when the crash occurred.
$X_{jt}^{(k)}$: is a set of environmental variables for county j at time t, where k=1,…,5. These include (Mean $PM_{2.5jt}$, Mean $Visibility_{jt}$, Mean $Temp_{jt}$, Mean Relative $Humidity_{jt}$, Wind $Speed_{jt}$)
$Z_{ijt}^{(l)}$: denotes individual-level characteristics for person i at county j and time t, where l=1,…,8. These include ($MeanAgeDummy_{ijt}^{One<25}$, $MeanAgeDummy_{ijt}^{Both<25}$, $MeanOverSpeed_{ijt}^{One\ OS}$, $MeanOverSpeed_{ijt}^{Both\ OS}$, $MeanDUI_{ijt}^{One\ DUI}$, $MeanDUI_{ijt}^{Both\ DUI}$, $MeanSexDummy_{ijt}^{Mixed}$, $MeanSexDummy_{ij}^{tBoth\ Female}$)
Interaction terms: To explore how environmental effects differ across individual characteristics, the model includes interaction terms between each environmental variable in $X_{jt}^{(k)}$ and key personal attributes in $Z_{ijt}^{(l)}$.

3.2 Hourly Car Crash Regression Model

$$\mathrm{Crash}_{\mathrm{jt}} = \mathrm{a}_0 + \mathrm{a}_{\mathrm{j}} + \sum_{\mathrm{k}=1}^{5} \beta_{\mathrm{k}} \mathrm{X}_{\mathrm{jt}}^{(\mathrm{k})} + \sum_{\mathrm{l}=1}^{8} \gamma_{\mathrm{l}} \mathrm{Z}_{\mathrm{ijt}}^{(\mathrm{l})} + \sum_{\mathrm{m}=1}^{6} \delta_{\mathrm{m}} \left(\mathrm{X}_{\mathrm{jt}}^{(\mathrm{k})} \cdot \mathrm{Z}_{\mathrm{ijt}}^{(\mathrm{l})}\right) + \mathrm{DOW}_{\mathrm{t}} + \mathrm{TOD}_{\mathrm{t}} + \varepsilon_{\mathrm{ijt}}$$

$$\mathrm{Crash}_{\mathrm{jt}} = \alpha_0 + \alpha_{\mathrm{j}} + \beta_1 \mathrm{Mean\ PM}_{2.5_{\mathrm{jt}}} + \beta_2 \mathrm{Mean\ Visibility}_{\mathrm{jt}} + \beta_3 \mathrm{Mean\ Temp}_{\mathrm{jt}} + \beta_4 \mathrm{Mean\ Relative\ Humidity}_{\mathrm{jt}} + \beta_5 \mathrm{Wind\ Speed}_{\mathrm{jt}} + \gamma_1 \mathrm{MeanAgeDummy}_{\mathrm{ijt}}^{\mathrm{One}<25} + \gamma_2 \mathrm{MeanAgeDummy}_{\mathrm{ijt}}^{\mathrm{Both}<25} + \gamma_3 \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}} + \gamma_4 \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{Both\ OS}} + \gamma_5 \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}} + \gamma_6 \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{Both\ DUI}} + \gamma_7 \mathrm{MeanSexDummy}_{\mathrm{ijt}}^{\mathrm{Mixed}} + \gamma_8 \mathrm{MeanSexDummy}_{\mathrm{ijt}}^{\mathrm{Both\ Female}} + \delta_1 \left(\mathrm{Mean\ PM}_{2.5_{\mathrm{jt}}} \cdot \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}}\right) + \delta_2 \left(\mathrm{Mean\ PM}_{2.5_{\mathrm{jt}}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) \delta_3\left(\mathrm{Mean\ Visibility}_{\mathrm{jt}} \cdot \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}}\right) + \delta_4\left(\mathrm{Mean\ Visibility}_{\mathrm{jt}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) + \delta_5\left(\mathrm{Mean\ Temp}_{\mathrm{jt}} \cdot \mathrm{MeanOverSpeed}_{\mathrm{ijt}}^{\mathrm{One\ OS}}\right) + \delta_6\left(\mathrm{Mean\ Temp}_{\mathrm{jt}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) + \delta_7\left(\mathrm{Mean\ Relative\ Humidity}_{\mathrm{jt}} \cdot \mathrm{MeanAgeDummy}_{\mathrm{ijt}}^{\mathrm{One}<25}\right) + \delta_8\left(\mathrm{Mean\ Relative\ Humidity}_{\mathrm{jt}} \cdot \mathrm{MeanDUI}_{\mathrm{ijt}}^{\mathrm{One\ DUI}}\right) + \mathrm{DOW}_{\mathrm{t}} + \mathrm{TOD}_{\mathrm{t}} + \varepsilon_{\mathrm{ijt}} \quad (2)$$

i: It represents individual drivers involved in the crash.
j: This index represents a geographic or spatial unit. In this context, it refers to a specific county in Colorado where the crashes occurred.
t: It denotes the temporal dimension of the analysis. Here, it represents the exact time in a day when the crash occurred.
$X_{jt}^{(k)}$: is a set of environmental variables for county j at time t, where k=1,…,5. These include (Mean $PM_{2.5jt}$, Mean $Visibility_{jt}$, Mean $Temp_{jt}$, Mean Relative $Humidity_{jt}$, Wind $Speed_{jt}$)
$Z_{ijt}^{(l)}$: denotes individual-level characteristics for person i at county j and time t, where l=1,…,8. These include ($MeanAgeDummy_{ijt}^{One<25}$, $MeanAgeDummy_{ijt}^{Both<25}$, $MeanOverSpeed_{ijt}^{One\ OS}$, $MeanOverSpeed_{ijt}^{Both\ OS}$, $MeanDUI_{ijt}^{One\ DUI}$, $MeanDUI_{ijt}^{Both\ DUI}$, $MeanSexDummy_{ijt}^{Mixed}$, $MeanSexDummy_{ij}^{tBoth\ Female}$)

Interaction terms: To explore how environmental effects differ across individual characteristics, the model includes interaction terms between each environmental variable in $X_{jt}^{(k)}$ and key personal attributes in $Z_{ijt}^{(l)}$.

Variable Definitions:

$Crash_{jt}$: Number of traffic crashes in county j at time t;

α_j: County fixed effects

$X_{jt}^{(k)}$: Environmental variables:

$MeanPM_{2.5jt}$, $MeanVisibility_{jt}$, $MeanTemp_{jt}$, $MeanRelativeHumidity_{jt}$: Air pollution and weather measures including visibility, temperature, relative humidity and wind speed in county j at time t;

$Z_{ijt}^{(l)}$: Driving behavior dummy variables (8 in total, two for each category):

Age Dummies (Control group: Neither driver <25 years old):

$MeanAgeDummy_{ijt}^{One<25}$: One driver <25 years old

$MeanAgeDummy_{ijt}^{Both<25}$: Both drivers <25 years old

Sex Dummies (Control group: Both male drivers):

$MeanSexDummy_{ijt}^{Mixed}$: One driver is male, the other one is female

$MeanSexDummy_{ijt}^{BothFemale}$: Both drivers are female

Overspeeding Dummies (Control group: Neither driver overspeeding):

$MeanOverSpeed_{ijt}^{OneOS}$: One driver overspeeding

$MeanOverSpeed_{ijt}^{BothOS}$: Both drivers overspeeding

DUI Dummies (Control group: Neither driver DUI):

$MeanDUI_{ijt}^{OneDUI}$: One driver DUI

$MeanDUI_{ijt}^{BothDUI}$: Both drivers DUI

$X_{jt}^{(k)} \cdot Z_{ijt}^{(l)}$: 6 interaction terms between each environmental variable and dummy variables

DOW_t: Temporal controls represent the day of the week for each record in the dataset from Sunday to Saturday.

TOD_t: Temporal controls represent the time when each crash occurred, categorized into specific periods of the day. Midnight(12 am-6 am), Morning Peak (6 am-9 am), Day (9 am-4 pm), Evening Peak (4 pm-7 pm), Night (7 pm-12 am)

ε_{ijt}: Error term.

This model examines how $PM_{2.5}$ levels and weather conditions influence the daily number of traffic accidents. It tests the hypothesis that higher PM concentrations and adverse weather increase crash frequency. The model also includes interactions to assess how risky driving behaviors like speeding and drunk driving modify the effect of environmental factors on accidents. Additionally, it accounts for systematic variations in crash counts by time of day and day of the week. Significant results would support the hypotheses regarding pollution, driver behavior, and temporal patterns that affect traffic safety. In the analysis of daily and hourly car crash counts, we include wind speed as an instrumental variable (IV) to address potential endogeneity issues. These behavioral variables may be endogenous due to omitted factors—such as unobserved driver risk attitudes—or reverse causality, where crash risk and driver behavior simultaneously influence each other. Wind speed serves as a suitable instrument because it affects drivers' decisions indirectly by influencing road conditions and driving comfort, thereby altering the propensity to speed or drive under the influence. However, wind speed is exogenous to the crash occurrence except through its effect on these driving behaviors.

4. Results

4.1 Daily Car Crash

Table 3.3 Regression Analysis of Daily Car Crash Frequency

Dependent variable: Daily Car Crash Count			
Variables	**Daily Car Crash Models**		
	FE OLS	**FE Poisson**	**FE Negative Binomial**
Intercept	21.60834***	1.56171***	1.98400***
	(4.37438)	(0.21062)	(0.31130)
Pollution & Weather			
$PM_{2.5}$	0.46171***	0.01380***	0.01163***
	(0.07049)	(0.00119)	(0.00194)
Visibility	-0.30582***	-0.00695***	-0.01195***
	(0.02760)	(0.00085)	(0.00143)
Temperature	-0.00248	-0.00038***	-0.00071***
	(0.00333)	(0.00011)	(0.00017)
Relative Humidity	0.04861***	0.00321***	0.00311***
	(0.00478)	(0.00018)	(0.00027)
Driver Characteristics			
Age Dummy (One<25)	-0.61110	-0.02415	-0.00009
	(0.53419)	(0.02152)	(0.03126)
Age Dummy (Both<25)	-1.91370***	-0.11920***	-0.09273***
	(0.36515)	(0.01359)	(0.02025)
Sex Dummy (One Female)	0.89460***	0.09263***	0.06395
	(0.21107)	(0.00826)	(0.01213)
Sex Dummy (Both Female)	1.65190***	0.15093***	0.12360***
	(0.25750)	(0.01002)	(0.01479)
OS Dummy (One Over Speed)	-8.60767***	-0.76359***	-0.54730***
	(2.62548)	(0.09969)	(0.15490)
OS Dummy (Both Over Speed)	-8.55558***	-0.70114***	-0.50170***
	(1.16439)	(0.04832)	(0.07017)
DUI Dummy (One DUI)	11.76900**	1.26522***	1.11700***
	(5.68203)	(0.24323)	(0.36350)
DUI Dummy (Both DUI)	15.86041***	1.82422***	1.58500***
	(4.29569)	(0.20837)	(0.30720)
Interaction Variables			
$PM_{2.5}$ × OS Dummy (One Over Speed)	0.22185**	0.02361***	0.02434***
	(0.08720)	(0.00319)	(0.00478)
$PM_{2.5}$ × DUI Dummy (One DUI)	-0.04073	0.00443	0.00037
	(0.09844)	(0.00388)	(0.00553)
Visibility × OS Dummy (One Over Speed)	0.55100**	0.04324***	0.02004
	(0.28031)	(0.01028)	(0.01619)

Visibility × DUI Dummy (One DUI)	-0.19917	-0.00398	0.00325
	(0.22537)	(0.00685)	(0.01118)
Temperature × OS Dummy (One Over Speed)	-0.03851*	-0.00311***	-0.00164
	(0.02195)	(0.00094)	(0.00134)
Temperature × DUI Dummy (One DUI)	0.04548	0.00490***	0.00398**
	(0.03307)	(0.00127)	(0.00188)
Relative Humidity × Age Dummy (One<25)	-0.00865	-0.00086**	-0.00108*
	(0.00988)	(0.00039)	(0.00058)
Relative Humidity × DUI Dummy (One DUI)	-0.00872	-0.00023	-0.00024
	(0.03163)	(0.00120)	(0.00177)
Instrument Variable			
Wind Speed	-0.48825***	-0.01574***	-0.01386***
	(0.07177)	(0.00120)	(0.00197)
Model Fit			
Pseudo R^2		0.68021	
Deviance		76,432.22329	31,866.13851
Log Likelihood		-109,961.30536	-100,808.58824
Observations	31,787	31,787	31,787
***p<0.001, **p<0.01, *p<0.05			

Note: The above table shows the daily car crash models with fixed effect on location. The data are all from the original relevant daily data.

Table 3.3 presents estimated effects of air pollution, weather, and characteristics of drivers on daily car crash numbers according to three competing fixed-effects models: OLS, Poisson, and Negative Binomial regression. Across models, there are robust and statistically significant findings.

Holding other variables constant, one-unit increase in concentration of $PM_{2.5}$ ($\mu g/m^3$) is associated with a 0.46171 increase in daily number of crashes ($p < 0.001$) in the OLS model, a 1.4% increase ($p < 0.001$) in the Poisson model, and a 1.2% increase ($p < 0.001$) in the Negative Binomial model. These results confirm that higher concentrations of particulate matter increase road safety hazards. Visibility has a uniformly negative coefficient: a one-unit decrease in visibility is associated with a reduction of 0.30582 in daily crashes ($p < 0.001$) in OLS, a decrease of 0.7% in Poisson ($p < 0.001$), and a decrease of 1.2% in Negative Binomial model. This supports the hypothesis that bad visibility increases the probability of accidents.

Temperature has a negative correlation with car crashes. In the Poisson model, the result is significant, with a coefficient of -0.00038 ($p < 0.001$), indicating that a one-unit increase in temperature leads to a 0.38% decrease in car crashes. In the Negative Binomial model, the coefficient is -0.00071 ($p < 0.001$). Relative humidity is positively related to crashes: a one-percentage-point increase is related to an increase by 0.04861 ($p < 0.001$, OLS), 0.32% ($p < 0.001$, Poisson), and 0.31% ($p < 0.001$, Negative Binomial).

With respect to when both the drivers are above 25 years of age, the correlation coefficient when both of the drivers are below 25 years of age is -1.91370 ($p < 0.001$) in the OLS model, -0.11920 ($p < 0.001$) in Poisson and -0.09273 ($p < 0.001$) in the Negative Binomial model. This shows that two young drivers are negatively associated with the number of accidents. This may be a sign of the correlation that the accident frequency is lower in groups of two young drivers, possibly due to a defensive driving culture or other factors. Relative to the model with two male drivers, a single female driver's coefficient is 0.89460 ($p < 0.001$) in the OLS model, 0.09263 ($p < 0.001$) in the Poisson model and 0.06395 ($p < 0.001$) in the Negative Binomial model; that of two female drivers is even larger at 1.65190 ($p < 0.001$), 0.15093 ($p < 0.001$) and 0.12360 ($p < 0.001$), respectively. This implies that there is a positive relationship between female motorists and the number of car accidents, specifically when both motorists are female, and this further suggests a strong relationship between the gender pair and the number of accidents. Relative to the baseline scenario where neither of the two drivers is speeding, the coefficient of "one speeding" in the OLS model is -8.60767 ($p < 0.001$) and that of “both speeding” is -8.55558 ($p < 0.001$), Poisson model is -0.76359 ($p < 0.001$), and that of "both speeding" is -0.70114 ($p < 0.001$); the corresponding coefficients in the Negative Binomial model are -0.54730 ($p < 0.001$) and -0.50170 ($p < 0.001$). This negative relationship indicates that speeding behavior is statistically associated with a lower rate of traffic crashes, but this result is confounded by selection and must not be read directly as speeding reducing crashes. The most likely explanation is selection error. Drivers who feel confident enough to speed are presumably more skilled in driving and therefore have fewer accidents. Relative to the case where neither of the drivers is drunk, drunk driving has a positive and significant effect on the number of traffic accidents. Under the Poisson model, the coefficients for "one driver in an intoxicated state due to alcohol" are 1.26522 ($p < 0.001$), and for "two drivers in an intoxicated state due to alcohol" are 1.82422 ($p < 0.001$); under the Negative Binomial model, the coefficients are 1.11700 ($p < 0.01$) and 1.58500 ($p < 0.001$), respectively. This suggests that there is a highly positive relationship between drunk driving and traffic accidents, as expected by the general view that drunk driving raises the risk of accidents.

The interaction term between $PM_{2.5}$ and single-passenger speeding is significantly positive in the OLS model (coefficient = 0.22185, $p < 0.01$), the Poisson model (coefficient = 0.02361, $p < 0.001$), and the Negative Binomial model (coefficient = 0.02434, $p < 0.001$). This suggests that higher $PM_{2.5}$ levels strengthen the positive association between single-passenger speeding and the number of traffic accidents, possibly because air pollution exacerbates the risks associated with speeding by impairing driver visibility or physiological performance. The single-passenger speeding by visibility interaction term is highly positive in both the OLS model (coefficient = 0.55100, $p < 0.01$) and the Poisson model (coefficient = 0.04324, $p < 0.001$), indicating that as visibility improves, the traffic accident by single-passenger speeding relationship also improves, presumably because drivers speed when visibility is better, thus making accidents more likely. Temperature and single-passenger speeding interaction term is significantly negative in the OLS (coefficient = -0.03851, $p < 0.05$) and Poisson model (coefficient = -0.00311, $p < 0.001$), which means that with an increase in temperature, the correlation between single-passenger speeding and number of accidents becomes slightly weaker, and the reason may be the complexity of speeding behavior impact due to the change in temperature. The temperature and single drunk driving interaction term is extreme and positive in the Poisson and negative binomial models with coefficients 0.00490 and 0.00398 ($p < 0.001$ and $p < 0.01$), respectively, which suggests that at higher temperatures, the relationship between single drunk driving and the number of traffic accidents becomes stronger, meaning that warm weather will increase the likelihood of drunk driving. The relative humidity-young driver interaction term (one of them is less than 25 years old) is very significant in the Poisson (coefficient = -0.00086, $p < 0.01$) as well as negative binomial (coefficient = -0.00108, $p < 0.05$) models, indicating that with higher humidity, the relationship with traffic accident risk caused by young drivers is weakened.

Wind speed is used as a proxy for the $PM_{2.5}$ sensor, and the findings show that higher wind speeds are associated with significantly lower crash rates. A unit increase in wind speed reduces the odds of crashes by 0.48825 ($p < 0.001$) and results in 1.4–1.6% lower crashes in Poisson and negative binomial models ($p < 0.001$). This aligns with the notion that pollutant dispersion enhances safety performance.

4.2 Hourly Car Crash

Table 3.4 Negative Binomial Regression Analysis of Hourly Car Crash Frequency

Dependent Variable: Hourly Car Crash Count		
Variables	**Hourly Car Crash Models**	
	FE Poisson	**FE Negative Binomial**
Intercept	-0.40410***	-0.40470***
	(0.13540)	(0.13560)

Pollution & Weather		
$PM_{2.5}$	0.10030***	0.10040***
	(0.00151)	(0.00152)
Visibility	-0.03066***	-0.03065***
	(0.00093)	(0.00093)
Temperature	0.00048***	0.00048***
	(0.00012)	(0.00012)
Relative Humidity	0.00007	0.00007
	(0.00013)	(0.00013)
Driver Characteristics		
Age Dummy (One<25)	0.04088***	0.04089***
	(0.00941)	(0.00944)
Age Dummy (Both<25)	0.10040***	0.10050***
	(0.00853)	(0.00855)
Sex Dummy (One Female)	-0.01271**	-0.01271**
	(0.00502)	(0.00503)
Sex Dummy (Both Female)	0.03096***	0.03908***
	(0.00610)	(0.00612)
OS Dummy (One Over Speed)	-0.11300**	-0.11280***
	(0.04204)	(0.04217)
OS Dummy (Both Over Speed)	0.12420***	0.12430***
	(0.03019)	(0.03026)
DUI Dummy (One DUI)	0.02588	0.02556
	(0.16970)	(0.17000)
DUI Dummy (Both DUI)	0.38440***	0.38450***
	(0.13310)	(0.13330)
Temporal Controls		
Week Dummy (Monday)	-0.03168***	-0.03166**
	(0.00657)	(0.00659)
Week Dummy (Saturday)	-0.09034***	-0.09033***
	(0.00723)	(0.00725)
Week Dummy (Sunday)	-0.16890***	-0.16880***
	(0.00872)	(0.00874)
Week Dummy (Thursday)	-0.00897	-0.00896
	(0.00624)	(0.00625)
Week Dummy (Tuesday)	-0.02770***	-0.02767***
	(0.00643)	(0.00644)
Week Dummy (Wednesday)	0.02505***	0.02503***
	(0.00628)	(0.00629)
Time Dummy (Morning Peak)	0.35400***	0.35400***
	(0.01274)	(0.01277)
Time Dummy (Day)	0.56300***	0.56310***
	(0.01242)	(0.01244)
Time Dummy (Evening Peak)	0.52800***	0.52800***

	(0.01256)	(0.01258)
Time Dummy (Night)	-0.01568	-0.01569
	(0.01456)	(0.01459)
Interaction Variables		
$PM_{2.5}$ × OS Dummy (One Over Speed)	0.01125***	0.01125***
	(0.00167)	(0.00167)
$PM_{2.5}$ × DUI Dummy (One DUI)	0.00400	0.00401
	(0.00250)	(0.00250)
Visibility × OS Dummy (One Over Speed)	-0.00408	-0.00410
	(0.00420)	(0.421
Visibility × DUI Dummy (One DUI)	0.03178***	0.03180***
	(0.00719)	(0.00721)
Temperature × OS Dummy (One Over Speed)	0.00126***	0.00126***
	(0.00048)	(0.00048)
Temperature × DUI Dummy (One DUI)	0.00038	0.00038
	(0.00082)	(0.00082)
Relative Humidity × Age Dummy (One<25)	0.00025	0.00025
	(0.00018)	(0.00018)
Relative Humidity × DUI Dummy (One DUI)	-0.00029	-0.00029
	(0.00071)	(0.00071)
Instrument Variable		
Wind Speed	-0.10330***	-0.10340***
	(0.00156)	(0.00156)
Model Fit		
Log Likelihood	-214,468.89081	-214,467.75527
Deviance	103,396.57475	102,805.19596
Observations	128,481	128,481

***p<0.001, **p<0.01, *p<0.05

Table 3.4 shows the results of two fixed effects count models, Poisson and Negative Binomial, that estimate the relationship between environmental conditions, driver characteristics, and hourly traffic crash counts in Colorado. The convergence of results across models suggests robustness.

Air quality and weather variables are statistically significant predictors of hourly crash counts. $PM_{2.5}$ level is highly significantly and positively related to the number of crashes. A one-unit increase in $PM_{2.5}$ level ($\mu g/m^3$) is associated with a 10% increase in predicted hourly crash frequency after adjusting for other variables. This robust and consistent effect may reflect impaired visibility, respiratory distress, or psychological factors such as distraction or irritation in dirty air, which contribute to the likelihood of crashes. Visibility has a significant and negative effect (−0.03066, -0.03065, $p < 0.001$), indicating that the frequency of crashes decreases by 3.1% for every one-unit increase in visibility. It is not surprising: improved weather conditions lead drivers to be more alert and responsive. Temperature, as opposed to previous models, has a minor but statistically significant positive coefficient (0.00048, $p < 0.001$). This implies that higher temperatures have a modest increase in crash frequency, possibly through greater travel demand, higher traffic volumes, or more dangerous driving in hot conditions. Relative humidity has a minor and positive coefficient (0.00007), which implies that wetter weather is associated with slightly increased crash counts. This might suggest weather-avoiding behavior or less travel in damp or rainy conditions.

Driver composition is significantly associated with the frequency of crashes. Crashes involving a single driver under 25 years of age are associated with a 4.1% increase in the number of crashes, and both drivers under 25 have a 10% increase, both highly significant ($p < 0.001$). The results reaffirm that young drivers are at higher risk, most likely due to inexperience, rash driving, or overconfidence. Gender composition is also a factor. Compared to those with only male drivers, having one female driver is associated with a 1.3% decrease in the frequency of crashes, whereas having two female drivers results in a 3.1% increase in the frequency of crashes. The trend may be due to differing risk behaviors: although female drivers are potentially more conservative on average, when both drivers are female, the greater involvement may be due to situational variables such as location or exposure by time of day. These effects are tiny but statistically relevant.

In the speed behavior, crashes where both drivers overspeed have a tremendous positive correlation with crash frequency (0.12420, 0.12430, both $p < 0.001$), while one driver's over-speeding has a negative coefficient (-0.11300, $p < 0.001$; -0.1120, $p < 0.001$). This asymmetrical relation supports a threshold mechanism: the joint commission of hazardous acts by the two drivers significantly increases crash risk, whereas one-sided speeding can be countered by defensive driving on the part of the other driver. Driving while intoxicated (DUI) is a strong predictor of involvement in crashes. When both drivers are DUI, crash risk is increased by over 38.5% ($p < 0.001$). The results confirm the amplified risk of alcohol-impaired driving and underscore its contribution to crash risk.

Temporal fixed effects identify robust diurnal and weekly patterns. Against midnight, morning peak (+35.4%), daytime (+56.3%), and evening peak (+52.8%) crash frequencies are notably higher, while nighttime crashes are lower (−1.6%). These results agree with traffic volume patterns and the cycle of congestion. Sunday (−16.9%), Saturday (−9.0%), Monday (-3.2%), Tuesday (−2.8%), and Thursday (-0.9%) have comparatively lower crash counts compared to Friday (the base), while Wednesday (+2.5%) has higher rates.

Interaction terms examine how environmental risks combine with driver behavior. $PM_{2.5}$ intensified the positive association between single-passenger speeding and traffic accidents (interaction coefficient = 0.01125, $p < 0.001$ in both the Poisson and Negative Binomial models). Higher levels of air pollution may compound the risks of speeding by impairing driver alertness or road visibility, making over-speeding more hazardous in polluted environments. The visibility effect is increased in DUI driver-involvement crashes. To be more specific, for every one-unit increase in visibility, there is a correlated 3.2% increase in the rate of crash involvement for DUI-involved crashes ($p < 0.001$). This may reflect overconfidence under improved visibility by impaired motorists. Temperature intensified the negative effect of Over Speeding (interaction coefficient = 0.00126, $p < 0.001$). Warmer temperatures may encourage increased speeds or more aggressive driving behavior, and therefore, over-speeding is more hazardous in such situations. Other interaction coefficients like temperature × DUI and relative humidity × age do not have statistical effects and suggest fewer cumulative effects among these factors.

Wind speed, as a potential instrument or control for atmospheric dispersion, is strongly negatively correlated with the risk of crashes (−0.10330, $p < 0.001$). This may be a result of the destabilizing effect of high winds on cars, reduced visibility due to dust, or changes in driving behavior.

4.3 Dose Response Function from Daily Negative Binomial Regression Model

Following the rationale in Chapter 1 and 2, dose-response functions are used to examine the nonlinear relationship between $PM_{2.5}$ concentrations and the frequency of traffic accidents. This method provides a detailed understanding of how incremental changes in pollution levels translate into accident risk, especially under different weather and driving conditions.

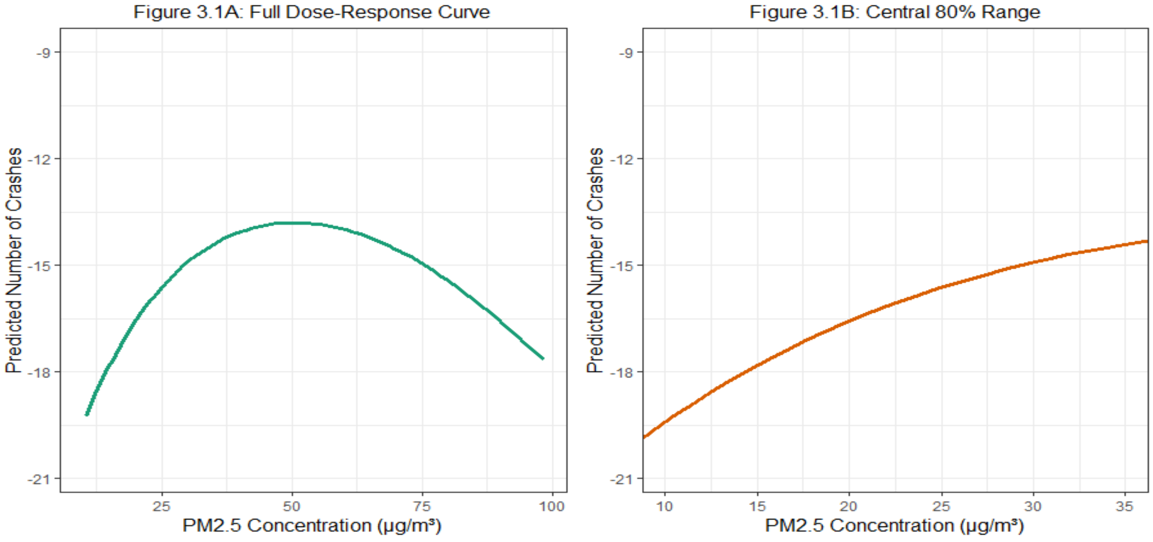

Figure3.1 Dose-Response Curve between $PM_{2.5}$ and Crashes

Figure 3.1 illustrates the dose-response curve for the number of predicted crashes versus the mean $PM_{2.5}$ concentration. The curve is a very strong nonlinear one: as the mean $PM_{2.5}$ concentration rises from about 10 to 50, the number of car crashes rises, then falls. At low levels of $PM_{2.5}$ (i.e., levels below 25), the slope of the increasing crash frequency is very high, suggesting a high-risk potential for crashes even with moderate exposure to pollution. At higher levels of $PM_{2.5}$, above 50, the curve starts to decline. Exposure to $PM_{2.5}$ can cause physiological responses, such as respiratory distress or cardiovascular stress, which can impair driving proficiency.

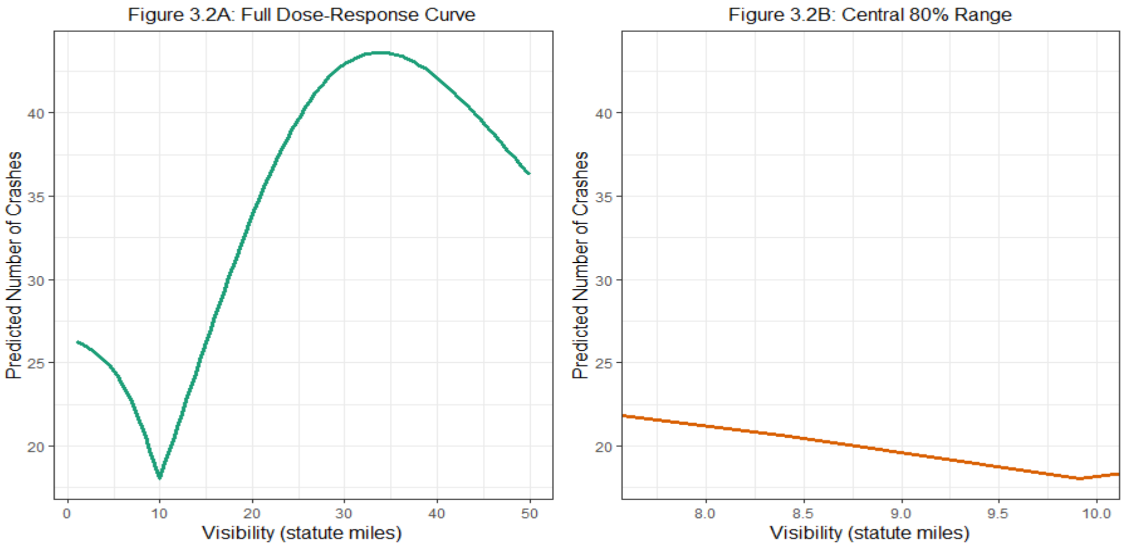

Figure 3.2 Dose-Response Curve between Visibility and Crashes

Figure 3.2 illustrates a non-linear relationship between visibility levels and the expected rate of crashes. First, with growing visibility from 0 to about 10 units, the number of estimated crashes drops to a minimum. However, for 10 to 35 units of visibility, the risk of a crash increases rapidly and peaks at around 35 units of visibility. Then, when visibility improves further, the number of crashes begins to fall again. This pattern suggests that drivers may be driving more safely in very low visibility and reducing crash risk. At moderate visibility, however, drivers continue to overestimate safety and drive more recklessly, thereby experiencing more crashes. Under greater visibility, however, better conditions most probably lead to safer driving and fewer accidents overall.

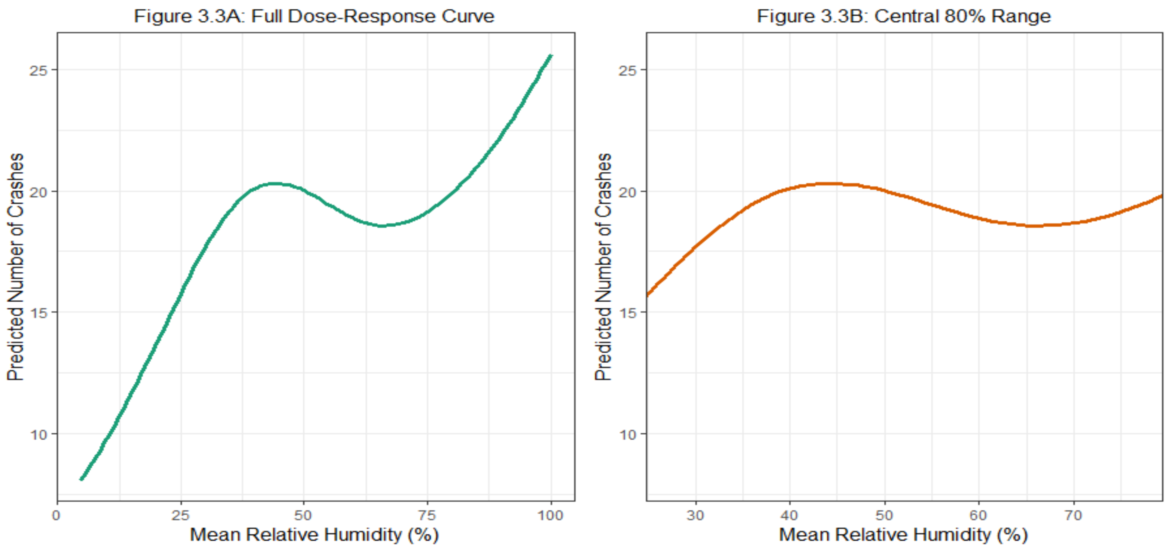

Figure 3.3 Dose-Response Curve between Relative Humidity and Crashes

Figure 3.3 displays the dose-response curve of mean relative humidity (%) vs. number of predicted crashes. The curve is nonlinear with three phases. Firstly, for 0% to approximately 45% relative humidity, the number of predicted crashes increases at a steady rate, and this indicates that drier conditions are perhaps associated with fewer crashes. Then, between the range of about 45% to 70%, the curve decreases somewhat or flattens, marking a plateau where humidity has a diminishing or leveling effect on crash risk. Finally, above 70%—most notably above 80%, the predicted number of crashes increases steeply again, reaching an inflection point near 100% relative humidity. This curve shows that high humidity is associated with increased crash risk, maybe because high humidity is often associated with fog, poor visibility, and slippery roads, all of which contribute to compromised driving safety. The middle range of humidity appears to be the safest range, illustrating a balance of environmental conditions.

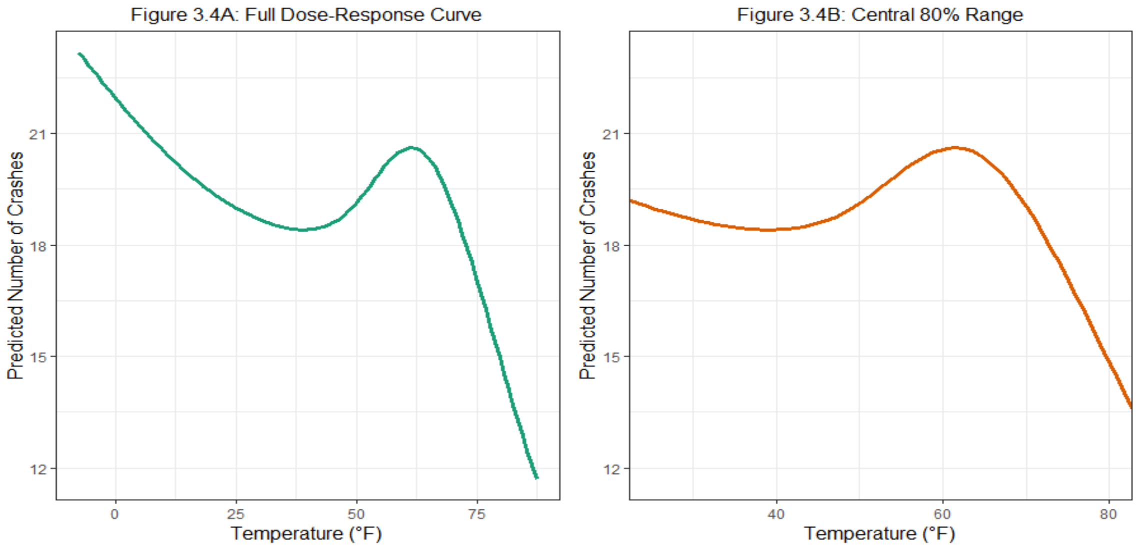

Figure 3.4 Dose-Response Curve between Temperature and Crashes

Figure 3.4 illustrates the dose-response curve of temperature and the number of Colorado traffic accidents. Overall, the pattern is nonlinear but downward, showing that as temperatures rise on average, there are fewer traffic accidents—a finding consistent with the earlier regression models. For temperatures in the range of 0–40°F, the decrease in the number of accidents is relatively modest, which suggests a less marginal relationship between temperature and crash probability. This could be because, for extremely cold temperatures, drivers already expect danger in the form of ice or snow and will drive with caution, partly offsetting the increased physical hazards. Between 63°F and 40°F, the curve trend is moderately upward, which indicates that accident rates increase as the temperature increases within this temperature range. Theory has it that this temperature interval often approximates in-between conditions (say spring or autumn), when road conditions may be inconsistent from time to time due to sporadic thawing, fog, or precipitation, and drivers fail to accordingly adjust. After 63°F, the pattern once again goes downward, with a steeper drop, indicating that higher temperatures are associated with fewer on-road incidents. It may be due to improved road conditions (e.g., good roads), improved performance of the car, and more stable weather, all of which reduce the probability of an accident. Hot weather may also enhance visibility and driver comfort, leading to improved driving conditions.

5. Discussion

5.1 Social Costs and Negative Externalities of Traffic Accidents

The social effects of motor vehicle crashes transfer intense adverse externalities, not all of which are transferred to the participants in the accidents, yet some are shifted to society and generate market failures. Economically, explicit costs transfer health care charges, wear and utilization of public infrastructures of the vehicles, all of which are variables of cost of some importance. The U.S. National Highway Traffic Safety Administration estimated the medical costs of traffic crashes in 2010 to be approximately $9.9 billion, 13% of which was reimbursed by government health programs, such as Medicaid and Medicare. In addition, property loss from traffic crashes in 2010 was valued at $7.6 billion, encompassing the cost of vehicle repairs, replacement value, and insurance claim fees paid, with rising insurance premiums contributing to the total societal costs. Furthermore, emergency response and police enforcement costs due to traffic crashes amounted to approximately US$1 billion, representing the utilization of public resources through police enforcement, accident scene management, medical care, and litigation (Blincoe et al., 2015).

In addition to the direct economic losses, accidents also generate huge indirect costs such as loss of productivity and congestion costs. Traffic crash losses in productivity amounted to $204 billion in 2010, which represents 1.4% of the U.S. GDP, with considerable losses in long-term recovery and diminished work capability (Blincoe et al., 2015). Simultaneously, traffic congestion due to car accidents undermines the effectiveness of the overall transportation system. The economic cost of traffic congestion in large U.S. cities is estimated by the Texas A&M Transportation Institute (TTI, 2019) to be $166 billion per year, with 25%-30% of the cost attributable to traffic accidents. Traffic accidents also lead to secondary environmental pollution, including fuel spills and increased emissions resulting from congestion caused by traffic accidents. Empirical analyses established that emissions of $PM_{2.5}$ from vehicles are 30-40 percent higher in the most congested traffic conditions, and this would result in air pollution and public health issues. The extent and magnitude of these economic losses establish that there is an impossibility of internalizing the social cost of accidents in a market, and government intervention is thus necessary. Policy measures, such as increasing fuel prices and promoting innovative traffic control technologies, not only decrease traffic accident rates but also reduce environmental pollution and improve the overall health of society.

Empirical evidence indicates that elevated $PM_{2.5}$ concentrations intensify traffic accident hazards and their social costs intensively. In the US, a rise of 1 $\mu g/m^3$ in $PM_{2.5}$ concentration raises same-day crash frequencies by 0.5%–1.2% by reducing visibility and acute health effects on drivers. Excesses beyond 62 $\mu g/m^3$ in $PM_{2.5}$ raise traffic fatality rates by 8%–12%, as revealed by China's research. The U.S. Environmental Protection Agency (EPA) estimates social costs of $3–5 billion per year due to traffic crashes caused by air pollution (including $PM_{2.5}$), such as medical expenses, property damage, and legal fees. European $PM_{2.5}$ crash costs range between 0.1% and 0.3% of GDP, and the most polluted regions (e.g., Poland) up to 0.5%. Additionally, Canadian research has shown that for every 10 $\mu g/m^3$ increase in PM2.5, the odds of injury in crashes rise by 6%. These findings highlight that reducing emissions of PM2.5 can lead to improved public health while offsetting economic losses associated with road accidents.

5.2 Limitation

There may be limitations in our research. Sample selection bias (only data on accidents): The model only uses observations of actual accidents and fails to include "zero accident" observations, which may lead to sample selection bias, especially in the marginal effect of variables such as PM, which may be overestimated. Another limitation may be that spatial heterogeneity is not fully controlled, despite controlling for fixed effects (such as counties). Spatial heterogeneity, including factors like enforcement of traffic rules and road structure, may still influence the relationship between accidents and pollution.

6. Conclusion

This study illustrates a strong and statistically significant relationship among air pollution, weather, risky driving behaviors, and the frequency of traffic accidents per hour. More specifically, high levels of $PM_{2.5}$ significantly increase the frequency of traffic accidents, underscoring the hazardous impact of air pollution on traffic safety. In contrast, greater visibility is associated with fewer accidents, suggesting a preventive role. However, there is a relation between high relative humidity and elevated accident rates, indicating that risk could be inherent in the humid atmosphere. Driving habits and driver traits are implicated as well. Female driver pairs both experience higher accident frequencies. Risky behavior, such as alcohol consumption, dramatically increases the number of accidents, with its effect compounded when both culprits engage in such behavior. Notably, $PM_{2.5}$ amplifies the risks associated with unsafe driving practices: the interaction between $PM_{2.5}$ and over-speeding shows a significant positive effect on accident rates, while the combination of $PM_{2.5}$ exposure and drunk driving further elevates crash risks. These findings suggest that air pollution not only independently contributes to accidents but also magnifies the danger of hazardous behavior on the road. Temporal patterns indicate that traffic accidents peak during morning and evening rush hours, whereas weekends, particularly Sundays, are associated with significantly lower accident rates, possibly due to reduced traffic volume and fewer travel purposes. Several interaction effects establish the synergistic effects of perilous conditions through pollution and unsafe driving. While encompassing many factors and making helpful inferences, there remains immense scope for further research. Future research should also analyze other variables, such as road conditions, fatigue, and driving history. An analysis of potential nonlinear effects of existing variables through extensive data set analysis and time series data can be used to gain a deeper understanding and more precise outcomes. Additionally, the application of advanced models, such as mixed-effects models, can be employed to obtain more precise estimations. Relevance to policy formulation and actual application of the results of this present work is relevant. The results can potentially be applied in recommending the formation of an effective driving policy to prevent crash-inducing driving behavior and promote a safe driving culture. Sensitization and training of the public are also significant. Through public awareness of risk causes as well as emphasizing the importance of safe driving, crash occurrence can be significantly reduced. In total, this study not only enhances our understanding of the determinants of motor vehicle accident rates but also offers actionable recommendations for improving road safety through policy interventions and public campaigns.

Bibliography

"Manchester Air Quality Index (AQI) and United Kingdom Air Pollution | AirVisual." Www.iqair.com, www.iqair.com/us/uk/england/manchester.

"Risk Attitude and Air Pollution: Evidence from Chess." Papers.ssrn.com, 20 May 2020, papers.ssrn.com/sol3/papers.cfm?abstract_id=3609237.

"空气质量、投资者情绪与股票收益率--《管理科学》2018 年 06 期." Www.cnki.com.cn, www.cnki.com.cn/Article/CJFDTotal-JCJJ201806013.htm. Accessed 10 Feb. 2023.

Affum, H. A., Akaho, E. H. K., Niemela, J. J., Armenio, V., & Danso, K. A. (2016). Validating the California Puff (CALPUFF) modelling system using an industrial area in Accra, Ghana as a case study. Open Journal of Air Pollution, 5(1), 27-36.

Ailshire, Jennifer, and Katrina M. Walsemann. "Education Differences in the Adverse Impact of $PM_{2.5}$ on Incident Cognitive Impairment among U.S. Older Adults." Journal of Alzheimer's Disease, vol. 79, no. 2, 19 Jan. 2021, pp. 615–625, https://doi.org/10.3233/jad-200765. Accessed 10 Aug. 2021.

Berman, Jesse D., and Keita Ebisu. "Changes in U.S. Air Pollution during the COVID-19 Pandemic." Science of the Total Environment, vol. 739, 15 Oct. 2020, p. 139864, www.sciencedirect.com/science/article/pii/S0048969720333842#t0005, 10.1016/j.scitotenv.2020.139864.

Bhattarai, G., Shrestha, S., Sim, H.-J., Lee, J.-C., & Kook, S. (2024). The Impact of PM, NO2, and SO2 on Respiratory Health and Relative Economics.

Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015). The economic and societal impact of motor vehicle crashes, 2010 (Revised). National Highway Traffic Safety Administration.

CA PIII, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Jr Heath CW: Particulate air pollution as a predictor of mortality in a prospective study of U.S. Adults. Am J Respir Crit Care Med. 1995, 151: 669-674. 10.1164/ajrccm/151.3_Pt_1.669.

Chang, T. Y., Zivin, J. G., Gross, T., & Neidell, M. (2019). The effect of pollution on high-skilled workers: Evidence from call center workers in China. American Economic Journal: Applied Economics, 11(1), 29-58. https://doi.org/10.1257/app.20170500

Chay, Kenneth Y., & Greenstone, Michael. (2003). Air Quality, Infant Mortality, and the Clean Air Act of 1970. (NBER Working Paper No. 10053.

Chay, Kenneth Y., & Greenstone, Michael. (2003). The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. The Quarterly Journal of Economics, 118(3), 1121-1167.

Chen, H., et al. (2018). Traffic-Related Air Pollution and Road Accidents in Toronto. Environmental Health Perspectives, 126(5), 057008. DOI: 10.1289/EHP2231

Chen, Jiu-Chiuan, and Joel Schwartz. “Neurobehavioral Effects of Ambient Air

Chen, Yuyu, Ebenstein, Avraham, Greenstone, Michael, & Li, Hongbin. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. Proceedings of the Proceedings of the National Academy of Sciences, 110(32), 12936-12941.

Christensen, Grace M., et al. “The Complex Relationship of Air Pollution and Neighborhood Socioeconomic Status and Their Association with Cognitive Decline.”

Environment International, vol. 167, Sept. 2022, p. 107416, https://doi.org/10.1016/j.envint.2022.107416. Accessed 22 July 2022.

Colorado Department of Transportation (CDOT) Data: Colorado Department of Transportation. (2023). Traffic Crash Data. Retrieved from CDOT website

Currie, Janet, & Walker, Reed. (2011). Traffic Congestion and Infant Health: Evidence from E-ZPass. American Economic Journal: Applied Economics, 3(1), 65-90.

Deng, Y., He, J., Shen, X., & Li, B. (2023). Does Air Pollution Cause More Car Accidents? Evidence from Auto Insurance Claims. Evidence from Auto Insurance Claims.

Dockery DW, Pope CA: Acute respiratory effects of particulate air pollution. Ann Rev Public Health. 1994, 15: 107-132. 10.1146/annurev.pu.15.050194.000543.

Du, Yixing, et al. "Air Particulate Matter and Cardiovascular Disease: The Epidemiological, Biomedical and Clinical Evidence." Journal of Thoracic Disease, vol. 8, no. 1, 2016, pp. E8–E19, www.ncbi.nlm.nih.gov/pmc/articles/PMC4740122/, 10.3978/j.issn.2072-1439.2015.11.37. Accessed 2 Feb. 2020.

Finkelstein, A., Gentzkow, M., & Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. American Economic Review: Applied Economics, 8(2), 228-248. https://doi.org/10.1257/app.20140100

Ghazali, N. A., Ramli, N. A., Yahaya, A. S., Yusof, N. F. F. M., Sansuddin, N., & Al Madhoun, W. A. (2010). Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. Environmental Monitoring and Assessment, 165(1-4), 475–489

Graff Zivin, J., & Neidell, M. (2012). The impact of pollution on worker productivity. American Economic Review, 102(7), 3652-3673. https://doi.org/10.1257/aer.102.7.3652

He, J., Liu, Y., & Tang, X. (2019). The Effect of Air Pollution on Road Traffic Accidents. National Bureau of Economic Research (NBER) Working Paper No. 26065. Retrieved from https://www.nber.org/papers/w26065

He, Y., Jiang, W., Gao, X. et al. Short-term effects and economic burden of air pollutants on acute lower respiratory tract infections in children in Southwest China: a time-series study. Environ Health 22, 6 (2023). https://doi.org/10.1186/s12940-023-00962-3

Heyes, A., Neidell, M., & Saberian, S. (2016). The effect of air pollution on investor behavior: Evidence from the S&P 500. National Bureau of Economic Research (NBER Working Paper No. 22753). https://doi.org/10.3386/w22753

Hoogeveen, O. (2023). Particulate matter on performance in professional Counter-Strike players (Master's thesis). California State University, Fullerton. https://doi.org/10.1287/mnsc.2022.4643

https://doi.org/10.1016/j.jeem.2023.102886JIAWEI MO+2IDEAS/RePEc+2X-MOL+2

Jana, Arup, et al. “Household Air Pollution and Cognitive Health among Indian Older Adults: Evidence from LASI.” Environmental Research, vol. 214, 1 Nov. 2022, p. 113880, www.sciencedirect.com/science/article/abs/pii/S0013935122012075, https://doi.org/10.1016/j.envres.2022.113880. Accessed 11 Feb. 2023.

Jiang, Yaqiong, et al. “Air Quality, Human Behavior and Urban Park Visit: A Case Study in Beijing.” Journal of Cleaner Production, vol. 240, Dec. 2019, p. 118000, 10.1016/j.jclepro.2019.118000.

Joo, S., Oh, C., Lee, S., & Lee, G. (2017). Assessing the impact of traffic crashes on near freeway air quality. Transportation research part D: transport and environment, 57, 64-73.

Kan HD, Chen BH: A case-crossover analysis of Air pollution and daily mortality in Shanghai. J Occu Health. 2003, 45: 119-124. 10.1539/joh.45.119.

Klingen, J., & van Ommeren, J. (2022). Risk-Taking and Air Pollution: Evidence from Chess. Environmental and Resource Economics, 81, 73–93. https://doi.org/10.1007/s10640-021-00618-1

Krzeszowiak, J., Stefanow, D., & Pawlas, K. (2016). The impact of particulate matter (PM) and nitric oxides (NOx) on human health and an analysis of selected sources accounting for their emission in Poland. Medycyna Środowiskowa-Environmental Medicine, 19(3), 7-15.

Künn, S., Palacios, J., & Pestel, N. (2023). Indoor air quality and strategic decision making. Management Science. https://doi.org/10.1287/mnsc.2022.4643

Levy, Tamir, and Joseph Yagil. “Air Pollution and Stock Returns in the US.” Journal of Economic Psychology, vol. 32, no. 3, June 2011, pp. 374–383, 10.1016/j.joep.2011.01.004. Accessed 6 Sept. 2020.

Li, Huichu, et al. “Particulate Matter Exposure and Stress Hormone Levels.” Circulation, vol. 136, no. 7, 15 Aug. 2017, pp. 618–627, 10.1161/circulationaha.116.026796. Accessed 8 May 2022.

Liu, C., et al. (2020). Associations between ambient particulate matter and hospitalizations for traffic injuries in China. The Lancet Planetary Health, 4(12), e555-e565. DOI: 10.1016/S2542-5196(20)30255-4

Liu, X. (2015). A more accurate method using MOVES (Motor Vehicle Emission Simulator) to estimate emission burden for regional-level analysis. Journal of the Air & Waste Management Association, 65(7), 837-843.

Lopez, D. P. (2003). The relationship between organizational culture and the job satisfaction and turnover intention of primary care providers (Publication No. 304541997) [Doctoral dissertation, University of Phoenix]. ProQuest Dissertations and Theses Global. https://www.proquest.com/docview/304541997?pq-origsite=gscholar&fromopenview=true

Maji, S., Yan, H., & Xia, W. (2021). Impact of PM2.5 on COVID-19 infection in highly polluted cities in Jiangsu province, China. Respiratory Research, 22(1), Article 91. https://doi.org/10.1186/s12931-021-01731-x

MARR, LINSEY C., and MATTHEW R. ELY. "Effect of Air Pollution on Marathon Running Performance." Medicine & Science in Sports & Exercise, vol. 42, no. 3, Mar. 2010, pp. 585–591, https://doi.org/10.1249/mss.0b013e3181b84a85. Accessed 3 Dec. 2019.

Mo, J., Wu, Z., & Yuan, Y. (2023). Air pollution kills competition: Evidence from eSports. Journal of Environmental Economics and Management, 122, 102886.

Musselwhite, C., Avineri, E., & Susilo, Y. (2017). Editorial JTH 10: The Coronavirus Disease COVID-19 and implications for transport and health. Journal of Transport & Health, 10, 1-2. https://doi.org/10.1016/j.jth.2017.03.001

Nair, K.C., Bihari, V., Pangtey, B.S. et al. Respiratory health problems associated to infrastructural development among residents living near Special Economic Zone in India. Clean Techn Environ Policy 13, 697–702 (2011). https://doi.org/10.1007/s10098-010-0337-x

Park, WooRam, and Yongmi Kim. "Air Pollution and Risk Preference." Journal of Economic Behavior & Organization, vol. 202, Oct. 2022, pp. 566–579, https://doi.org/10.1016/j.jebo.2022.08.024. Accessed 3 Nov. 2022.

Pollution on Cognitive Performance in US Adults." NeuroToxicology, vol. 30, no. 2, Mar. 2009, pp. 231–239, https://doi.org/10.1016/j.neuro.2008.12.011. Accessed 2 May 2019.

Quah, Euston, and Tay Liam Boon. "The economic cost of particulate air pollution on health in Singapore." Journal of Asian Economics 14.1 (2003): 73-90.

Sager, L. (2019). Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management, 98*, 102250. https://doi.org/10.1016/j.jeem.2019.102250

Sager, L. (2021). The Social Cost of Air Pollution-Related Traffic Accidents. Journal of Environmental Economics and Management, 105, 102400. DOI: 10.1016/j.jeem.2020.102400

Shen, J., Barbera, J., & Shapiro, C. M. (2019). Distinguishing sleepiness and fatigue: Focus on definition and measurement. Sleep Health, 5(3), 209-216. https://doi.org/10.1177/0748730419854197

South Carolina Department of Health and Environmental Control (SCDHEC. (n.d.). Nitrogen Dioxide & Sulfur Dioxide Fact Sheet. Retrieved from here.

Suman Chakrabarti, Mohammed Tajuddin Khan, Avinash Kishore, Devesh Roy, Samuel P Scott, Risk of acute respiratory infection from crop burning in India: estimating disease burden and economic welfare from satellite and national health survey data for 250 000 persons, International Journal of Epidemiology, Volume 48, Issue 4, August 2019, Pages 1113–1124, https://doi.org/10.1093/ije/dyz022

Sun, Z., An, X., Tao, Y. et al. Assessment of population exposure to PM10 for respiratory disease in Lanzhou (China) and its health-related economic costs based on GIS. BMC Public Health 13, 891 (2013). https://doi.org/10.1186/1471-2458-13-891

Texas A&M Transportation Institute (TTI). (2019). Urban Mobility Report. Texas A&M University.

U.S. Environmental Protection Agency (EPA). (2019). Integrated Science Assessment for Particulate Matter. EPA/600/R-19/188. Retrieved from https://www.epa.gov/isa

United States Environmental Protection Agency (EPA) Data: United States Environmental Protection Agency. (2023). Air Quality Data. Retrieved from EPA website

US EPA. "Health and Environmental Effects of Particulate Matter (PM)." US EPA, 9 May 2019, www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm.

Wang, T., Wang, Y., & Cui, N. (2022). Traffic costs of air pollution: the effect of PM2. 5 on traffic violation. Environmental Science and Pollution Research, 29(48), 72699-72717.

Wolf, Kathrin, et al. "Associations between Short-Term Exposure to Particulate Matter and Ultrafine Particles and Myocardial Infarction in Augsburg, Germany." International Journal of Hygiene and Environmental Health, vol. 218, no. 6, Aug. 2015, pp. 535–542, 10.1016/j.ijheh.2015.05.002. Accessed 30 Mar. 2021.

Zeka A, Zanobetti A, Schwartz J: Short term effects of particulate matter on cause specific mortality: effects of lags and modification by city characteristics. Occup Environ Med. 2005, 62: 718-725. 10.1136/oem.2004.017012.

Zhang, Xin, et al. "The Impact of Exposure to Air Pollution on Cognitive Performance." Proceedings of the National Academy of Sciences, vol. 115, no. 37, 27 Aug. 2018, pp. 9193–9197, www.pnas.org/content/115/37/9193, 10.1073/pnas.1809474115.

Zhang, Z., Wang, J., & Lu, W. (2018). Exposure to nitrogen dioxide and chronic obstructive pulmonary disease (COPD) in adults: a systematic review and meta-analysis. Environmental science and pollution research international, 25(15), 15133–15145. https://doi.org/10.1007/s11356-018-1629-7

Appendix

The repository can be accessed at:

https://github.com/StrawberrySoap/CSU-Dissertation/blob/3a5ab55c76d32f6f6bdf511eaf4b4931cf548a34/Dissertation%20R%20code

The repository includes:

R scripts for data cleaning, variable construction, and regression modeling;

Code for generating all figures and tables included in the thesis;

Researchers, reviewers, and readers are encouraged to access the repository and are welcome to reuse the materials for non-commercial research purposes, subject to the licensing terms indicated in the repository.

Appendix A. Dose-Response Function Specifications and Marginal Effects (Chapter 1)

This appendix provides the full dose-response function specifications and marginal effect expressions for all key explanatory variables included in the Beta regression model of Chapter 1. The dependent variable is the adjusted daily winning rate in chess, bounded between 0 and 1.

The estimated model is as follows:

$$\mathrm{logit}(\mu_i) = a_0 + a_y + a_m + a_d + \beta_1 \mathrm{DPMlag}_i + \beta_2 \mathrm{DPM}_i + \beta_3 \mathrm{DDBT}_i + \beta_4 \mathrm{DP}_i + \beta_5 \mathrm{DRH}_i + \beta_6 \mathrm{DSLP}_i + \gamma_1 \mathrm{RATE}_i + \gamma_2 \mathrm{AGE}_i + \gamma_3 \mathrm{INC}_i + \gamma_4 \mathrm{SEX}_i + \delta_1(\mathrm{DPM}_i \mathrm{RATE}_i) + \delta_2(\mathrm{DPM}_i \mathrm{AGE}_i) + \delta_3(\mathrm{DPM}_i \mathrm{INC}_i) + \delta_4(\mathrm{DPM}_i \mathrm{SEX}_i)$$

Where:

μ_i: expected winning rate for player i

DPM_i: $PM_{2.5}$ concentration

DPMlag_i: $PM_{2.5}$ 20-hour concentration

DDBT_i: daily dry bulb temperature

DRH_i: relative humidity

DP_i: daily precipitation

DSLP_i: sea level pressure

AGE_i: player's age

INC_i: income level

SEX_i: player’s gender

RATE_i : rating difference between player and opponent

A.1 $PM_{2.5}$ Concentration

Model Terms:

$\beta_2\mathrm{DPM_i} + \delta_1(\mathrm{DPM_iRATE_i}) + \delta_2(\mathrm{DPM_iAGE_i}) + \delta_3(\mathrm{DPM_iINC_i}) + \delta_4(\mathrm{DPM_iSEX_i})$

Marginal Effect:

$$\frac{\partial\mu_i}{\partial\mathrm{DPM_i}} = \mu_i(1-\mu_i)\cdot(\beta_2 + \delta_1(\mathrm{RATE_i}) + \delta_2(\mathrm{AGE_i}) + \delta_3(\mathrm{INC_i}) + \delta_4(\mathrm{SEX_i}))$$

A.2 Relative Humidity

Model Term: $\beta_5\mathrm{DRH_i}$

Marginal Effect: $\frac{\partial\mu_i}{\partial\mathrm{DRH_i}} = \mu_i(1-\mu_i)\cdot\beta_5$

A.3 Age

Model Term: $\gamma_2\mathrm{AGE_i} + \delta_2(\mathrm{DPM_iAGE_i})$

Marginal Effect: $\frac{\partial\mu_i}{\partial\mathrm{AGE_i}} = \mu_i(1-\mu_i)\cdot(\gamma_2 + \delta_2(\mathrm{DPM_i}))$

A.4 Income

Model Term: $\gamma_3\mathrm{INC_i} + \delta_3(\mathrm{DPM_iINC_i})$

Marginal Effect: $\frac{\partial\mu_i}{\partial\mathrm{INC_i}} = \mu_i(1-\mu_i)\cdot(\gamma_3 + \delta_3(\mathrm{DPM_i}))$

A.5 Rating Difference

Model Term: $\gamma_1\mathrm{RATE_i} + \delta_1(\mathrm{DPM_iRATE_i})$

Marginal Effect: $\frac{\partial\mu_i}{\partial\mathrm{RATE_i}} = \mu_i(1-\mu_i)\cdot(\gamma_1 + \delta_1(\mathrm{DPM_i}))$

A.6 $PM_{2.5}$: Sex

Model Term: $\delta_4(\mathrm{DPM_iSEX_i})$

Marginal Effect: $\frac{\partial \mu_i}{\partial \mathrm{DPM_i RATE_i}} = \mu_i(1 - \mu_i) \cdot \delta_4$

A.7 $PM_{2.5}$: Rating Difference

Model Term: $\delta_1(\mathrm{DPM_i RATE_i})$

Marginal Effect: $\frac{\partial \mu_i}{\partial \mathrm{DPM_i RATE_i}} = \mu_i(1 - \mu_i) \cdot \delta_1$

Appendix B. Robustness Checks (Chapter 1)

To ensure the reliability of the empirical findings in Chapter 1, I conducted a series of robustness checks across all three model specifications: logistic regression, OLS, and beta regression.

B.1 Logistic Regression Model

Multicollinearity: After removing the wet bulb temperature due to high correlation with other weather variables, all remaining covariates showed acceptable multicollinearity levels (VIF < 3), except for $PM_{2.5}$ and its 10-hour lag, which were expectedly correlated.

Fixed Effects Robustness: I included fixed effects for year, month, day, hour, zip code, and player ID to absorb unobserved temporal and spatial heterogeneity as well as individual-specific characteristics. This ensured that estimated pollution effects are not confounded by fixed contextual differences.

Heteroskedasticity: Breusch–Pagan tests indicated strong evidence of heteroskedasticity ($p < 2.2e-16$). To correct this, I employed cluster-robust standard errors at the player ID level.

Endogeneity: I use hourly wind speed as instrument variable trying to solve potential endogenous problem, in the first stage, coefficient of hourly wind speed is -0.10350 (0.00272) with $p<0.001$, F-value is 3432 < 10, showing it’s a good IV, but in the second stage, the insignificant results show that the model does not have a significant endogeneity problem.

B.2 Ordinary Least Squares (OLS) Model

Nonlinearity: Linearity assumptions were tested and rejected (p-value extremely small), indicating a nonlinear relationship. To improve model flexibility and fit, I used natural splines with six degrees of freedom (i.e., ns(x, df = 6)) to model continuous predictors including $PM_{2.5}$.

Collinearity: The pairwise correlation coefficients between variables were all below 0.1, indicating no serious multicollinearity concerns.

Heteroskedasticity: The Breusch–Pagan test again showed strong evidence of heteroskedasticity. I addressed this by using cluster-robust standard errors clustered at the player level.

Endogeneity: I use daily wind speed as instrument variable trying to solve potential endogenous problem, in the first stage, coefficient of daily wind speed is 0.05010 (0.00309) with $p<0.001$, F-value is $8038 < 10$ for the original OLS model, and 0.04717 (0.00363) with $p<0.001$, F-value is $5988 < 10$ in the model excluding 0% and 100% winning rate, both showing it's a good IV, but in the second stage, the insignificant results show that the model does not have a significant endogeneity problem.

B.3 Beta Regression Model

Model Fit: Given the bounded nature of the dependent variable (adjusted winning rate between 0 and 1), I applied a beta regression. The pseudo-R-squared of 0.245 indicated a moderate fit and improved over OLS and logistic alternatives.

Multicollinearity: After removing the wet bulb temperature, all remaining variables had VIF < 3, with the exception of $PM_{2.5}$ and its lag.

Heteroskedasticity: Test results indicated statistically significant heteroskedasticity ($p < 0.01$), addressed using cluster-robust standard errors at the player ID level.

These robustness checks collectively affirm the stability of the results across different estimation strategies. Each model addresses potential statistical and econometric concerns, ensuring that the main conclusions regarding the effects of air pollution on chess performance are not artifacts of modeling assumptions.

Endogeneity: I use daily wind speed as instrument variable trying to solve potential endogenous problem, in the first stage, coefficient of daily wind speed is -0.07986 (0.00377) with $p<0.001$, F-value is 6308 < 10, showing it’s a good IV, but in the second stage, the insignificant results show that the model does not have a significant endogeneity problem.

Appendix C. Dose-Response Function Specifications and Marginal Effects (Chapter 2)

In Chapter 2, I investigated the relationship between environmental pollution and asthma using logistic regression. The dependent variable is a binary indicator of current asthma status $\mathrm{CASTHMA}_{ijt}$, where the subscripts i, j, and t refer to the individual, county, and date, respectively.

The estimated model is:

$$\log\left(\frac{\mathrm{P(CASTHMA_i=1)}}{1-\mathrm{P(CASTHMA_i=1)}}\right) = \alpha_0 + \alpha_y + \alpha_m + \alpha_d + a_j + \beta_1 \cdot \mathrm{PMAve_i} + \beta_2 \cdot \mathrm{SOAve_i} + \beta_3 \cdot \mathrm{RHAve_i} + \beta_4 \cdot \mathrm{TEMPAve_i} + \beta_5 \cdot \mathrm{PMlag_i} + \beta_6 \cdot \mathrm{SOlag_i} + \gamma_1 \cdot \mathrm{BMI_i} + \gamma_2 \cdot \mathrm{AGE_i} + \gamma_3 \cdot \mathrm{SMOKER_i} + \gamma_4 \cdot \mathrm{GENHLTH_i} + \gamma_5 \cdot \mathrm{SEX_i} + \gamma_6 \cdot \mathrm{EXERANY_i} + \delta_1(\mathrm{PMAve_i} \cdot \mathrm{SMOKER3_i}) + \delta_2(\mathrm{PMAve_i} \cdot \mathrm{GENHLTH_i}) + \eta_1(\mathrm{SOAve_i} \cdot \mathrm{SMOKER_i}) + \eta_2(\mathrm{SOAve_i} \cdot \mathrm{GENHLTH_i}) + \eta_3(\mathrm{SOAve_i} \cdot \mathrm{AGE_i}) + \eta_4(\mathrm{SOAve_i} \cdot \mathrm{SEX_i}) + \eta_5(\mathrm{SOAve_i} \cdot \mathrm{BMI_i})$$

Where:

$\alpha_y,\alpha_m,\alpha_d,\alpha_j$: Year, month, day, and county fixed effects.

CASTHMA: Binary indicator for current asthma attack of individual.

PMAve, SOAve, TEMP, RHAve: Pollution and weather variables.

PMlag, SOlag : 20-hour lagged pollution concentrations.

X: Individual characteristics (e.g., SEX, BMI4, AGE_G, SMOKER3, GENHLTH, EXERANY2).

C.1 $PM_{2.5}$

Model Term:

$$\beta_1 \cdot \mathrm{PMAve_i} + \delta_1(\mathrm{PMAve_i} \cdot \mathrm{SMOKER3_i}) + \delta_2(\mathrm{PMAve_i} \cdot \mathrm{GENHLTH_i})$$

Marginal Effect:

$$\frac{\partial \mathrm{logit}(\mathrm{P}(\mathrm{CASTHMA}=1))}{\partial \mathrm{PMAve_i}} = \beta_1 + \delta_1(\mathrm{SMOKER3_i}) + \delta_2(\mathrm{GENHLTH_i})$$

C.2 SO_2

Model Term:

$$\beta_2 \cdot \mathrm{SOAve_i} + \eta_1(\mathrm{SOAve_i} \cdot \mathrm{SMOKER_i}) + \eta_2(\mathrm{SOAve_i} \cdot \mathrm{GENHLTH_i}) + \eta_3(\mathrm{SOAve_i} \cdot \mathrm{AGE_i}) + \eta_4(\mathrm{SOAve_i} \cdot \mathrm{SEX_i}) + \eta_5(\mathrm{SOAve_i} \cdot \mathrm{BMI_i})$$

Marginal Effect:

$$\frac{\partial \mathrm{logit}(\mathrm{P}(\mathrm{CASTHMA}=1))}{\partial \mathrm{SOAve_i}} = \beta_2 + \eta_1(\mathrm{SMOKER_i}) + \eta_2(\mathrm{GENHLTH_i}) + \eta_3(\mathrm{AGE_i}) + \eta_4(\mathrm{SEX_i}) + \eta_5(\mathrm{BMI_i})$$

C.3 Temperature

Model Term: $\beta_4 \cdot \mathrm{TEMPAve_i}$

Marginal Effect: $\frac{\partial \mathrm{logit}(\mathrm{P}(\mathrm{CASTHMA}=1))}{\partial \mathrm{TEMPAve_i}} = \beta_4$

C.4 Sex

Model Term: $\gamma_5 \cdot \mathrm{SEX_i} + \eta_4(\mathrm{SOAve_i} \cdot \mathrm{SEX_i})$

Marginal Effect: $\frac{\partial \mathrm{logit}(\mathrm{P}(\mathrm{CASTHMA}=1))}{\partial \mathrm{SEX_{CNTY_i}}} = \gamma_5 + \eta_4(\mathrm{SOAve_i})$

C.5 Smoking Status

Model Term: $\gamma_3 \cdot \mathrm{SMOKER_i} + \delta_1(\mathrm{PMAve_i} \cdot \mathrm{SMOKER3_i}) + \eta_1(\mathrm{SOAve_i} \cdot \mathrm{SMOKER_i})$

Marginal Effect: $\frac{\partial \mathrm{logit}(\mathrm{P}(\mathrm{CASTHMA}=1))}{\partial \mathrm{SMOKER3_i}} = \gamma_3 + \delta_1(\mathrm{PMAve_i}) + \eta_1(\mathrm{SOAve_i})$

C.6 Age

Model Term: $\gamma_2 \cdot \mathrm{AGE_i} + \eta_3(\mathrm{SOAve_i} \cdot \mathrm{AGE_i})$

Marginal Effect: $\frac{\partial \mathrm{logit(P(CASTHMA=1))}}{\partial \mathrm{AGE_{G_i}}} = \gamma_2 + \eta_3(\mathrm{SOAve_i})$

C.7 General Health Status

Model Term: $\gamma_4 \cdot \mathrm{GENHLTH_i} + \delta_2(\mathrm{PMAve_i} \cdot \mathrm{GENHLTH_i}) + \eta_2(\mathrm{SOAve_i} \cdot \mathrm{GENHLTH_i})$

Marginal Effect: $\frac{\partial \mathrm{logit(P(CASTHMA=1))}}{\partial \mathrm{GENHLTH_i}} = \gamma_4 + \delta_2(\mathrm{PMAve_i}) + \eta_2(\mathrm{SOAve_i})$

C.8 SO_2 : Age

Model Term: $\eta_3(\mathrm{SOAve_i} \cdot \mathrm{AGE_i})$

Marginal Effect: $\frac{\partial \mathrm{logit(P(CASTHMA=1))}}{\partial (\mathrm{SOAve_i} \cdot \mathrm{AGE_i})} = \eta_3(\mathrm{SOAve_i})$

Appendix D. Robustness Checks (Chapter 2)

To ensure the robustness of the asthma analysis in Chapter 2, I conducted a series of diagnostic and validation tests across both logistic and multinomial logistic models.

D.1 Logistic Regression Model

Multicollinearity: Variance Inflation Factor (VIF) values were generally low for all variables (VIF < 2), except for BMI, height, and weight, which exhibited natural correlation due to their physiological relationship.

Fixed Effects Robustness: To control for unobserved heterogeneity across time and location, I included fixed effects at the year, month, day, state, and county levels.

Heteroskedasticity: A Breusch-Pagan test suggested the presence of heteroskedasticity ($p < 0.05$). To address this, I employed cluster-robust standard errors clustered at the county × date level, accounting for spatial and temporal correlation in errors.

Endogeneity: To assess potential endogeneity in pollution exposure ($PM_{2.5}$ and SO_2), I used wind speed as an instrument. First-stage results confirmed wind speed as a strong instrument,(coefficient: -0.73570 (0.00094) ***, 0.47641 (0.00838)***), the F-statistics: 256.3 and 724, much bigger than 10, and second-stage results indicated that the original logistic model is not significantly biased by endogeneity.

D.2 Multinomial Logistic Regression Model

Multicollinearity: All VIF values were extremely low (< 0.001), suggesting no significant multicollinearity issues.

Fixed Effects Robustness: Due to model constraints, I included only state-level fixed effects to account for spatial differences.

Endogeneity: Wind speed was also used as an instrument for pollution variables ($PM_{2.5}$ and SO_2), confirming consistency of results. The coefficient: -0.01334 (0.00064)*** and 0.01654 (0.00054)***, the F-statistics: 272.1 and 904.6, much bigger than 10. The second stage shows there’s no potential endogeneity problem since the results is insignificant.

Appendix E. Dose-Response Function Specifications and Marginal Effects (Chapter 3)

The outcome variable is the hourly number of crashes, modeled using a Poisson regression with a log-link:

$$
\begin{aligned}
&\text{Crash}_i = \alpha_0 + \alpha_j + \beta_1 \text{Mean PM}_{2.5_i} + \beta_2 \text{Mean Visibility}_i + \beta_3 \text{Mean Temp}_i + \\
&\beta_4 \text{Mean Relative Humidity}_i + \gamma_1 \text{MeanAgeDummy}_i^{\text{One}<25} + \gamma_2 \text{MeanAgeDummy}_i^{\text{Both}<25} + \\
&\gamma_3 \text{MeanOverSpeed}_i^{\text{One OS}} + \gamma_4 \text{MeanOverSpeed}_i^{\text{Both OS}} + \gamma_5 \text{MeanDUI}_i^{\text{One DUI}} + \\
&\gamma_6 \text{MeanDUI}_i^{\text{Both DUI}} + \gamma_7 \text{MeanSexDummy}_i^{\text{Mixed}} + \gamma_8 \text{MeanSexDummy}_i^{\text{Both Female}} + \\
&\delta_1\left(\text{Mean PM}_{2.5_i} \cdot \text{MeanOverSpeed}_i^{\text{One OS}}\right) + \delta_2\left(\text{Mean PM}_{2.5_i} \cdot \right. \\
&\left. \text{MeanDUI}_i^{\text{One DUI}}\right)\delta_3\left(\text{Mean Visibility}_i \cdot \text{MeanOverSpeed}_i^{\text{One OS}}\right) + \delta_4\left(\text{Mean Visibility}_i \cdot \right. \\
&\left. \text{MeanDUI}_i^{\text{One DUI}}\right) + \delta_5\left(\text{Mean Temp}_i \cdot \text{MeanOverSpeed}_i^{\text{One OS}}\right) + \delta_6\left(\text{Mean Temp}_i \cdot \right. \\
&\left. \text{MeanDUI}_i^{\text{One DUI}}\right) + \delta_7\left(\text{Mean Relative Humidity}_i \cdot \text{MeanAgeDummy}_i^{\text{One}<25}\right) + \\
&\delta_8\left(\text{Mean Relative Humidity}_i \cdot \text{MeanDUI}_i^{\text{One DUI}}\right)
\end{aligned}
$$

Variable Notes:

Crash: Number of traffic crashes in county j at time t;

α_j: County fixed effects

$X^{(k)}$: Environmental variables:

$\text{MeanPM}_{2.5i}$, MeanVisibility_i, MeanTemp_i, $\text{MeanRelativeHumidity}_i$: Air pollution and weather measures including visibility, temperature, and relative humidity.

$Z^{(l)}$: Driving behavior dummy variables (8 in total, two for each category):

Age Dummies (Control group: Neither driver <25 years old):

$\text{MeanAgeDummy}_i^{\text{One}<25}$: One driver <25 years old

$\text{MeanAgeDummy}_i^{\text{Both}<25}$: Both drivers <25 years old

Sex Dummies (Control group: Both drivers are male):

$\text{MeanSexDummy}_i^{\text{OneFemale}}$: One driver is female

$\text{MeanSexDummy}_i^{\text{BothFemale}}$: Both drivers are female

Overspeeding Dummies (Control group: Neither driver overspeeding):

$\text{MeanOverSpeed}_i^{\text{OneOS}}$: One driver overspeeding

$\text{MeanOverSpeed}_i^{\text{BothOS}}$: Both drivers overspeeding

DUI Dummies (Control group: Neither driver DUI):

$\text{MeanDUI}_i^{\text{OneDUI}}$: One driver DUI

$\text{MeanDUI}_i^{\text{BothDUI}}$: Both drivers DUI

$X^{(k)} \cdot Z^{(l)}$: 6 interaction terms between each environmental variable and dummy variables

E.1 $PM_{2.5}$

Model Term: $\beta_1 \text{Mean PM}_{2.5_i}$

Marginal Effect: $\frac{\partial \text{Crash}_i}{\partial \text{Mean PM}_{2.5_i}} = \beta_1$

E.2 Visibility

Model Term:

$$\beta_2 \text{Mean Visibility}_i + \delta_1\left(\text{Mean Visibility}_i \cdot \text{MeanOverSpeed}_i^{\text{One OS}}\right) + \delta_2\left(\text{Mean Visibility}_i \cdot \text{MeanDUI}_i^{\text{One DUI}}\right)$$

Marginal Effect:

$$\frac{\partial \text{Crash}_i}{\partial \text{Mean Visibility}_i} = \beta_2 + \delta_1(\text{MeanOverSpeed}_i^{\text{One OS}}) + \delta_2(\text{MeanDUI}_i^{\text{One DUI}})$$

E.3 Relative Humidity

Model Term:

$$\beta_4 \text{Mean Relative Humidity}_\text{i} + \delta_5\big(\text{Mean Relative Humidity}_\text{i} \cdot \text{MeanAgeDummy}_\text{i}^{\text{One<25}}\big) + \delta_6\big(\text{Mean Relative Humidity}_\text{i} \cdot \text{MeanDUI}_\text{i}^{\text{One DUI}}\big)$$

Marginal Effect:

$$\frac{\partial \text{Crash}_\text{i}}{\partial \text{Mean Relative Humidity}_\text{i}} = \beta_4 + \delta_5(\text{MeanAgeDummy}_\text{i}^{\text{One<25}}) + \delta_6(\text{MeanDUI}_\text{i}^{\text{One DUI}})$$

E.4 Temperature

Model Term:

$$\beta_3 \text{Mean Temp}_\text{i} + \delta_3\big(\text{Mean Temp}_\text{i} \cdot \text{MeanOverSpeed}_\text{i}^{\text{One OS}}\big) + \delta_4\big(\text{Mean Temp}_\text{i} \cdot \text{MeanDUI}_\text{i}^{\text{One DUI}}\big)$$

Marginal Effect:

$$\frac{\partial \text{Crash}_\text{i}}{\partial \text{Mean Temp}_\text{i}} = \beta_3 + \delta_3(\text{MeanOverSpeed}_\text{i}^{\text{One OS}}) + \delta_4(\text{MeanDUI}_\text{i}^{\text{One DUI}})$$

Appendix F. Robustness Checks (Chapter 3)

To ensure the validity and robustness of our empirical results regarding the effect of particulate matter on traffic accidents, we perform a series of robustness checks for the three main models (OLS, Poisson, and Negative Binomial). These checks address several potential statistical concerns, including linearity, heteroskedasticity, endogeneity, and overdispersion, as well as the appropriateness of fixed effects and standard errors.

F.1 OLS Model Robustness

We first assess the robustness of the linear regression model:

Linearity Test: A nonlinearity test rejects the null hypothesis of a linear relationship ($p < 2.2e-16$), indicating potential model misspecification under strict linearity assumptions.

Heteroskedasticity Test: The Breusch–Pagan test reveals strong evidence of heteroskedasticity ($p < 2.2e-16$), suggesting that standard OLS inference may be biased without adjustment.

Fixed Effects Robustness: We include county-level fixed effects to control for unobserved time-invariant regional factors. The inclusion of factor(county) does not change the key coefficient signs or significance levels.

Endogeneity: To address potential reverse causality or omitted variable bias related to pollution levels, we instrument $PM_{2.5}$ using wind speed. The instrumental variable (IV) approach confirms the direction and statistical strength of the pollution effects. During the first stage, the coefficient of instrumental variable is -0.39261 (0.01627)***, F-statistic is 64.52 >10, showing it's a strong IV.

Clustered Standard Errors: All standard errors are clustered at the county-date level to account for correlated shocks within time and space.

F.2 Poisson Model Robustness

Given that the dependent variable (traffic accident count) is non-negative and discrete, we also apply Poisson regression:

Overdispersion Test: The variance of the outcome greatly exceeds its mean ($Var(Y) >> E(Y)$, overdispersion statistic > 8.7), which violates the Poisson assumption of equidispersion.

Heteroskedasticity: The model exhibits significant heteroskedasticity ($p < 2.2e-16$), but the results remain robust under cluster-robust standard errors.

Fixed Effects: County-level fixed effects are included to absorb unobserved regional heterogeneity.

Clustered Standard Errors: Errors are clustered at the county-date level to ensure valid inference under grouped data.

Endogeneity: I used hourly wind speed and daily wind speed as IV for two Poisson models. For the hourly Poisson model, coefficient of IV is -0.71967 (0.01695)***, and F-statistic is $245.6 > 10$ in the first stage; for the daily Poisson model, coefficients of IV is -0.26413 (0.00591)***, F-statistic is $490.2 > 10$ in the first stage, both showing it’s a strong IV.

F.3 Negative Binomial Model Robustness

The negative binomial model is designed to accommodate overdispersion by relaxing the assumption that $Var(Y) = E(Y)$. Our robust results include:

Overdispersion Check: The estimated overdispersion parameter confirms that Var(Y) >> E(Y), justifying the use of the negative binomial over Poisson.

Heteroskedasticity: Tests again reveal heteroskedasticity ($p < 2.2e-16$), but inference remains robust with clustered standard errors.

Fixed Effects: We include county fixed effects (factor(county)), and the key estimates remain consistent.

Endogeneity: Wind speed is used as an instrument for pollution variables, and the IV-negative binomial results confirm a significant pollution-accident relationship. In the first stage, the coefficient of Instrument variable is -0.26413 (0.00591)* in daily negative binomial model and -0.71967 (0.01695)*** in hourly negative binomial model. F-statistics are 490.2 and 245.6 individually, both are over 10, showing that wind speed is a strong instrument variable for $PM_{2.5}$.

Clustered Standard Errors: Standard errors are clustered at the county-date level throughout.