Title
Three Essays on Environmental Economics

Author
Jiang, Zhiyun

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Three Essays on Environmental Economics

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Economics

by

Zhiyun Jiang

Committee in charge:

Professor Richard T. Carson, Chair
Professor Judson Boomhower
Professor Jennifer Burney
Professor Mark Jacobsen
Professor Katharine Ricke

2022

The dissertation of Zhiyun Jiang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my family.

TABLE OF CONTENTS

Dissertation Approval Page ... iii

Dedication ... iv

Table of Contents ... v

List of Figures ... vii

List of Tables ... ix

Acknowledgements ... xi

Vita ... xii

Abstract of the Dissertation ... xiii

Chapter 1 Impact of Smoke from Fires on Agriculture ... 1
- 1.1 Introduction ... 2
- 1.2 Literature Review ... 6
- 1.3 General Modelling Strategy and Data ... 8
  - 1.3.1 Agriculture Data ... 10
  - 1.3.2 Smoke Plume and Fire Location Data ... 11
  - 1.3.3 Weather Data ... 16
  - 1.3.4 Summary Statistics ... 16
- 1.4 Empirical Specification ... 17
- 1.5 Results ... 19
  - 1.5.1 Main Results ... 19
  - 1.5.2 Differential Impacts ... 20
  - 1.5.3 Robustness Check ... 24
  - 1.5.4 Farmer's Responses ... 26
- 1.6 Economic and Policy Implications ... 29
  - 1.6.1 Economic Loss ... 29
  - 1.6.2 Wildfire Management Options ... 32
- 1.7 Conclusion ... 34
- 1.8 Appendix ... 36
- 1.9 References ... 46

Chapter 2 The Interaction Relation of Temperature and Precipitation with Outdoor Recreation ... 51
- 2.1 Introduction ... 52

2.2 Literature Review 54
2.3 Conceptual Framework and Data 56
2.4 Empirical Strategy 59
2.5 Results 61
2.5.1 Main results 61
2.5.2 Results for Different Trip Types and Individual Groups 67
2.5.3 Robustness Check 75
2.6 Conclusion 76
2.7 Appendix 78
2.8 References 87

Chapter 3 The Role of Ridesharing in Changing Urban Trip Patterns 89
3.1 Introduction 90
3.2 Literature Review 92
3.3 Conceptual Framework and Data 95
3.4 Empirical Strategy 100
3.5 Results 102
3.6 Conclusion 108
3.7 Appendix 110
3.8 References 120

LIST OF FIGURES

Figure 1.1: Smoke Days in 2008 ... 14

Figure 1.2: Smoke Days in 2019 ... 15

Figure 1.3: Average Smoke Days ... 36

Figure 2.1: Histogram of Maximum Temperature and Total Rain ... 58

Figure 2.2: Response Surfaces from Binning ... 62

Figure 2.3: Response Surfaces from GAM ... 63

Figure 2.4: Cross Sections of Response Surface at Different Levels of Total Rain ... 65

Figure 2.5: Cross Sections of Response Surface at Different Maximum Temperature ... 66

Figure 2.6: Response Surfaces for Different Types of Trips ... 68

Figure 2.7: Response Surfaces for Male and Female ... 70

Figure 2.8: Response Surfaces for Different Life Stages ... 71

Figure 2.9: Response Surfaces for Individuals with Car(s) or Not ... 73

Figure 2.10: Response Surfaces for Individuals with Dog(s) or Not ... 74

Figure 2.11: Histogram for Maximum Temperature and Total Rain During Daylight ... 78

Figure 2.12: Response Surface with Total Rain and Snow from GAM ... 79

Figure 2.13: Response Surface for Observations Without Snow ... 80

Figure 2.14: Response Surface with Has Visits as Outcome ... 81

Figure 2.15: Response Surface following Negative Binomial Regression ... 82

Figure 2.16: Response Surface Using Year-Quarter Fixed Effects ... 83

Figure 2.17: Response Surface Using Year and Month Fixed Effects ... 84

Figure 3.1: Number of Trips and Entry Length from NHTS 2017 ... 99

Figure 3.2: Change in Rate of Taxi Driver Out of Total Employment vs. Entry Time ... 110

Figure 3.3: Percentage Change in Bus Ridership against Entry Time ... 111

Figure 3.4: Percentage Change in Rail Ridership against Entry Time ... 112

Figure 3.5: Number of Trips Over Surveys ... 113

Figure 3.6: Number of Taxi/Rideshare Trips Predicted by Different Functional Forms of Entry Length ... 114

LIST OF TABLES

Table 1.1: Summary Statistics 17

Table 1.2: Main Results for Yields of Corn and Soybeans 20

Table 1.3: Results from Different Stages in Growing Season 22

Table 1.4: Results from Different Smoke Density 22

Table 1.5: Results from Smoke Days Interacting with County Income Level 23

Table 1.6: Results from Smoke Days Interacting with County Farm Size Distribution 24

Table 1.7: Results of Farm Production Expenses for All Counties 27

Table 1.8: Results of Hired Farm Labor for All Counties 29

Table 1.9: Per Acre Cost Comparison for Fuel Management 33

Table 1.10: Results for Yields Dropping Southern States 37

Table 1.11: Results for Yields Dropping Observations Where There is a Fire in the County .. 38

Table 1.12: Results for Yields Regressing on Smoke Residuals 39

Table 1.13: Results for Yields with Controls of Growing Degree Days and Precipitation Interacting with Irrigation Status 40

Table 1.14: Results for Yields with Alternative Fixed Effects and Time Trends 41

Table 1.15: Results for Level of Yields of Corn and Soybean 42

Table 1.16: Results for Yields at Different Smoke Coverage over Crop Fields 43

Table 1.17: Results of Farm Production Expenses for Corn or Soybean Counties 44

Table 1.18: Results of Hired Farm Labor for Corn or Soybean Counties 45

Table 2.1: Summary Statistics 59

Table 2.2: Summary Statistics of Individual Control Variables 85

Table 3.1: Summary Statistics for Number of Trips and Entry Length 98

Table 3.2: Impacts of Entry Length on Number of Trips in Different Modes 103

Table 3.3: Impacts of Entry Length on Public Transit 104

Table 3.4: Impacts of Entry Length on Average Daily Rideshare 105

Table 3.5: Impacts of Entry Length on Number of Trips in Different Modes and Distances 106

Table 3.6: Impacts of Entry Length on Number of Trips in Different Modes and Starting Time 107

Table 3.7: Summary Statistics for Control Variables 115

Table 3.8: Change in Number of Trips as % of Weighted Mean 116

Table 3.9: Impacts of Entry Length on Number of Trips in Different Functional Forms 117

Table 3.10: Impacts of Entry Length on Number of Trips Using Count Data Model 118

Table 3.11: Impacts of Entry Length on Number of Trips in Different Modes Using Year-Month Fixed Effects 119

ACKNOWLEDGEMENTS

I would like to thank my advisor, Richard T. Carson, for all the guidance and support throughout my PhD life. I am deeply grateful for the considerable time of him guiding me in developing my research, as well as providing comments and suggestions in writing of this dissertation. I would also like to thank my committee members Judson Boomhower, Jennifer Burney, Mark Jacobsen, and Katharine Ricke for the feedback and suggestions to help improve my research.

I would like to thank my family for their continuous support and encouragement throughout the years. I am also thankful to my friends, in and outside of UCSD, for their help and support in this journey.

VITA

2012 Bachelor of Science, University of Wisconsin-Madison

2014 Master of Science, Stanford University

2022 Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Three Essays on Environmental Economics

by

Zhiyun Jiang

Doctor of Philosophy in Economics

University of California San Diego, 2022

Professor Richard T. Carson, Chair

This dissertation studies three distinctive aspects of environmental economics. Chapter 1 examines the impact of smoke from fires on agriculture production of the two main U.S. cash crops: corn and soybeans. Linking smoke plume maps derived from satellite images with county-level information on corn and soybean yields, I use a panel data approach to estimate exposure to smoke plumes treating their exact frequency, timing, and location in any year as exogenous shocks. Exposure to one more day of smoke, on average, reduces yields of corn and soybeans by 0.31% and 0.23%, respectively. To help put these results in an economic context for corn and soybeans, a 10% increase in smoke relative to 2019 results in an annual loss of almost $1 billion.

Chapter 2 explores the interaction relation of temperature and precipitation with number of outdoor recreation trips. Using detailed information on outdoor recreation trips in England over a four-year period, I use a semi-parametric response surface approach to examine the interaction relation. I found that although daily visits increase with temperature and decrease with rain, these gradients only have small variations across rain or temperature. Interaction of the two variables plays a small role in outdoor recreation.

Chapter 3 examines how the introduction of ridesharing services such as Uber and Lyft into the U.S. urban market influences trip choice decisions. Using data from the 2009 and 2017 National Household Travel Surveys, I show that the longer Uber and Lyft have been in an urban market, the greater the increase in the 2017 survey trips that were made using taxi/rideshare services relative to the 2009 survey benchmark. This increase is driven by an upward shift that is more pronounced for short and longer distance trips than for middle distance trips and is also more pronounced on weekdays relative to weekends. Ridesharing services are shown to be a substitute for short haul bus trips, but a complement with longer rail trips.

Chapter 1

Impact of Smoke from Fires on Agriculture

Abstract: With wildfires projected to substantially increase under climate change, there is growing interest in understanding their economic cost. While wildfires directly burn forests and homes, they also produce smoke that can cover wide areas and travel long distances, physiologically harming people and plants. These injuries have economic consequences. This paper provides the first national economic estimates of the impact of this smoke on the production of the two main U.S. cash crops: corn and soybeans. To do this, I link smoke plume maps derived from satellite images with county-level information on corn and soybean yields. A panel data approach is used to estimate exposure to smoke plumes treating their exact frequency, timing, and location in any year as exogenous shocks. This allows separation of smoke impacts from the better studied air pollutant impacts, which is important because, while pollution from vehicles and power plants is falling, that from smoke is increasing. Exposure to one more day of smoke, on average, reduces yields of corn and soybeans by 0.31% and 0.23%, respectively. Using these estimated yield response functions for policy purposes requires specific future smoke generation

scenarios. To help put these results in an economic context for corn and soybeans, a 10% increase in smoke relative to 2019 results in an annual loss of almost $1 billion. Some climate change scenarios for 2050 involve considerable larger increases in U.S. smoke levels. If other U.S. agricultural production is equally sensitive to smoke and, if the findings reported here hold for other countries with wildfires, then total agricultural losses from likely increase in smoke over the next 30 years through 2050 are likely to be on order of several hundred billion dollars. The estimates provided here are likely to be of use in consideration of policies to reduce wildfires. As an example, I consider the debate over mechanical removal of fuel wood versus prescribed burning to reduce wildfire risks. Taking account of impacts of prescribed burning smoke on corn and soybean yields can shift the preferred option from a benefit-cost perspective.

1.1 Introduction

The U.S. has been experiencing record-breaking wildfire events, with large economic damages occurring from loss of structures and, in some instances, life. Wildfires have always existed as part of the ecosystem. They have been increasing over time in acreage burned (Burke et al., 2007) for reasons related to fire suppression efforts and the increasing encroachment of urban development into areas of high fire risk (Radeloff et al., 2018). Climate change is now further exacerbating the wildfire activities. For example, the work of Abatzogloua and Williams (2016) shows climate change contributed to about half of the total burned area by wildfires in Western U.S. between 1984 to 2015. Projection shows that the frequency of wildfires and length of fire season will increase on roughly three-quarters of the world's land areas by the end of this century, with the U.S. being one of the countries likely to experience the most substantial impacts (Sun et al., 2019).

With concern over increasing wildfires under climate change, there is growing interest in understanding their full economic costs. These costs go beyond the dramatic destruction of homes media audience see and the direct loss of life. Wildfires also generate enormous amounts of smoke. Smoke can travel for long distances, with its reach extending for thousands of miles (Miller et al., 2017). The fires themselves can last for days or even weeks. Smoke from such fires hangs in the air even longer. There is a growing literature, mostly in the biological sciences, looking at different physiological impacts of smoke. Many of these seek to understand underlying mechanisms or look at impacts associated with specific wildfire events. Moving in the policy direction, attention has largely been directed at health effects on farm workers and local residents (Reid et al., 2016; Vo et al., 2021). Here I look at the impact of wildfires on agriculture demonstrating that smoke from wildfires can reduce crop yields. To the best of my knowledge, this paper provides the first national estimates of the economic damage done by smoke to the two major U.S. cash crops: corn and soybeans.

Surprisingly, the intersection between wildfires and agriculture production yields has been little explored by economists. Work on prescribed burning, which deliberately sets fires in favorable weather conditions, to reduce later wildfire risk (and hence creates smoke similar to wildfires), often mentions possible local health effects and public opposition (Mercer et al., 2007; Florec et al., 2020), but has not considered the possibility of substantial yield impacts on major crops across the entire United States.

In contrast, the role of air pollutants in causing economic damage to major U.S. crops has been well studied (e.g., Garcia et al., 1986; Westenbarger and Frisvold, 1995; Boone et al., 2019). Turning to smoke, the difficulty comes in separating the effect of air pollution from all sources (measured at U.S. Environmental Protection Agency (EPA)'s monitoring stations) and from that of smoke, whose chemical components evolve in complex atmospheric reactions, into various pollutants such as ozone that are known to reduce crop yields. Solving this problem of separating the impact of smoke from air pollution from all sources requires a different data source. I use smoke plume maps derived from satellite images for this purpose.

I link this satellite smoke plume imaging and fire location with county-level information on agricultural yields and production expenses. Using variation in exposure to smoke plumes as exogenous shocks providing the source of identification, I find that exposure to an additional day of smoke on average decreases corn and soybean yields by 0.31% and 0.23%, respectively. The estimation coefficients are reasonably robust to range of alternative specifications.

For policy purposes, the dose response relationships identified in this paper need to be coupled with projections of wildfires and smoke dispersion under different climate change scenarios. That is beyond the scope of this paper. Nevertheless, it is likely to be useful to cast results in terms of economic loss rather than yield reductions. I do this by considering the annual loss from a 10% increase relative to 2019 smoke levels, which is well within current year to year variation. For corn, this loss estimate is just under $600 million and for soybeans almost $300 million. This 10% change is small relative to projected increases in wildfires in many regions in the contiguous U.S., where some estimates range up to 75% by mid-century (2040-2069)[1] and the greatest increases are in the areas where corn and soybean production is concentrated (Gao et al., 2021). In the absence of other information, a reasonable assumption is that yields of other U.S. crops are as sensitive to smoke as corn and soybeans and that smoke from wildfires in other countries have similar impacts. Corn and soybeans are well studied crops, where switching the particular varietal grown, as well as crop switching behavior in response to pollution exposure

[1] The 75% increase is compared with baseline period (1971–2000) in Gao et al. (2021).

have long been recognized (Griliches, 1957; Kopp et al., 1985). Because of corn and soybeans' prominence, comprising almost half of all U.S. total crop cash receipts, research at agricultural experiment stations and commercial entities aimed at implicit adaptation to smoke is likely to be more advanced than other crops and it is less likely to think that other commercial crops are less sensitive to smoke. Smoke may also impact agriculture output in other major agricultural producers such as Australia, Brazil, Canada, China, European Union, India, Russia, and Turkey. These producers are among the top 30 in terms of crop production value in the world (Food and Agriculture Organization of the United Nations, 2018) with the U.S. output comprises about 8% of the world. These regions are projected to have increases in wildfires in the future and some may experience substantial impacts (Sun et al., 2019). Referring to 2050 as short hand for mid-century and based on the assumptions that smoke also affect other crops and other countries, agricultural losses from the likely increase in wildfire-related smoke could be in the several hundred-billion-dollar range cumulated to 2050.

My analysis strategy also allows me to decompose the smoke days over three phases of the growing season (i.e., planting, cultivation, and harvest). Smoke impacts are concentrated in cultivation and, to a lesser significance, the planting stages. More detailed smoke data available for the later part of my sample period suggests that heavier smoke levels reduce yields more. I also look at whether farmers adapt to the smoke. Here, farmers respond to more smoke days by applying more agricultural chemicals and fertilizer.

My results may also be helpful to a contentious debate over deploying measures to reduce wildfire risks by reducing the fuel wood (e.g., brush and dead timber) that increase the intensity of wildfires and help direct their path. This debate, fundamentally different from climate consideration and building on the urban wildland interface, focuses on prescribed burning versus mechanical removal. The negative smoke externality effect on corn and soybean yields from prescribed burning follows from my modeling as it generates similar type of smoke wildfires do. Based on a simplified calculation, adding in these economic losses implies that mechanical removal from a welfare standpoint is likely to be the preferred option even though it is generally several times more expensive from the immediate financial outlay perspective of a local, state or federal agency. Adding in the monetary value of health effects only strengthen this case.

1.2 Literature Review

The impact of smoke from fires rises from the air pollutants it contains. The main suspects are those pollutants already known to do harm to human health and plants: particulates and ozone. Particulate matter is one of the most obvious air pollutants emitted and the one known to do the most harm to human health. According to the 2017 National Emissions Inventory (NEI) Data from U.S. EPA, wildfires are major source of $PM_{2.5}$ contributing to almost 30% of primary $PM_{2.5}$ emissions. U.S. EPA Air Pollutant Emissions Trends Data indicate that while contribution of anthropogenic sources to $PM_{2.5}$ in recent years have been decreasing, the contribution of wildfires has been increasing. Ford et al. (2018) predicts that, under some climate change scenarios, by the end of the century, emission from fires will contribute to more than half of annual $PM_{2.5}$ in the contiguous U.S. states. Wildfires also emit over 10% of primary PM_{10} according to the U.S. EPA 2017 NEI Data.

The science literature has looked at the impact of smoke from fires on plant productivity by studying different mechanisms. Particulates can both absorb and diffuse light. On the absorption side, aerosols from fires can reduce total radiation reaching plants and reduce photosynthesis (Park et al., 2018; Yamasoe et al., 2005). On the other hand, in some instances diffusion of solar radiation due to particulate matter from fires appears to increase plant productivity (Park et al., 2018; Yamasoe et al., 2005). Studies have found mixed results for the overall impacts of particulate matters from fires under different scenarios or geographic scales (Hemes et al., 2020; Yue and Unger, 2018).

Less obvious wildfires emit a host of other less visible chemicals. Some of these are known to be ozone "precursors". Ozone, a colorless gas long known to be harmful to plants (Krupa and Manning, 1988). It is both toxic to plants and reduces photosynthesis (Ainsworth et al., 2012; Yue and Unger, 2018). Ozone is a not a pollutant directly released by combustion. Instead, it forms as the result of complex atmospheric chemistry that converts nitrogen oxides (NOx) and non-methane organic carbons (NMOC) into ozone under sunlight (Jaffe and Wigder, 2012). Importantly, ozone levels can be highly variable over time and space (Ainsworth et al., 2008) and, in this sense, are the antithesis of carbon dioxide, a pollutant characterized by uniform mixing. This interacts with the much sparser pollution monitoring network in rural areas to make quantifying how substantive the role ozone plays in reducing crop yields more difficult. In contrast to particulates, controlled plant experiments show that increasing ozone concentrations are detrimental to plants. A question following such work is how harmful is ozone to particular crops under field conditions. Real world agricultural production can be affected by farmer decisions.

Economists have looked at agriculture and air pollution for a long time. Earlier papers find expected negative, and substantial, harmful relationship between ozone and agricultural production (Garcia et al., 1986; Westenbarger and Frisvold, 1995). In more recent work, Boone et al. (2019) provides evidence of non-linear impacts of ozone on corn production. Da et al. (2021) show effects of ozone and climate conditions on a set of crops. Zhou et al. (2018) show significant negative impacts of $PM_{2.5}$ on corn (and wheat) yields in China. There have been papers more generally look into climate change and agriculture (Schlenker and Roberts, 2009; Deschênes and Greenstone, 2007). However, the specific link between agriculture yields and smoke from wildfires has not been explored by economists.

Outside of agriculture, there is a small but growing literature in economics and epidemiology aiming to understand impacts of smoke from fires on other sectors. The most studied area is health. Jayachandran (2009) received widespread attention by showing that Indonesia large wildfires were causally linked to early childhood mortality. A number of other studies have shown linkages between wildfires and respiratory morbidity and mortality (Reid et al., 2016; Miller et al., 2017). Richardson et al. (2012) look specifically at the health impacts of western wildfires on Los Angeles residents. A recent study looking into labor market shows smoke from fires reduce earnings and a third of those earning losses can be explained by employment loss (Borgschulte et al., 2020). Lastly, there is work on how individuals respond to air pollution. Most of this literature is related to pollution avoidance, such as increasing face masks purchases or reduce outdoor activities to reduce pollution exposure (Zhang and Mu, 2018; Neidell, 2009). Individuals also increase medication expenditures as defensive investments (Deschênes et al., 2017). There is also specific work on how people engage in various types of avoidance behavior in response to wildfires (Richardson et al., 2012; Santana et al., 2021). My contribution to this literature will be to look at how farmers receiving smoke from wildfires respond after that shock has been received.

1.3 General Modelling Strategy and Data

The general modeling strategy employed here will assume that the number of acres harvested with the crop of interest is observed and, as is eventually, average yield across those acres. I can observe farms aggregated at the county level. This leads me to concentrate on corn and soybeans, the two crops with the largest quantity of planted acreage. According to U.S. Department of Agriculture (USDA) statistics in 2019, corn and soybean together accounted for more than 40% of total U.S. crop cash receipts. Both have reasonably wide county-level distribution but clear concentrations that are not in areas where there is substantial direct fire risk to burning fields. Corn and soybeans are also among the U.S. major agricultural exports. There is an effective world price, after taking tariffs and transportation costs, making the assumption of even a large corporate agricultural operation in a county being firms in a competitive industry tenable.

The initial assumption is that farmers have taken their perceived probability distribution (conditional on currently available information) of being hit by wildfire impacts into account at the time their crop was planted. This distribution intersects with the same farmer's priors on the timing and magnitudes of different components of climate change. I abstract from these longer run considerations and focus on what happens after that decision has been made. This is done by using county-level fixed effects to control for those components that are idiosyncratic to location. Climate region specific fourth order yearly polynomial is used to remove region specific technological improvement in agriculture.

There are two potential selection effects. The first is that I exclude counties that are not producing corn or soybeans in any given year. This paper focus on what happens after planting when the wildfire shocks and smoke exposure are treated as exogenous from a farmer's perspective. My analysis can therefore not answer the question of whether the farmer would have made a different planting decision under a different baseline level of wildfire risk. The other is imposed by USDA data availability. Data for some individual counties or some states are not reported by USDA. These appear to be small fraction and available county level data accounts for about 80% or more of the total production.

Wildfire shocks can affect a location in two forms. One is through actual fire, where a local fire spread to crop land can lead to direct loss. Wildfires also generate smoke plumes which float through space and time. The specification of this stimulus variable will be discussed further in later sections, whose impact on corn and soybean yields will be the target of my primary focus. Yield is then regressed on the two wildfire stimulus variables, the local binary indicator, and a measure of how many days there had been smoke over the crop land in the county. In between, weather takes its influential course and weather variables are also controlled. This provides an estimate of how a change in the number of days of smoke at a location changes yield.

Implicitly included in this estimate are actions taken by farmers after the wildfire shocks and smoke exposure. Since total annual expenditures can be disaggregated to several broad categories, it is possible to look at differences across years with appropriate controls. The rest of this section describes where particular variables come from. Some of the issues involved in assembling the dataset used for analysis are also discussed.

1.3.1 Agriculture Data

Agricultural yield is the main outcome variable. Yields of corn and soybeans are obtained from the USDA's National Agricultural Statistics Service (NASS) Survey. County-level yield data was collected each year in the study period of 2006 to 2019. Yields are in unit of bushels per acres, representing bushels of production divided by acres of area harvested.

In addition to yield data, I collect county-level data of farm production expenses, farmland areas, and number of hired farm workers through the USDA's Census of Agriculture. Since the Census of Agriculture is quinquennial, these data are only available for three years within the study period, in 2007, 2012 and 2017. Farm production expenses includes all farm related expenses such as agricultural chemicals, fertilizer, fuel costs, hired labor costs, and etc. These expenses are not only for the production of corn and soybeans, the dominant crops produced in most of the counties included in the analysis, but also for other crops grown and livestock operations. The Census of Agriculture provides total production expenses, as well as expenses broken out by particular categories. Some of these are reasonably assignable time-wise to a particular phase of the growing season such as agricultural chemicals and fertilizer. Farm land area within a county is the acreage designated as land in farms by USDA, used for crops or grazing. Production expenses are divided by farmland areas in acres, to create variables of per acre expenses. These are used as outcome variables to when examining how farmers respond to wildfire shocks. The Census of Agriculture also provides information on number of hired farm workers, which is defined to include full time and part-time workers, and paid family members, but excludes contract labor.

The USDA NASS Cropland Data Layer (CDL) is used to build up geographic information on crop and other agricultural land. It provides land use maps in 30-meter resolution. These maps contain different agricultural land cover types, including corn and soybean fields, land cover for other types of crops as well as for grassland/pastures. The CDL covers entire contiguous United States and goes back to 2008. For 2006 and 2007 in the study period, CDL 2008 is used as a proxy since changes at the annual level tend to be small. The CDL allows the spatial connection of crop fields, smoke plumes and fire locations.

1.3.2 Smoke Plume and Fire Location Data

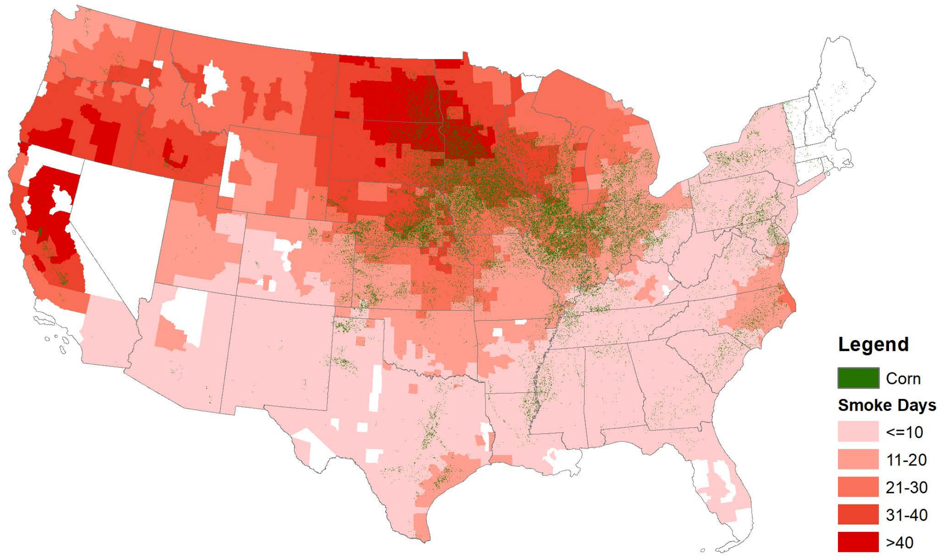
Smoke plumes and fire location data are collected from the U.S. National Oceanic and Atmospheric Administration (NOAA) Hazard Mapping System (HMS). HMS process information from multiple satellites to provide daily maps of smoke plumes and active fires. Smoke plume maps show polygons of smoke cover over North America. These are produced by analysis based on visual classification of plumes using satellite imagery (U.S. NOAA, 2021). Fire maps show points representing corresponding satellite pixels where a potential fire event was detected. These fire pixels are derived from automated fire detection process using satellite images, followed by quality control by expert image analysts (U.S. NOAA, 2021). For both smoke and fire data, HMS is not able to differentiate the source of fire between wildfire, prescribed burning, or agricultural burning. According to U.S. EPA 2017 NEI Data for criteria air pollutants, among the three types of fires, wildfires account for about 68% of these emissions, prescribed burning, 31%, and agricultural burning only about 1%. Thus, the smoke emissions of interest are largely from wildfires, followed by prescribed burning, whose smoke should be quite similar in chemical composition because it is the result of burning the same type of biomass.

The smoke plume data is used to construct the variable of smoke days, which is the variable of interest to measure smoke exposure. This is defined as number of days that a particular crop land in a county is entirely (100%) under a smoke plume in growing season in a year. This definition of smoke days is similar to that used in Borgschulte et al. (2020). Smoke plume maps are available at daily level. The CDL is available annually and corn or soybean land cover map can be obtained from this layer each year. For each day, the smoke plume map is overlayed with corn or soybean land cover map as well as the county boundary map. This allows to check whether corn or soybean land in a county is entirely under a smoke plume on that day. If so, that day can be counted towards the smoke days. Smoke days over the course of a year are constructed by summing all of smoke days during corn or soybeans growing season in that year. Growing seasons for crops are state specific, following the USDA usual planting dates (USDA, 2010).

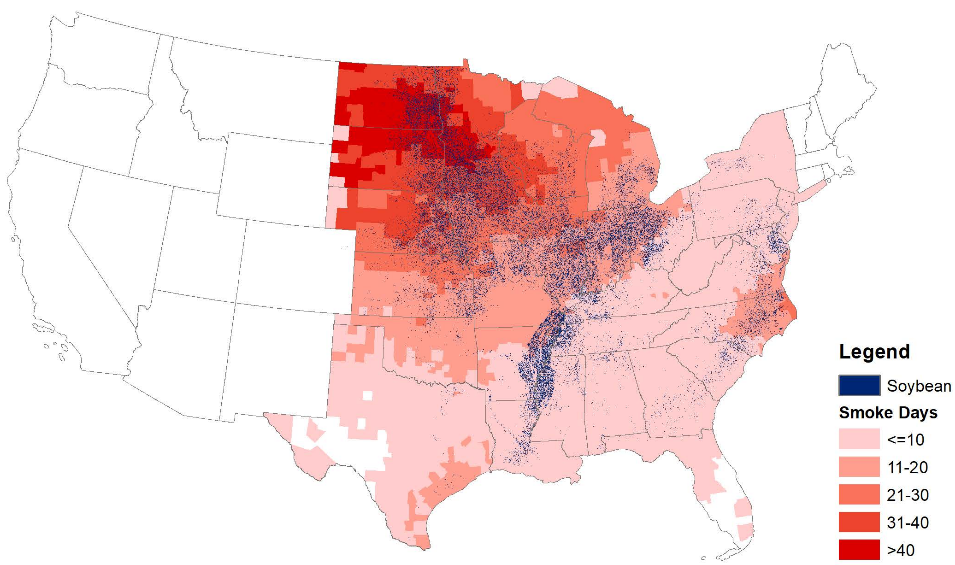
USDA usual planting dates provides begin and end dates for when crops are planted and harvested in most years. Begin dates are when planting or harvesting is at about 5% complete and end dates are at about 95% complete. Growing season in this paper is then defined from the begin dates of planting to the end dates of harvesting. This further allows me to divide the growing season into three stages, planting, cultivation and harvest. Planting and harvest stages follow the corresponding begin to end dates and the cultivation stage is the time in between.

Calculated smoke days for corn and soybeans for 2008 and 2019 are shown in maps in Figure 1.1 and Figure 1.2, where different red color corresponds to different ranges of smoke days. The white areas are either states that USDA does not usually collect survey data for corn or soybean, or when a particular county does not have corn or soybean land cover[2]. Figure 1.1 and Figure 1.2 show variation of exposure to smoke days for both crops across counties and years. 2008 represents a year with fewer smoke days exposure while 2019 represents a year with more smoke days exposure. Since corn and soybeans have similar geographic distribution, they also experience similar distribution of smoke days. I also calculate average number of smoke days for corn and soybean over the 14 years study period, shown in maps in Figure 1.3 in the Appendix. An indicator variable for local fire is also generated to account for whether there has been any active fire in corn or soybean land in a county in the growing season in a year. Similar to smoke, for each day, fire maps with points of active fire location are overlayed with corn or soybean land cover map and county boundary map. If for any day in the growing season in a year, there is fire point occur on corn or soybean land in a county, then, the fire indicator variable for that county for that year becomes 1.

[2] Calculation of smoke days are based on corn or soybean land cover. The regression sample for corn and soybean contains fewer number of counties than non-white counties shown in Figure 1.1 or Figure 1.2 as yield data is not reported by USDA for some counties.



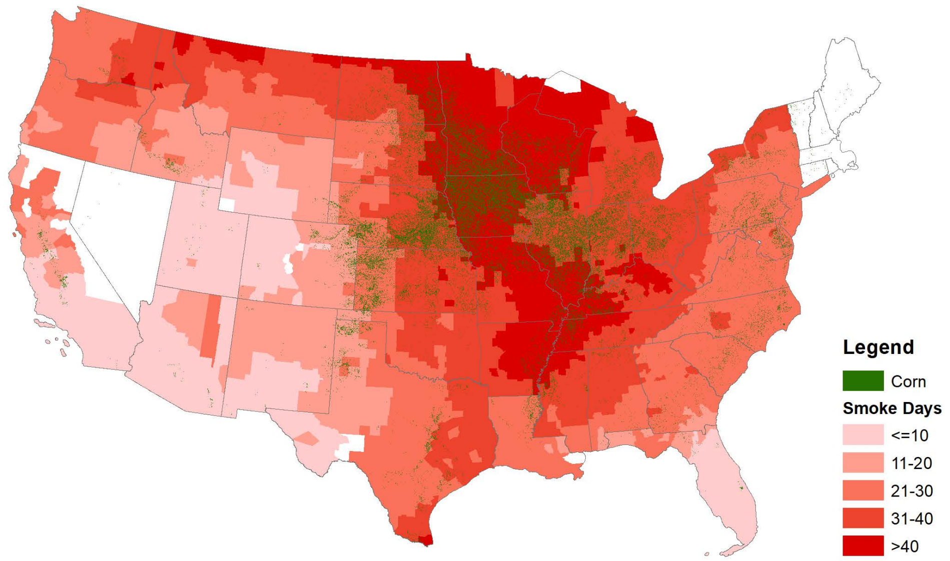
(a) Smoke Days for Corn



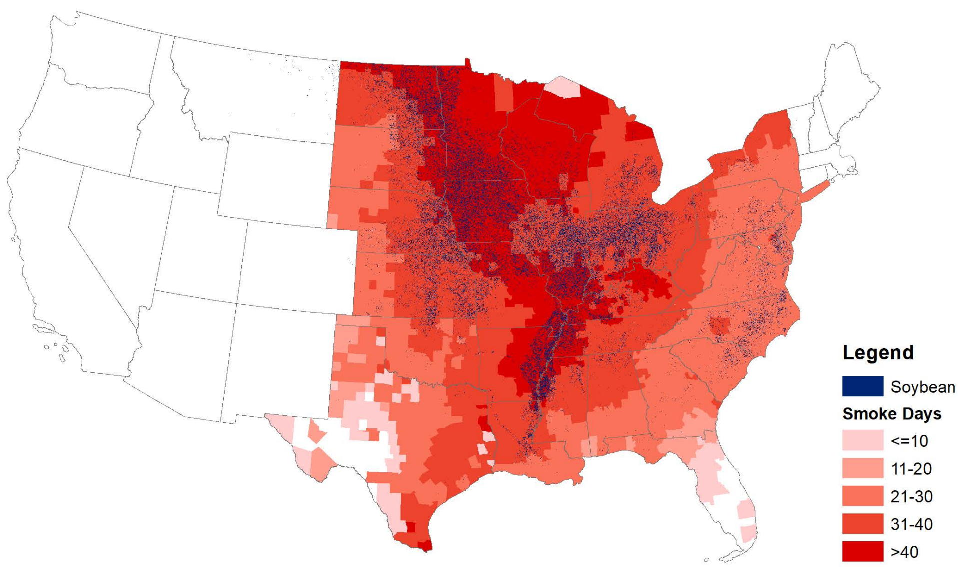
(b) Smoke Days for Soybeans

Figure 1.1: Smoke Days in 2008

Note: This figure shows location of crop land cover and experienced smoke days in each county in different ranges. (a) shows corn land cover and smoke days in 2008 and (b) shows soybean land cover and smoke days in 2008.



(a) Smoke Days for Corn



(b) Smoke Days for Soybeans

Figure 1.2: Smoke Days in 2019

Note: This figure shows location of crop land cover and experienced smoke days in each county in different ranges. (a) shows corn land cover and smoke days in 2019 and (b) shows soybean land cover and smoke days in 2019.

1.3.3 Weather Data

Weather variables, daily mean temperature and precipitation were obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) and PRISM provides daily temperature and precipitation information spanning the study period in 4km grids, across the contiguous United States (PRISM Climate Group). Daily mean temperature and daily precipitation from PRISM is averaged over corn or soybean land cover within each county, depending on the nature of the dependent variable. For corn and soybeans, average daily temperature is then converted to degree days following previous literature (Deschênes and Greenstone, 2007), which uses a base of 8 ℃ and a ceiling of 32 ℃. Degree days are defined to reflect those plants that are only able to absorb heat above a threshold, and cannot absorb more heat above a ceiling when temperature is too high (Deschênes and Greenstone, 2007). A daily temperature below 8 ℃ contributes to 0-degree days. For temperature between base and ceiling, it contributes the number above 8 ℃. If daily temperate exceeds the ceiling temperature, it contributes ceiling less the base degrees. Then, degree days are summed over the growing season to provide an estimate of yearly growing degree days. Growing season precipitation is constructed by summing daily precipitation over the growing season.

1.3.4 Summary Statistics

The yield of corn and soybean is linked to exposure to smoke, fire and weather variables by county by year. Table 1.1 shows the summary statistics for the variables to be used in the empirical modeling exercise. There are 22,045 observations for corn and 19,236 observations for soybeans. Most of corn and soybean production occurs in the counties in the eastern part of the United States. Corn field and soybean field both on average experience about 28 smoke days each year. They also experience similar percentage of fire exposure. Corn fields are exposed to slightly higher growing degree days, as well as higher growing precipitation.

Table 1.1: Summary Statistics

	Corn		Soybean	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Yield (bu/acre)	140.18	39.30	42.11	11.01
Smoke Days	28.30	18.65	28.35	18.54
$\mathbb{1}$(Fire)	0.48	0.50	0.44	0.50
Growing Degree Days	2399.92	618.52	2319.36	553.53
Growing Precipitation (mm)	711.36	218.16	682.04	198.30
N	22,045		19,236	

Note: This table shows summary statistics for corn and soybean sample separately. Yields are in bushels per harvested acres. Smoke days, $\mathbb{1}$(fire), growing degree days, and growing precipitations are calculated following definition in the previous sections.

1.4 Empirical Specification

The regression model forming the core of my empirical specification is given by:

$$\log(Y_{ct}) = \alpha + \beta \text{Smoke Days}_{ct} + \delta \mathbb{1}(\text{Fire})_{ct} + \gamma X_{ct} + \mu_c + f(t) + \varepsilon_{ct} \tag{1.1}$$

where Y_{ct} is outcome variable of yield in county c in year t. Smoke Days_{ct} is the main variable of interest, representing number of days being exposed to smoke, and that variable is defined in the previous section. The direct impacts from a local fire burning are controlled for using $\mathbb{1}(\text{Fire})_{ct}$, an indicator of whether there has been fire in the corn or soybean land in the county, as defined in the previous section. Weather variables are likely to be correlated with smoke exposure and agriculture production. These variables are contained in the X_{ct} set of controls. In my main specification, they are operationalized using a quadratic in growing degree days and a quadratic in growing precipitation. County fixed effects, μ_c, are used to account for characteristics that are time invariant at the county level over my sample period. f(t) represents controls for national or regional level temporal effects and these are operationalized in different variants of Equation 1.1 using polynomial time trends or year fixed effects. The year fixed effects can remove national shocks each year. 4th order polynomial time trend aims to control for technological improvement over the years while the regional 4th order polynomial time trend allows to capture different technological improvement in each region. Where regional level controls are used, they are defined by the nine U.S. climate regions developed by National Centers for Environmental Information for the contiguous United states, where each is climatically consistent (Karl and Koss, 1984). Thus Equation 1.1 uses year to year county-level variation to identify impacts of smoke from fires. Since yield is measured by production per harvested acres, Equation 1.1 is estimated using weights defined by harvested acres. Robust standard errors are clustered at the state level.

Corn and soybean fields are in eastern U.S. as can be seen in Figure 1.1 and Figure 1.2. Brey et al. (2018) link observed smoke plumes to sources of active fires and their results indicate that much of the smoke in eastern U.S. comes from other regions. For the northern parts of the eastern U.S., smoke tends to be produced in fires occurring western U.S. or other countries such as Canada. The southern part of the eastern U.S. has more smoke produced internally, but large share of the smoke it receives comes from western U.S. and outside of U.S. With much of the smoke coming from outside of areas where corn and soybeans are being produced rather than by local fire activity, it is less likely for smoke to be correlated with unobserved local environment conditions influencing yields. This should help with identification.

Estimates based on variants of Equation 1.1 implicitly incorporate post planting responses by farmers after wildfire shocks occur. I now turn to the question of whether, and, if so, how do farmers respond to smoke. It is possible to estimate a regression similar as Equation 1.1, but now using farm production expenses and number of hired farm workers as the dependent variables. This analysis is substantially more limited for two main reasons. First, production expenses data are only available through the U.S. Agriculture Census administered every five years, and hence for only three years, in 2007, 2012 and 2017. Second, production expenses and hired workers data are for all farm production activities, including other crops and livestock, and hence do not directly correspond to my corn or soybean production models. Congruent with this new definitional basis, Smoke Days_{ct} and $\mathbb{1}(\text{Fire})_{ct}$ are now calculated based on all agricultural land cover in a county instead of that specific to corn or soybeans, and for the entire year, rather than for corn or soybean's growing season. Weather variables are now specified in terms of quadratic functions of yearly average temperature and yearly total precipitation.

1.5 Results

1.5.1 Main Results

Table 1.2 displays the results from estimating variants of Equation 1.1. The first three columns show results for corn, while the last three columns are results for soybeans. Each column uses a different approach to account for temporal effects as indicated in the table. Coefficients and standard errors for smoke days and fire indicator are scaled by 100, so they represent the percentage change in yield. For both corn and soybeans, the parameter estimates on the main variables of interest are all statistically significant and reasonably insensitive to the particular approach used to control for temporal effects. Column (3) and column (6) using regional 4th order polynomial in years is my preferred specification, since it captures both technological improvement and allows for regional differences.

These results show that exposure to an additional smoke day decreases the corn yield by 0.313% and the soybean yield by 0.232%. Corresponding elasticities at means are estimated to be -0.117 for corn and -0.082 for soybeans. This effect is somewhat smaller than the impacts found for temperature changes found. Schlenker and Roberts (2009) for instance show that yields for corn and soybeans gradually increases with temperature to about 30 ℃ and then sharply decreases, with exposed to one day of temperature above this decreasing corn and soybean yields in the range of about 1% to 6%.

Table 1.2: Main Results for Yields of Corn and Soybeans

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Days	-0.203**	-0.340***	-0.313***	-0.328***	-0.253***	-0.232***
	(0.082)	(0.083)	(0.073)	(0.063)	(0.046)	(0.044)
$\mathbb{1}$(Fire)	-0.464	-0.156	-0.092	-0.601	-1.030	-0.712
	(0.346)	(0.438)	(0.447)	(0.564)	(0.669)	(0.600)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
4th Order	No	Yes	No	No	Yes	No
Regional 4th Order	No	No	Yes	No	No	Yes
N	22,045	22,045	22,045	19,236	19,236	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Elasticity at means	-0.076	-0.127	-0.117	-0.115	-0.089	-0.082

Note: This table shows main regression results following Equation 1.1 using temporal effects as indicated. Column (1) to (3) show results for corn and column (4) to (6) show results for soybeans. All the coefficients and standard errors are scaled by 100. The values for elasticity at means are reported in the bottom in original scale. Standard errors are clustered at the state level.

1.5.2 Differential Impacts

My treatment variable, the number of smoke days, can be decomposed in various ways that provide additional insight into the nature of its differential impacts by looking at when the smoke occurs, the second how much smoke occurs, and the third where the smoke appears. I can look at the when issue since I have temporal information on when the smoke days occur. As explained in the previous section, I divided the growing seasons into three time periods of planting, cultivation and harvest. Statistical identification comes from seeing different patterns of smoke days. Table 1.3 shows the impact of smoke for both corn and soybeans during the three parts of the growing seasons. Results for cultivation show the most significance for both crops. For corn, the magnitude of coefficient is largest during the planting season, but significance is less than cultivation. Results for harvest season is insignificant. It is important to note that the effect is better defined during the cultivation period. This time period is larger and hence the variation in the number of smoke days being used for identification is larger. For soybeans, the coefficients on smoke during the planting and cultivation period are almost identical in magnitude, but like corn the smoke effect is better defined during the cultivation phase. In contrast to corn, there is a positive, but very noisy and insignificant effect during the harvest season for soybeans.

Secondly, main results provide estimates for impacts of number of days exposed by smoke, but does not differentiate by smoke intensity. There are two reasons for this. It is not clear how the bundle of pollutants that impact plants changes with smoke intensity and NOAA's HMS data does not provide detailed smoke intensity measurement for my entire sample period. I can, however, look at smoke intensity measure from 2011, when HMS starts to provide qualitative information on smoke density. Smoke plume coverage is classified into thin, medium, and thick smoke bins. The number of thick days (e.g., close proximity to a large wildfire) is too small for analysis, thus are grouped with medium smoke days. Table 1.4 shows results of switching the original undifferentiated number of smoke days. These results suggest a dose-response relationship for both corn and soybeans with higher smoke intensity having more deleterious impacts. It is important to note though that these effects are much noisier statistically.

Table 1.3: Results from Different Stages in Growing Season

(Scaled by 100)	(1) Corn log(Yield)	(2) Soybean log(Yield)
Planting Smoke Days	-0.768**	-0.256*
	(0.338)	(0.134)
Cultivation Smoke Days	-0.290***	-0.257***
	(0.059)	(0.036)
Harvest Smoke Days	-0.490	0.238
	(0.293)	(0.236)
County FE	Yes	Yes
Regional 4th Order	Yes	Yes
N	22,045	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results by decomposing smoke days in the growing season into three time periods of planting, cultivation and harvest. Coefficients and standard errors are all scaled by 100. Standard errors are clustered at the state level.

Table 1.4: Results from Different Smoke Density

(Scaled by 100)	(1) Corn log(Yield)	(2) Soybean log(Yield)
Thin Smoke Days	-0.280**	-0.206*
	(0.135)	(0.112)
Mid & Thick Smoke Days	-0.332	-0.364**
	(0.209)	(0.161)
County FE	Yes	Yes
Regional 4th Order	Yes	Yes
N	13,449	11,961

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results by decomposing smoke days into different density category of thin smoke days and mid & thick smoke days. The regression is conducted over 2011 to 2019 due to data availability. Coefficients and standard errors are all scaled by 100. Standard errors are clustered at the state level.

Table 1.5: Results from Smoke Days Interacting with County Income Level

(Scaled by 100)	(1) Corn log(Yield)	(2) Soybean log(Yield)
Smoke Days	-0.264***	-0.208***
	(0.065)	(0.042)
Smoke Days* 𝟙(Low Income)	-0.240***	-0.081
	(0.067)	(0.059)
County FE	Yes	Yes
Regional 4th Order	Yes	Yes
N	22,045	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results by adding interaction term of smoke days with indicator of whether a county is below the median income. Coefficients and standard errors are all scaled by 100. Standard errors are clustered at the state level.

Lastly, I interact the number of smoke days with two types of county level indicators. One is indicator of whether a county is below the median in terms of income to examine whether lower income counties are more negatively affected. Results are shown in Table 1.5. The coefficients for interaction of smoke days and lower income counties are negative for both commodities but only the one for corn is statistically significant. Another indicator is whether the county has more smaller farms. This indicator is based on the median farm size in each county and the indicator is 1 if the median farm size in a county is below the median for all counties. Thus, the indicator can be interpreted as whether the farm has more smaller farms. Table 1.6 shows negative coefficients for the interaction term but again, only the one for corn is statistically significant.

Table 1.6: Results from Smoke Days Interacting with County Farm Size Distribution

(Scaled by 100)	(1) Corn log(Yield)	(2) Soybean log(Yield)
Smoke Days	-0.283***	-0.221***
	(0.067)	(0.048)
Smoke Days* 𝟙(More Smaller Farms)	-0.172** (0.073)	-0.052 (0.038)
County FE	Yes	Yes
Regional 4th Order	Yes	Yes
N	22,045	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results by adding interaction term of smoke days with indicator of whether a county has more smaller farms. Coefficients and standard errors are all scaled by 100. Standard errors are clustered at the state level.

1.5.3 Robustness Check

I conduct several additional robustness checks to improve confidence in the main results. First, most of the smoke in northern parts of eastern U.S. states are from outside that area while a large proportion of smoke in the southern part of the eastern U.S. states is generated internally and hence may hide some unknown important source of endogeneity that would influence the interpretation of my results. I rerun my main specification excluding southern regions of eastern U.S. to only consider impacts of smoke if smoke is coming from outside regions. These results are shown in Table 1.10 in the Appendix. The coefficients are similar to the main results. Since the HMS data is not able to distinguish between wildfire, prescribed burning or agricultural burning, another concern to the specification is that farmers conduct agricultural burning before planting according to the expectation of yields, creating potential endogeneity. To address this issue, I drop observations where there is fire in the county in a particular year. Table 1.11 in

Appendix shows these results. Most results are similar to my main specifications, although I note that the coefficients under year-FE become less significant for corn. This limits the concern of potential endogeneity of agricultural burning. Many papers have looked into the impact of temperature and climate change on agriculture production. There may be concerns of whether the impact of smoke identified in this paper has been captured as part of the temperature impact. I regress log(yield) on residual of smoke days removing temperature and precipitation to examine impact of smoke apart from temperature and precipitation. To be specific, I regress smoke days on quadratic growing degree days and quadratic growing precipitation to obtain smoke residual and then run regression of Equation 1.1, but replacing smoke days and weather variables with smoke residual. Results can be seen in Table 1.12 in Appendix. The coefficients for smoke residual are similar to main regression results, indicating that the impact shown in this paper is likely from smoke itself instead of temperature or precipitation.

Next, I turn to the potential role of irrigation. Previous research on climate change impacts has shown that areas with and without irrigation systems may be differently impacted with respect to yield outcomes (Schlenker and Roberts, 2009). The irrigation status of a county is defined following Schlenker and Roberts (2009), where counties east of the 100 degree meridian are considered non-irrigated counties and those west of the 100 degree meridian are considered irrigated counties. Table 1.13 shows results allowing for each of the weather variables to differ between irrigated vs. non-irrigated counties, i.e., providing interaction of weather variables with irrigation status. These are largely consistent with the main results. In part, this occurs because most corn and soybeans are planted in eastern counties, and hence classified as non-irrigated counties.

My main results in Table 1.2 displays several specifications involving different specifications for modeling the temporal component (e.g., fixed effects vs time trends). I explore this specification issue further here by considering region-year fixed effects, quadratic time trends, as well as regional quadratic time trends. These results are shown in Table 1.14 in the Appendix. Again, parameter estimates for smoke days are similar to my main regression results.

I also look at using yield rather than log(yield) as the dependent variable. Table 1.15 in Appendix shows coefficients from models using yield as the dependent variable. When the estimated coefficients are divided by weighted mean yield to obtain the percent change in yields from exposure to one more day of smoke, the results are similar to my main specifications.

For the purposes of this paper, I defined a smoke day as occurring when the entire (100%) area where a particular crop is planted in a county is covered by smoke plumes, similar to Borgschulte et al. (2020). I also explore alternatives to use smoke days measures based on different percentage of area devoted to corn or soybeans that was covered by smoke. I look at smoke days defined by having at least a percentage cover of 75% or 50% of the crop's area in a county covered by smoke plumes. These results are displayed in Table 1.16 in Appendix, and show very little change to the coefficients for the smoke days.

1.5.4 Farmer's Responses

In addition to corn and soybean yields, I can also examine how farmers experiencing smoke adjust their production inputs. The key caveat noted earlier is that farm production expenses and hired labor data are only available for total farm operations in each county, not for specific crops. Further, county level expenses and hired labor data are only available for the three USDA Census of Agriculture years: 2007, 2012 and 2017. I run a regression similar to Equation 1.1 for these three years. Smoke days and fire indicators are now counted over the entire year rather than a specific growing season and over all agricultural land instead of those planted with corn (or soybeans). I also replace crop specific growing degree days and growing precipitation with average yearly temperature and total annual precipitation as the weather controls. Three years of data do not define 4th order polynomial in time, so I only look at models with year fixed effects for f(t). While regressions are still clustered at the state level, they are now weighted using farm land areas.

Table 1.7: Results of Farm Production Expenses for All Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Production	Agricultural Chemical	Fertilizer	Labor	Fuel	Others
(Scaled by 100)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)
Smoke Days	0.109**	0.266**	0.351**	0.028	0.035	0.108**
	(0.044)	(0.128)	(0.143)	(0.064)	(0.071)	(0.050)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	8,808	8,808	8,808	8,808	8,808	8,808

Standard errors in parentheses

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

Note: This table shows regression results for farm production expenses. Column (1) shows result for total farm production expenses. Column (2) to (5) show results for selected detailed categories of expenses. Column (6) shows result for rest of the other expenses. Regression is conducted over 2007, 2012 and 2017. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

I first look at farm production expenses. The expenses are divided by farm land areas to obtain per acre expenses and then log of per acre expenses are used as outcome variable. Regression results are displayed in Table 1.7. The first column is for total production expense, which includes all types of expenses for production. One additional day of smoke exposure increases total farm production expense per acre by 0.109%. Columns (2) to (5) show similar results for selected categories of production expenses. Agricultural chemical, which includes insecticides, herbicides, fungicides and other pesticides (USDA, 2017), and fertilizer expenses are substantially more responsive and increase by 0.266% and 0.351%, respectively. On the other hand, the coefficients on labor and fuel expenses are positive, but quite small and insignificant. An "other" expense category provides similar results to rest of other production expenses. Taken as a whole, these results suggest that farmers do engage in some adjustment of production inputs as a result of exposure to smoke and that these changes are concentrated in the agricultural chemicals and fertilizer categories. Table 1.17 in Appendix splits out corn and soybean counties (there is some overlap) and provides a noisier but perhaps more nuanced view. It suggests that the agricultural chemical response is concentrated in corn counties, that allowing the separate effects on fertilizer are similar but now insignificant. There is now a hint of employing more labor for corn and a highly significant "other" expense category effect for soybeans. I take the combination of Table 1.7 and Table 1.17 as suggestive that there are farm level responses that try to compensate for receiving more smoke days. Sorting those effects out is a useful extension of this paper and will require detailed farm level data used in crop level studies of agricultural production.

Lastly, I look at the impacts of the number of smoke days on number of workers hired. There could be three underlying mechanisms behind the need for more labor. One is through farm workers becoming less productive due to smoke, but smoke exposure in this paper is usually not thick smoke day as depicted over California fields during wildfire season. The second is that more labor is needed to apply agricultural chemicals and fertilizer. The third comes from noting that one way to deal with falling yields is to hire more labor to increase the fraction of the crop successfully harvested. The coefficient on smoke days in Table 1.8 is positive, but small and insignificant. This suggests that hired farm labor is not one of the main channels for smoke impacts on the input side. Table 1.18 in the Appendix suggests to the extent that there is a hired labor effect it is concentrated in soybeans rather than corn.

Table 1.8: Results of Hired Farm Labor for All Counties

(Scaled by 100)	(1) log(Hired Labor)
Smoke Days	0.055 (0.048)
County FE	Yes
Year FE	Yes
N	8,741

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results for number of hired farm workers. Regression is conducted over 2007, 2012 and 2017. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

1.6 Economic and Policy Implications

1.6.1 Economic Loss

Using the results reported in the previous sections, the economic costs of the damages associated with a given change in smoke days can be estimated by a simple back-of-the-envelope calculation using the following equation:

$$\Delta\% \text{ in Smoke Days} \times \text{2019 Average Smoke Days} \times \Delta\% \text{ in Yield per Smoke Day} \times$$
$$\text{2019 Average U.S. Yield} \times \text{2019 Total U.S. Harvested Acres} \times \text{2019 Average U.S. Price} \quad (1.2)$$

Assuming a 10% increase in number of smoke days relative to 2019, this leads to an annual loss of \$576 million for yields of corn and \$262 million for yields of soybeans. Adding the two leads to \$838 million. Thus, the total loss of corn and soybeans is almost \$1 billion.

This calculation assumes that average U.S. prices for corn and soybeans do not change. This may not be a conservative assumption, as it would be natural to think that agricultural commodity prices would increase, if world production decreases. For both corn and soybeans, the U.S. produces about third of world production, so the change above is on the order of 3% of world production. There would no doubt be adjustments, however, there is already a reasonable amount of price variation, with corn and soybean prices being volatile.

A 10% increase is used in the calculation as a base to consider economic loss. Detailed damage estimates for particular climate change scenarios would require a model that predicts changes in wildfire location, timing and duration. An air dispersion model calibrated to current satellite-based plume maps projecting number of location-specific smoke days for those scenarios would also be required. This has not been done and is beyond the scope of this paper. However, many of the individual components exist for such projections and these could be included as comprehensive assessment of overall economic damages under climate change in future studies.

While such projection is not available, a simple analysis here may shed some useful light. I conduct a simple projection by fitting a straight line to average number of growing season smoke days each year from 2006 to 2019 for corn and soybeans. Then I use this model to predict number of smoke days for years forward from 2019, end of my sample period, using the fitted time trend. This simple projection suggests that the coming decade will lead to a 38% increase in smoke days between 2019 and 2029 for both corn and soybeans. A further extrapolation to

2050 suggests over 80% increase in smoke days compared to 2019 for both crops. These predictions are substantially greater than 10% increase used in the back-of-the-envelope calculation above. The larger increases are consistent with predicted increases in wildfires. Gao et al. (2021) projected increases in fire probability by mid-century (2040-2069) compared to baseline period (1971–2000) in most regions in the contiguous U.S., where many regions experience substantial increases and largest increase is up to 75%. This also suggests smoke levels under climate change are likely to be much higher, although further modeling is required to link increases in wildfires to specific smoke levels. Projections by Gao et al. (2021) also suggest that greatest increases are in regions where corn and soybean production are concentrated. This is likely to lead to exposure to smoke with higher density by corn and soybean and lead to further damage. All these projections suggest that the economic loss due to increase in smoke levels in the future is likely to be larger than the back-of-the-envelope calculation using 10% increase in smoke.

Furthermore, according to USDA statistics, corn and soybeans together account for 43% of total crop cash receipts (USDA, 2019). This paper focus on impacts of smoke on production of corn and soybeans and similar impacts are likely to impose on other crops. Corn and soybeans are major cash crops and have been well studied. There have been earlier evidences of varietal switching as well as crop switching behavior in response to pollution exposure (Griliches, 1957; Kopp et al., 1985). Research conducted for corn and soybeans aimed at implicit adaptation to smoke is likely to be more advanced than other crops and it is less likely to think that other crops are less sensitive to smoke.

Further economic losses may also occur in countries other than U.S. Calculating from data in Food and Agriculture Organization of the United Nations (2018), in terms of gross production value for total crops, U.S. accounts for 8% of the world value. Other major agricultural producers such as Australia, Brazil, Canada, China, European Union, India, Russia, and Turkey, which are among the top 30 for gross production value of total crops in the world (Food and Agriculture Organization of the United Nations, 2018) may also be affected by smoke. They are projected to have increases in wildfires and some of them may experience substantial impacts (Sun et al., 2019).

Current back-of-the-envelope calculation shows that 10% increase in smoke relative to 2019 leads to annual economic loss of corn and soybeans to be almost $1 billion. If assuming similar damage of smoke on other crops, as well as damage in other countries, the potential increase in wildfire-related smoke over the next 30 years through 2050 are likely to lead to impacts on agriculture in the order of hundred-billion-dollar magnitude.

1.6.2 Wildfire Management Options

My empirical results also have policy implication for wildfire management. Government expenditures on dealing with the threat of wildfires has grown over time, reaching $6.11 billion in FY2020 (Hoover, 2020). Fuel reduction programs play an important role in mitigating the threat. The purpose of these programs is to remove grasses, shrubs, and trees to preserve ecosystems and constrain the damages of wildfires (U.S. Department of the Interior, Office of Wildland Fire, 2021). There are two main and often competing variants: prescribed burning and mechanical removal. Prescribed burning deliberately sets fire to vegetation likely to amplify wildfires or direct such fires toward high value targets like homes. The vegetation that serves as the fuel (e.g., scrubs and dead timber) is almost identical to that burned during a wildfire. The major difference is that prescribed burns are set during favorable weather conditions for containment. Prescribed burning is usually less costly than mechanical removal (Wade and

Lunsford, 1989), which uses machines and people to remove the same vegetation. There is public controversy over prescribed burning often relates to concerns of risks in escaping of fires (Ryan et al., 2013) as well as the obvious air pollution in the form of visible smoke.

Table 1.9 provides a highly simplified cost comparison of prescribed burning vs. mechanical treatment. The per acre costs for fuel management follow estimates from Calkin and Gebert (2006) and are converted to $2019. With the per acre cost of mechanical removal more than three times larger than prescribed burning, the attraction of government agencies to prescribed burning is obvious. However, prescribed burning also generates smoke and air pollution. This should be taken into account in making a choice between the two approaches. Table 1.9 adds per acre costs of loss of corn and soybeans from smoke based on back-of-the-envelope calculation. Using the equation in the previous section, a 31% of smoke is used to estimate total loss from prescribed burning as according to U.S. EPA 2017 NEI Data. Out of agricultural burning, prescribed fire and wildfires, prescribed burning accounts for about 31% emission out of the three types of fires. The loss is then divided by total prescribed burning acres in 2019 of 6.06 million acres (National Interagency Fire Center, 2019) to obtain a per acre cost.

Table 1.9: Per Acre Cost Comparison for Fuel Management

(Per Acre)	Prescribed Burning	Mechanical Treatment
Cost of Fuel Management	$83	$296
Cost of Yields of Corn	$296	-
Cost of Yields of Soybean	$135	-
Total Cost	$515	$296

Note: This table shows cost comparison between prescribed burning and mechanical treatment as fuel management methods. All costs are in per acre, based on acres that are applied with fuel management treatments.

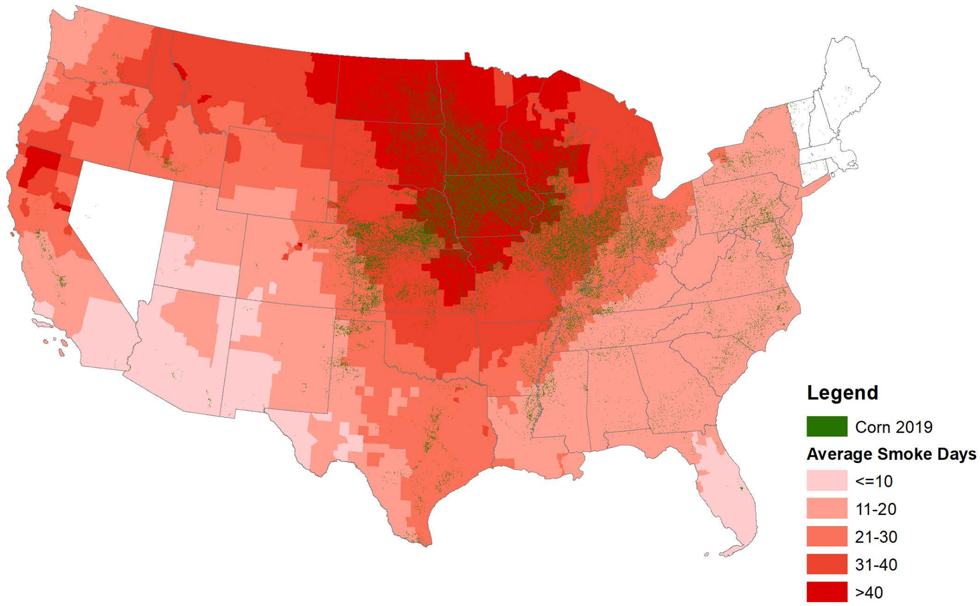
Under current assumptions, calculations show that while cost of fuel management itself for prescribed burning is less expensive than mechanical treatment, the total cost considering negative impacts of yield is higher for prescribed burning. Policy makers need to take into account the costs associated with smoke from fires when making decisions. Other than yield loss, smoke also generates health costs, including to elderly, child, farm workers, etc. Miller et al. (2017) used similar smoke plume data and estimated annual death of elderly (over 65) from smoke to be 489. This can be converted to a cost by multiplying the value of statistical life of \$7.4 million (\$2006), according to mortality risk valuation from U.S. EPA. This can be further converted to \$2019, then similarly, multiplying by 31% and diving by 6.06 million acres to obtain a per acre cost of \$236 (in \$2019). This health cost is estimated for the elderly and the total health costs can be even larger. Adding these costs will make the prescribed burning option less favorable. The current cost comparison is derived from a highly simplified calculation and based on various assumptions, including implicitly assuming per acre application of the two fuel management treatments have similar impacts in reducing wildfires. More detailed cost benefit analysis is needed for future research.

1.7 Conclusion

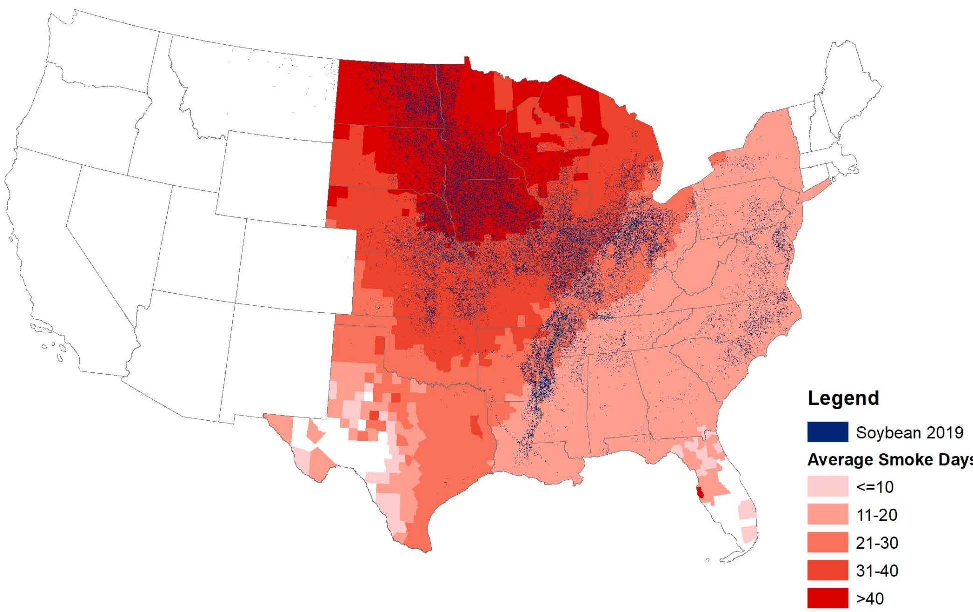
In this paper, I show that exposure to one more day of smoke from fires reduces yields of corn and soybeans by 0.31% and 0.23%, respectively. I estimate that the annual damage associated with a 10% increase in smoke relative to 2019 for these two crops from reduced yields to be \$838 million. Various projections suggest that wildfires and smoke levels are likely to increase in the future. If assuming there will be similar impacts of smoke on other crops and the impact in U.S. may also occur in other countries, the potential increase in smoke over the next 30 years through 2050 are likely to lead to agricultural losses on order of several hundred billion dollars.

My results suggest that the smoke effects are concentrated in cultivation stage. This makes intuitive sense. Application of agricultural chemicals and fertilizer are classic ways for farmers to respond to adverse shocks in an effort to try to drive expected yields back up. On the other hand, demand for labor doesn't seem to decrease as a result of decreased yields. More detailed smoke data in later years also suggests that heavier smoke levels reduce yields more. There is some empirical support for the proposition that lower income U.S. counties and counties with a greater concentration of small farms suffer proportionately larger adverse yield reductions. Exploring this issue further requires farm level data. The current study also leaves an open question for the longer term, to understand how much room is left for farmers to respond to increasing levels of smoke by switching to alternative cultivar without substantial reductions to the profits, in a sector that already has low margins.

1.8 Appendix



(a) Smoke Days for Corn



(b) Smoke Days for Soybeans

Figure 1.3: Average Smoke Days

Note: This figure shows location of crop land cover in 2019 and average experienced smoke days over 14 years from 2006 to 2019 in each county. (a) shows corn land cover in 2019 and average smoke days and (b) shows soybean land cover in 2019 and average smoke days.

Table 1.10: Results for Yields Dropping Southern States

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Days	-0.161	-0.359***	-0.332***	-0.333***	-0.269***	-0.248***
	(0.095)	(0.093)	(0.082)	(0.075)	(0.048)	(0.046)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
4th Order	No	Yes	No	No	Yes	No
Regional 4th Order	No	No	Yes	No	No	Yes
N	15,192	15,192	15,192	13,351	13,351	13,351

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results of yields for corn and soybean dropping southern states in the sample. Specific states dropped are Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee and Texas. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.11: Results for Yields Dropping Observations Where There is a Fire in the County

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Days	-0.203*	-0.277***	-0.253***	-0.372***	-0.250***	-0.233***
	(0.106)	(0.094)	(0.087)	(0.090)	(0.044)	(0.047)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
4th Order	No	Yes	No	No	Yes	No
Regional 4th Order	No	No	Yes	No	No	Yes
N	8,106	8,106	8,106	6,216	6,216	6,216

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results of yields for corn and soybean dropping observations that there is fire in a county in the year. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.12: Results for Yields Regressing on Smoke Residuals

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Residuals	-0.247***	-0.465***	-0.461***	-0.194***	-0.283***	-0.284***
	(0.086)	(0.084)	(0.081)	(0.060)	(0.065)	(0.062)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
4th Order	No	Yes	No	No	Yes	No
Regional 4th Order	No	No	Yes	No	No	Yes
N	22,045	22,045	22,045	19,236	19,236	19,236

Standard errors in parentheses

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

Note: Smoke residuals are the residuals of regressing smoke days on quadratic growing degree days and quadratic growing precipitation. This table shows regression results from Equation 1.1 but replacing smoke days, quadratic growing degree days and quadratic growing precipitation with smoke residuals. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.13: Results for Yields with Controls of Growing Degree Days and Precipitation Interacting with Irrigation Status

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Days	-0.210**	-0.341***	-0.311***	-0.325***	-0.253***	-0.232***
	(0.083)	(0.085)	(0.074)	(0.065)	(0.046)	(0.044)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
4th Order	No	Yes	No	No	Yes	No
Regional 4th Order	No	No	Yes	No	No	Yes
N	22,045	22,045	22,045	19,236	19,236	19,236

Standard errors in parentheses

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

Note: This table shows regression results of yields for corn and soybean, where controls for each term of quadratic growing degree days and quadratic growing precipitation allowed to be interacted with irrigation status. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.14: Results for Yields with Alternative Fixed Effects and Time Trends

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Days	-0.264***	-0.384***	-0.373***	-0.274***	-0.302***	-0.285***
	(0.092)	(0.100)	(0.095)	(0.059)	(0.047)	(0.049)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	No	No	Yes	No	No
2nd Order	No	Yes	No	No	Yes	No
Regional 2nd Order	No	No	Yes	No	No	Yes
N	22,045	22,045	22,045	19,236	19,236	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results of yields for corn and soybean, where each column follows alternative fixed effects and time trends as specified. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.15: Results for Level of Yields of Corn and Soybean

	Corn			Soybean		
	(1) Yield	(2) Yield	(3) Yield	(4) Yield	(5) Yield	(6) Yield
Smoke Days	-0.209**	-0.394***	-0.375***	-0.114***	-0.099***	-0.098***
	(0.100)	(0.097)	(0.083)	(0.020)	(0.018)	(0.018)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
4th Order	No	Yes	No	No	Yes	No
Regional 4th Order	No	No	Yes	No	No	Yes
N	22,045	22,045	22,045	19,236	19,236	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Coefficient/ mean yield	-0.130%	-0.245%	-0.233%	-0.249%	-0.218%	-0.214%
Elasticity at means	-0.048	-0.091	-0.087	-0.088	-0.077	-0.075

Note: This table shows regression results of yields for corn and soybean, using level of yields instead of log(yield) as outcome. Standard errors are clustered at the state level.

Table 1.16: Results for Yields at Different Smoke Coverage over Crop Fields

	Corn			Soybean		
	(1)	(2)	(3)	(4)	(5)	(6)
	≥50% Cover	≥75% Cover	Full Cover	≥50% Cover	≥75% Cover	Full Cover
(Scaled by 100)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)	log(Yield)
Smoke Days	-0.314***	-0.314***	-0.313***	-0.227***	-0.228***	-0.232***
	(0.073)	(0.072)	(0.073)	(0.044)	(0.043)	(0.044)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Regional 4th Order	Yes	Yes	Yes	Yes	Yes	Yes
N	22,045	22,045	22,045	19,236	19,236	19,236

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results of yields for corn and soybean, using alternative smoke coverage as specified. Column (3) and (6) are same as the main specification. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.17: Results of Farm Production Expenses for Corn or Soybean Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Production	Agricultural Chemical	Fertilizer	Labor	Fuel	Others
(Scaled by 100)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)	log(Expense /Acre)
Panel A: Regression for Corn Counties						
Smoke Days	0.111*	0.198*	0.051	0.142*	-0.037	0.110*
	(0.064)	(0.101)	(0.127)	(0.083)	(0.104)	(0.061)
N	4,980	4,980	4,980	4,980	4,980	4,980
Panel B: Regression for Soybean Counties						
Smoke Days	0.139**	0.010	0.048	0.159	-0.017	0.164***
	(0.057)	(0.127)	(0.122)	(0.094)	(0.050)	(0.059)
N	4,270	4,270	4,270	4,270	4,270	4,270
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

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

Note: This table shows regression results for farm production expenses in restricted samples. Panel A shows regression results only in corn counties and Panel B shows regression results only in soybean counties. Regressions are conducted over 2007, 2012 and 2017. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

Table 1.18: Results of Hired Farm Labor for Corn or Soybean Counties

(Scaled by 100)	(1) log(Hired Labor)
Panel A: Regression for Corn Counties	
Smoke Days	0.055
	(0.089)
N	4,971
Panel B: Regression for Soybean Counties	
Smoke Days	0.105
	(0.067)
N	4,266
County FE	Yes
Year FE	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows regression results for number of hired farm workers in restricted samples. Panel A shows regression results only in corn counties and Panel B shows regression results only in soybean counties. Regressions are conducted over 2007, 2012 and 2017. All the coefficients and standard errors are scaled by 100. Standard errors are clustered at the state level.

1.9 References

Abatzogloua, J. T. and Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42):11770–11775.

Ainsworth, E. A., Rogers, A., and Leakey, A. D. B. (2008). Targets for crop biotechnology in a future high-CO2 and high-O3 world. *Plant Physiology*, 147(1):13–19.

Ainsworth, E. A., Yendrek, C. R., Sitch, S., Collins, W. J., and Emberson, L. D. (2012). The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annual Review of Plant Biology*, 63:637–661.

Boone, C. D. A., Schlenker, W., and Siikamäki, J. (2019). *Ground-level ozone and corn yields in the United States*. Working Paper. Center for Environmental Economics and Policy, Columbia University.

Borgschulte, M., Molitor, D., and Zou, E. Y. (2020). *Air pollution and the labor market: Evidence from wildfire smoke*. Working Paper. Department of Economics, University of Illinois at Urbana-Champaign.

Brey, S. J., Ruminski, M., Atwood, S. A., and Fischer, E. V. (2018). Connecting smoke plumes to sources using Hazard Mapping System (HMS) smoke and fire location data over North America. *Atmospheric Chemistry and Physics*, 18(3):1745–1761.

Burke, M., Driscoll, A., Heft-Neal, S., Xue, J., Burney, J., and Wara, M. (2007). The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences*, 118(2):e2011048118.

Calkin, D. and Gebert, K. (2006). Modeling fuel treatment costs on forest service lands in the western United States. *Western Journal of Applied Forestry*, 21(4):217–221.

Da, Y., Xu, Y., and McCarl, B. (2021). Effects of surface ozone and climate on historical (1980–2015) crop yields in the United States: Implication for mid-21st century projection. *Environmental and Resource Economics*, 81:355–378.

Deschênes, O. and Greenstone, M. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97:354–385.

Deschênes, O., Greenstone, M., and Shapiro, J. S. (2017). Defensive investments and the demand for air quality: Evidence from the NOx budget program. *American Economic Review*, 107(10):2958–2989.

Florec, V., Burton, M., Pannell, D., Kelso, J., and Milne, G. (2020). Where to prescribe burn: The costs and benefits of prescribed burning close to houses. *International Journal of Wildland Fire*, 29:440–458.

Food and Agriculture Organization of the United Nations (2018). *FAOSTAT*. Available from: https://www.fao.org/statistics/en/.

Ford, B., Val Martin, M., Zelasky, S. E., Fischer, E. V., Anenberg, S. C., Heald, C. L., and Pierce, J. R. (2018). Future fire impacts on smoke concentrations, visibility, and health in the contiguous United States. *GeoHealth*, 2(8):229–247.

Gao, P., Terando, A. J., Kupfer, J. A., Varner, J. M., Stambaugh, M. C., Lei, T. L., and Hiers, J. K. (2021). Robust projections of future fire probability for the conterminous United States. *Science of the Total Environment*, 789:147872.

Garcia, P., Dixon, B. L., Mjelde, J. W., and Adams, R. M. (1986). Measuring the benefits of environmental change using a duality approach: The case of ozone and Illinois cash grain farms. *Journal of Environmental Economics and Management*, 13(1):69–80.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522.

Hemes, K. S., Verfaillie, J., and Baldocchi, D. D. (2020). Wildfire-smoke aerosols lead to increased light use efficiency among agricultural and restored wetland land uses in California's Central Valley. *Journal of Geophysical Research: Biogeosciences*, 125:e2019JG005380.

Hoover, K. (2020). *Federal wildfire management: Ten-year funding trends and issues (FY2011-FY2020)*. (R46583). Congressional Research Service.

Jaffe, D. A. and Wigder, N. L. (2012). Ozone production from wildfires: A critical review. *Atmospheric Environment*, 51:1–10.

Jayachandran, S. (2009). Air quality and early-life mortality evidence from Indonesia's wildfires. *Journal of Human Resources*, 44(4):916–954.

Karl, T. R. and Koss, W. J. (1984). *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983*. (Historical Climatology Series 4-3). National Climatic Data Center.

Kopp, R. J., Vaughan, W. J., Hazilla, M., and Carson, R. (1985). Implications of environmental policy for U.S. agriculture: The case of ambient ozone standards. *Journal of Environmental Management*, 20:321–331.

Krupa, S. V. and Manning, W. J. (1988). Atmospheric ozone: Formation and effects on vegetation. *Environmental Pollution*, 50(1-2):101–137.

Mercer, D. E., Prestemon, J. P., Butry, D. T., and Pye, J. M. (2007). Evaluating alternative prescribed burning policies to reduce net economic damages from wildfire. *American Journal of Agricultural Economics*, 89(1):63–77.

Miller, N., Molitor, D., and Zou, E. (2017). *Blowing smoke: Health impacts of wildfire plume dynamics*. Working Paper. Department of Finance, University of Illinois at Urbana-Champaign.

National Interagency Fire Center (2019). *Wildland fire summary and statistics annual report 2019*. https://www.predictiveservices.nifc.gov/intelligence/2019_statssumm/annual_report_2019.pdf.

Neidell, M. (2009). Information, avoidance behavior, and health: The effect of ozone on asthma hospitalizations. *The Journal of Human Resources*, 44(2):450–478.

Park, S. B., Knohl, A., Lucas-Moffat, A. M., Migliavacca, M., Gerbig, C., Vesala, T., Peltola, O., Mammarella, I., Kolle, O., Lavrič, J. V., Prokushkin, A., and Heimann, M. (2018). Strong radiative effect induced by clouds and smoke on forest net ecosystem productivity in central Siberia. *Agricultural and Forest Meteorology*, 250-251:376–387.

PRISM Climate Group. Oregon State University. Available from: http://prism.oregonstate.edu.

Radeloff, V. C., Helmers, D. P., Kramer, H. A., Mockrin, M. H., Alexandre, P. M., Bar-Massada, A., Butsic, V., Hawbaker, T. J., Martinuzzi, S., Syphard, A. D., and Stewart, S. I. (2018). Rapid growth of the US wildland-urban interface raises wildfire risk. *Proceedings of the National Academy of Sciences*, 115(13):3314–3319.

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., and Elliott, C. T. (2016). Critical review of health impacts of wildfire smoke exposure. *Environmental Health Perspectives*, 124(9):1334–1343.

Richardson, L. A., Champ, P. A., and Loomis, J. B. (2012). The hidden cost of wildfires: Economic valuation of health effects of wildfire smoke exposure in Southern California. *Journal of Forest Economics*, 18(1):14–35.

Ryan, K. C., Knapp, E. E., and Varner, J. M. (2013). Prescribed fire in North American forests and woodlands: History, current practice, and challenges. *Frontiers in Ecology and the Environment*, 11(Online Issue 1):e15–e24.

Santana, F. N., Gonzalez, D. J., and Wong-Parodi, G. (2021). Psychological factors and social processes influencing wildfire smoke protective behavior: Insights from a case study in Northern California. *Climate Risk Management*, 34:100351.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37):15594–15598.

Sun, Q., Miao, C., Hanel, M., Borthwick, A. G., Duan, Q., Ji, D., and Li, H. (2019). Global heat stress on health, wildfires, and agricultural crops under different levels of climate warming. *Environment International*, 128:125–136.

U.S. Department of Agriculture. *National Agricultural Statistics Service Surveys*. Available from: https://www.nass.usda.gov/Surveys/.

U.S. Department of Agriculture. *USDA National Agricultural Statistics Service Cropland Data Layer*. Available from: https://nassgeodata.gmu.edu/CropScape/.

U.S. Department of Agriculture (2007). *2007 Census of Agriculture*. Available from: https://www.nass.usda.gov/AgCensus/.

U.S. Department of Agriculture (2010). *Field crops: Usual planting and harvesting dates*. National Agricultural Statistics Service, Agricultural Handbook Number 628.

U.S. Department of Agriculture (2012). *2012 Census of Agriculture*. Available from: https://www.nass.usda.gov/AgCensus/.

U.S. Department of Agriculture (2017). *2017 Census of Agriculture*. Available from: https://www.nass.usda.gov/AgCensus/.

U.S. Department of Agriculture (2019). *Farm income and wealth statistics*. https://www.ers.usda.gov/data-products/farm-income-and-wealth-statistics/.

U.S. Department of the Interior, Office of Wildland Fire (2021). *Fuels management*. https://www.doi.gov/wildlandfire/fuels.

U.S. Environmental Protection Agency. *Air Pollutant Emissions Trends Data*. https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data.

U.S. Environmental Protection Agency. *Mortality risk valuation*. https://www.epa.gov/environmental-economics/mortality-risk-valuation.

U.S. Environmental Protection Agency (2017). *2017 National Emissions Inventory (NEI) Data*. https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-data.

U.S. National Oceanic and Atmospheric Administration. *Hazard Mapping System*. Available from: https://www.ospo.noaa.gov/Products/land/hms.html.

U.S. National Oceanic and Atmospheric Administration (2021). *Product information: Hazard Mapping System*. https://www.ospo.noaa.gov/Products/land/hms.html#about.

Vo, J., Acmio, N. C., and Landon, C. (2021). Wildfire smoke in strawberry fields: Evaluating fire-associated PM2.5 exposure effects on Ventura County farmworkers. *Chest*, 160(4):a1928.

Wade, D. D. and Lunsford, J. D. (1989). *A guide for prescribed fire in southern forests*. Technical Publication R8-TP 11, U.S. Department of Agriculture, Forest Service, Southern Region.

Westenbarger, D. A. and Frisvold, G. B. (1995). Air pollution and farm-level crop yields: An empirical analysis of corn and soybeans. *Agricultural and Resource Economics Review*, 24(2):156–165.

Yamasoe, M., von Randow, C., Manzi, A., Schafer, J., Eck, T., and Holben, B. (2005). Effect of smoke on the transmissivity of photosynthetically active radiation inside the canopy. *Atmospheric Chemistry and Physics Discussions, European Geosciences Union*, 5(4):5909–5934.

Yue, X. and Unger, N. (2018). Fire air pollution reduces global terrestrial productivity. *Nature Communications*, 9(5431):1–9.

Zhang, J. and Mu, Q. (2018). Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks. *Journal of Environmental Economics and Management*, 92:517–536.

Zhou, L., Chen, X., and Tian, X. (2018). The impact of fine particulate matter ($PM_{2.5}$) on China's agricultural production from 2001 to 2010. *Journal of Cleaner Production*, 178:133–141.

Chapter 2

The Interaction Relation of Temperature and Precipitation with Outdoor Recreation

Abstract: Impacts of weather variables on various outcomes have been studied in a large amount of literature. Many studies usually focus on one weather variable as variable of interest and use others as controls. This paper explores whether the interaction of two weather variables plays a role in outdoor recreation trips. Using detailed information on outdoor recreation trips in England over a four-year period and a semi-parametric response surface approach, this paper examines the interaction relation of temperature and precipitation with number of outdoor recreation trips. It was found that although daily visits increase with temperature and decrease with rain, these gradients only have small variations across rain or temperature. Interaction of the two variables plays a small role in outdoor recreation.

2.1 Introduction

The impact of weather variables and climate change on economic performance has been a key area of research. Many studies have been conducted and a large literature has looked into the impacts from different perspectives, including agriculture, health, energy, conflict, and economic growth, etc. These papers help provide insights on economic impacts of climate change in the future and facilitate policy making involving climate change mitigation and adaptation. In literature looking into effects of weather variables, many studies focus on the average impact of one weather variable and using others as controls. The interaction effects of weather variables are less explored, though the relationship between one weather variable and outcome variable may vary depending on other weather variables. Using outdoor recreation as an example, people may take different decisions of trips in response to temperature depending on the presence or different levels of rainfall. People may also have different tolerance of rainfall to take outdoor activities depending on temperature. This study examines the interaction relation of weather variables, in particular, the impact of interaction of temperature and precipitation on England outdoor recreation trips. Using survey data for outdoor recreation in England and UK climate data, I apply a semi-parametric regression to address this question. Making temperature and precipitation a non-parametric term in the regression allows the exploration of non-linear interaction relation of the two weather variables without restriction of function form or how they interact. A response surface is generated to examine how temperature and precipitation influence number of trips. Results show a general pattern of increase in daily visits along maximum temperature and decrease in daily visits along precipitation. However, variation for gradients along temperature across different levels of total rain is small and there is also little variation of gradients along rain at different maximum temperature. These results suggest that interaction of temperature and precipitation plays a small role for outdoor recreation. The interaction relation is also further explored in different types of trips and different groups of individuals. Interaction plays a small role in most of the types and groups though they may have different shape of response surfaces. More visits are made to town and countryside vs. seaside or coastline. For different groups, female is more sensitive to temperature and precipitation. Individuals with car(s) or dog(s) are slightly more sensitive to temperature at higher temperature and lower level of rain compared to individuals without car or dog. At different life stages, interaction plays a bigger role in the family group compared to independents and empty nesters. They strongly favor outdoor recreation trips at a peak around 25 °C at lower level of rain and the peak becomes much smaller at higher level of rain.

This study focuses on outdoor recreation as outcome, which is an important sector that is highly sensitive to weather variables. In U.S., outdoor recreation industry was $459.8 billion in 2019, which accounted for about 2% of gross domestic product (GDP) (U.S. Bureau of Economic Analysis, 2020). In Great Britain, expenditure on outdoor-related tourism and leisure activities was £36.8 billion in 2019 (Office for National Statistics, 2021b). This is a key spending sector where the total annual family spending is about £820 billion (averaged over financial year ending 2018 to 2020) (Office for National Statistics, 2021a). Furthermore, looking into the impact of interaction of temperature and precipitation, this study focuses on outdoor recreation trips in England, which is in the region considered to have large amount of precipitation and lots of variation. According to State of the UK Climate 2020, UK has shown large annual variation in precipitation data, where recent period of 2011 to 2020 is 4% wetter than period of 1981 to 2010. Monthly and seasonal rainfall patterns also vary strongly within a year (Kendon et al., 2021). Rainfall is also likely to be affected by global warming, projections show that daily rainfall will increase under warming scenarios across UK and number of days with extreme rainfall will also increase (Hanlon et al., 2021). The large variation in precipitation facilitates the study of interaction of temperature and precipitation and results may help provide further insights for climate change.

This paper uses an alternative way to empirically estimate response surface to interaction of weather variables by semi-parametric approach other than regression with certain functional forms of interaction terms. This allows the examination of the non-linear interaction relationship without assuming for specific functional forms. While this paper focuses on outdoor recreation, the empirical method can be used for other economic sectors where temperature and precipitation play an important role, such as agriculture yields, air conditioning demand, etc. The method can also be useful in other settings where interaction of other variables is of interest.

2.2 Literature Review

This paper directly links to other literature that study impact of weather variables and climate change on outdoor recreation. This study closely follows Fezzi et al., 2018, who show that weather patterns largely affect the number of trips of outdoor recreation in England and their projection indicates that there will be greater number of outdoor recreation trips and increase in welfare in the future. This paper further explores the interaction effect of temperature and precipitation in outdoor recreation in England. It also adds knowledge to other literature that aims to understand the impacts of weather and climate change on outdoor recreation from perspective of different types of activities. For example, skiing activities has been one of the research interests. Wake, et al. (2006) suggest that there are fewer skiers' visits during low snow years and ski ticket sales decrease. Scott et al. (2006) examine how adaptation strategies help with ski tourism under climate change. There have also been literature focusing on impacts of weather and climate change on park visitation. Scott et al. (2007) suggest that due to direct impact of climate change, visits to the Waterton Lakes National Park in Canada will increase in mid-century. Hewer and Gough (2016) examined visits to a zoo in Canada and found that temperature is the weather variable that imposes largest impact on zoo visitation. Other than these, impacts on beach and coastal zone visits (Moreno et al., 2008), golf participation (Scott and Jones, 2006), recreational fishing (Dundas and von Haefen, 2020), has also been explored.

This paper is also related to the broad stream of literature that focus on impacts of weather and climate change on various economics performance in different sectors. Outdoor recreation is most closely related to literature focusing on amenity value and leisure. Albouy et al. (2016) examine households' willingness to pay to live in an area depending on local climates in U.S. and their projections show that by the end of the century, there will be 1% to 4% yearly welfare losses of income due to changes in climate amenities. Meier and Rehdanz (2017) explore the willingness to pay for changes in climates in Britain in both housing and labor market and show that compensation for climate amenities mainly comes from housing market. On the other hand, research related to leisure has examined the time allocation between working and leisure hours. Graff Zivin and Neidell (2014) found that increase in higher end temperature leads to decrease in outdoor leisure time for the non-employed and increase in indoor leisure while increase of temperature at the lower end increases outdoor leisure and decreases indoor leisure. Another study focusing on impacts of rain shows that on rainy days, men allocate half an hour to work from leisure (Connolly, 2008). Beyond amenity and leisure, research has also provided evidence of impacts from weather variables and climate change on agriculture sector (Deschênes and Greenstone, 2007, Schlenker and Roberts, 2009), health and mortality (Deschênes and Greenstone, 2011), etc.

2.3 Conceptual Framework and Data

This paper uses panel data analysis to examine the interaction relation of temperature and precipitation on number of outdoor recreation visits. Using county fixed effects, time invariant characteristics are controlled from visitation patterns and the variation of weather variables identifies the impacts on outdoor recreation trips. Detailed daily outdoor recreation trip data used in this paper allows for treatment of weather variables precise to daily level, which facilitates identification and estimation of impacts. To study the interaction relation of temperature and precipitation, the econometric model follows a semi-parametric approach. Keeping various control variables and fixed effects linearly in the model, temperature and precipitation is included as a non-parametric term. This allows the estimation of the non-linear interaction relation for temperature and precipitation, imposing no restrictions on how they interact.

Datasets used by this paper follows Fezzi et al., 2018[1], where outdoor recreation trip data is linked with weather data. The first data source is the Monitor of Engagement with the Natural Environment (MENE)[2] from the UK government. Four years of survey data from March 2009 to February 2013 are used. The MENE survey collects detailed information on visits to the natural environment by people over 16 in England (Natural England, 2013). Each surveyed individual was asked to provide information on their outdoor visits for each of the last 7 days before the survey time, including number of visits and type of visits. Demographic and other individual characteristics are also collected in the survey. About 40,000 individuals are surveyed in England every year.

[1] I thank Carlo Fezzi, Richard Carson, Silvia Ferrini, and Amii Harwood for suppling their data with outdoor recreation trips linked with weather data.
[2] From https://www.gov.uk/topic/outdoor-access-recreation/recreation.

Weather data comes from the British Met Office, where 1.5 km grid weather data is generated by UKV[3] including various weather characteristics. Variables used in the study include maximum temperature, total rain during daylight, total snow during daylight, hours of bright sunshine, which are hours where the median solar radiation is greater than 120W/m^2, hours of cloudy skies, which are hours where median solar radiation between 0 and 120W/m^2 and mean wind speed during daylight (Fezzi et al., 2018). The weather variables are linked with outdoor recreation trips according to residence postcode of individuals surveyed.

This paper focuses on the interaction impact of temperature and precipitation. To be specific, temperature refers to the maximum temperature and precipitation mostly refers to the total rain during daylight in the main specification. Maximum temperature is used in this study considering outdoor visits are mostly likely to be taken during day time. For the same reason, total rain during daylight is used for precipitation. Another form of precipitation to be considered in snowfall. In the main specification, total snow during daylight is included as a control variable and this paper focuses on the interaction of temperature and rainfall. In robustness check, regression using combined total rain and snow as precipitation is explored. The maximum temperature data is distributed with high frequencies at middle range of temperature and the total rain data is highly skewed with most frequencies concentrated at zero and low level of total rain. Since the long right tail of total rain has only few observations but span to as high as over 100 mm, for this study, the observations with total rainfall above 13 mm are excluded to remove the last 1% data at tail. To combine total rain and snow, a new variable is created with name total rain and snow. From conversion table by U.S. National Oceanic and Atmospheric

[3] From https://www.metoffice.gov.uk/ research/news/2012/ukv.

Administration, total snow is converted to a water equivalent measure. The total rain and snow variable is calculated by summing the total rain and converted water equivalence of total snow.

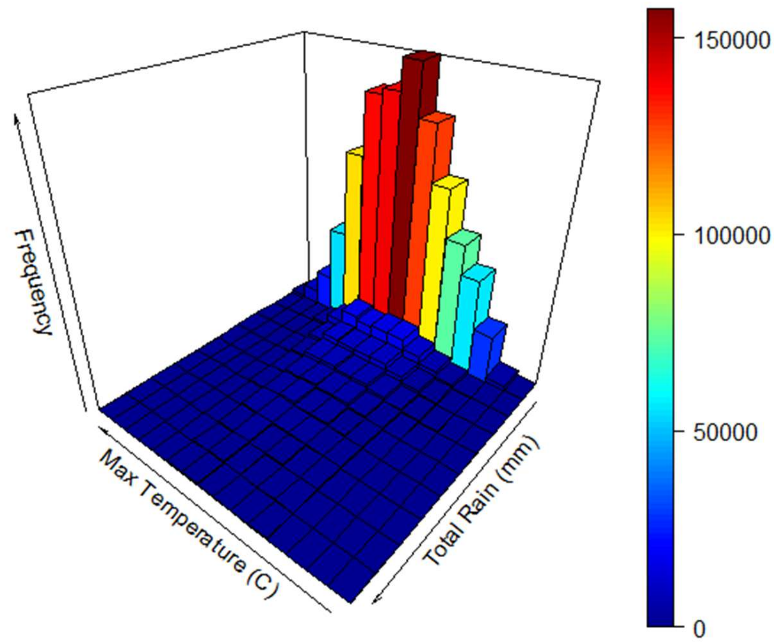


Figure 2.1: Histogram of Maximum Temperature and Total Rain

Note: This figure shows a histogram with frequency of maximum temperature and total rain during daylight in the sample. Higher frequency is represented by darker red and lower frequency is represented by darker blue.

Figure 2.1 shows a two-way histogram representing frequency of maximum temperature and total rain during daylight in the sample. Separate histogram for the two variables can be seen in Figure 2.11 in Appendix. The final sample has 1,272,625 observations. Table 2.1 shows the summary statistics for dependent variable and weather variables. Number of daily visits are from 0 to 3. People do not go out for outdoor recreation for about 85% days, even if people go out, usually only one trip is made. The maximum temperature is from about -7 ℃ to 32 ℃, with mean of about 13 ℃. Mean total rain during daylight is at about 1 mm.

Table 2.1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Daily Visits				
Number of Visits	0.162	0.410	0	3
Percent Visits=0	0.851	0.356	0	1
Percent Visits=1	0.137	0.344	0	1
Percent Visits=2	0.009	0.095	0	1
Percent Visits=3	0.002	0.048	0	1
Weather Variables				
Max Temperature (°C)	13.067	6.103	-6.699	31.799
Total Rain (mm)	0.954	1.993	0	13
Total Rain and Snow (mm)	0.957	1.993	0	13.274
Total Snow (mm)	0.031	0.310	0	15.935
Hours of Sunshine (hrs)	8.554	4.057	0	15
Hours of Cloudy Skies (hrs)	4.428	1.918	1	15
Mean Wind Speed (m/s)	4.326	1.990	0.099	19.126

Note: This table shows the summary statistics for outcome variable and weather variables. 1,272,625 observations are included in the sample.

2.4 Empirical Strategy

To study the interaction relationship of temperature and precipitation, an empirical strategy with the following specification is used:

$$Y_{ipct} = \alpha + g(T_{pct}, P_{pct}) + \delta X_{pct} + \gamma Z_{ipct} + \mu_c + \theta S_t + \varepsilon_{ipct} \tag{2.1}$$

The dependent variable Y_{ipct} is the number of outdoor recreation visits taken for individual i, who lives in postcode p, county c, in day t. The $g(T_{pct}, P_{pct})$ represents the non-parametric term containing maximum temperature and total rain during daylight for a specific postcode on a particular day. These are the variables of interest. Since the $g(T_{pct}, P_{pct})$ is non-parametric, this has advantage over using $T_{pct} + P_{pct} + T_{pct}P_{pct}$ by allowing non-linear relation for temperature and rain, as well as imposing no restrictions on how they interact. X_{pct} includes a group of control for other weather variables indicated in the previous section. Z_{ipct} is individual specific control variables capturing individual characteristics, using variables shown in Table 2.2 in Appendix.

μ_c is the county fixed effects. S_t is used to represent several time fixed effects and indicators, including year fixed effects, quarter fixed effects, day of week fixed effects, as well as indicators of whether it is a holiday and indicator of whether it is during summer school holiday. Year and quarter fixed effects are used instead of month-year fixed effects to control underlying macroeconomics with temperature highly correlated with months.

To estimate the specification, I start from using a binning method for the non-parametric term. Temperature and rain are divided into grids and the interaction term is represented as indicators of each grid. From the temperature dimension, grids are divided according to less than 0 ℃, then each 2.5 ℃ increment until 27.5 ℃ and greater than 27.5 ℃. Along the total rain dimension, grids are divided according to 0 mm, each 1 mm increment until 5 mm, then 5 to 10 mm, and greater than 10 mm. Some grids with fewer observations are further grouped. With binned temperature and precipitation, two types of regressions are run. Following exactly Equation 2.1, an Ordinary Least Square (OLS) is conducted as baseline. Considering the dependent variable is number of trips, being non-negative and has discrete values, a count data model of Poisson regression is also run. Then, to allow for a smooth response surface, a second non-parametric method is used with the Generalized Additive Model (GAM). In the GAM, non-parametric term of temperature and precipitation are included as smooth terms using thin plate regression splines and the rest of the control variables are included as linear terms. Similar to the binning, the GAM is also run in two cases, firstly running regression following normal distribution as a baseline case, then running Poisson regression.

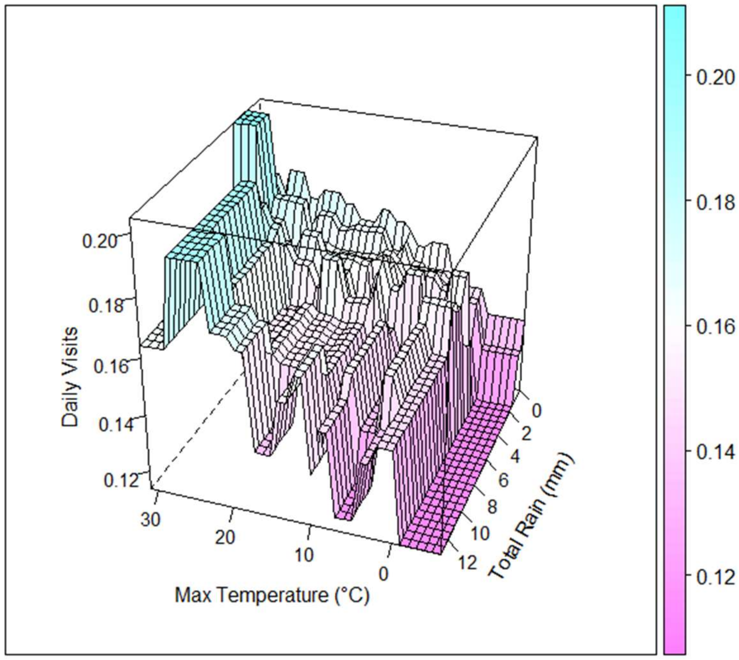
2.5 Results

2.5.1 Main results

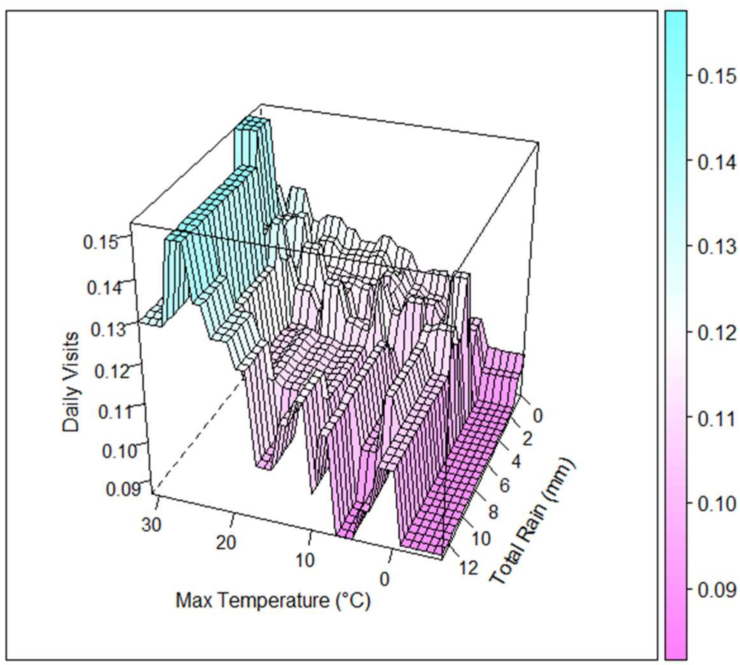
The estimation results using binning for the non-parametric term is shown in Figure 2.2. This figure represents a response surface with the z axis being the daily visits predicted from different levels of maximum temperature and total rain during daylight, while holding other control variables at mean[4]. Figure 2.2 a) shows the 3D response surface of predicted daily visits following OLS regression. The general shape follows expectation, where daily visits increase with the maximum temperature and daily visits decrease with total rain during daylight. At the lower level of rain, there is sharp increase in daily visits in maximum temperature below about 5 °C. The increase trend then becomes gentler until about 25 °C, and there is sharp increase again. While the binned OLS results have variations and not being smooth, it can be seen that the general increase in daily visits along the maximum temperature have similar shape across different levels of total rain. Figure 2.2 b) uses a similar response surface as Figure 2.2 a) to show binning results using Poisson regression. The magnitude of predicted daily visits from Poisson are smaller than predicted values from OLS in general. However, the shape of the response surface in Figure 2.2 b) is very similar to Figure 2.2 a).

The results from GAM with thin plate regression spline for non-parametric term are shown in Figure 2.3. Figure 2.3 a) represents results of response surface from GAM following normal distribution while Figure 2.3 b) shows results of GAM using Poisson regression. As in Figure 2.2, the z axis is predicted number of daily visits. The two response surfaces follow a similar pattern as results in the binning method. They also show increase of daily visits along

[4] Predicted response surfaces are shown in this paper and prediction intervals can be computed by bootstrap.



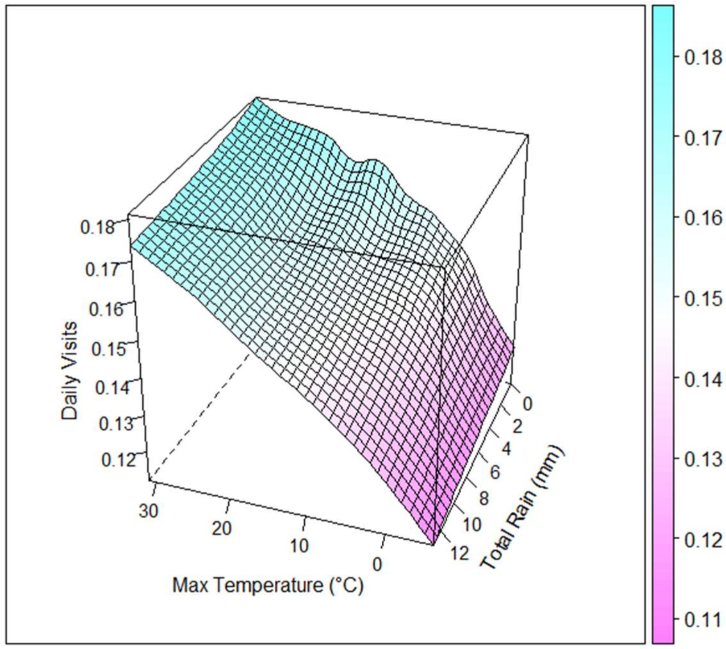
(a) Baseline



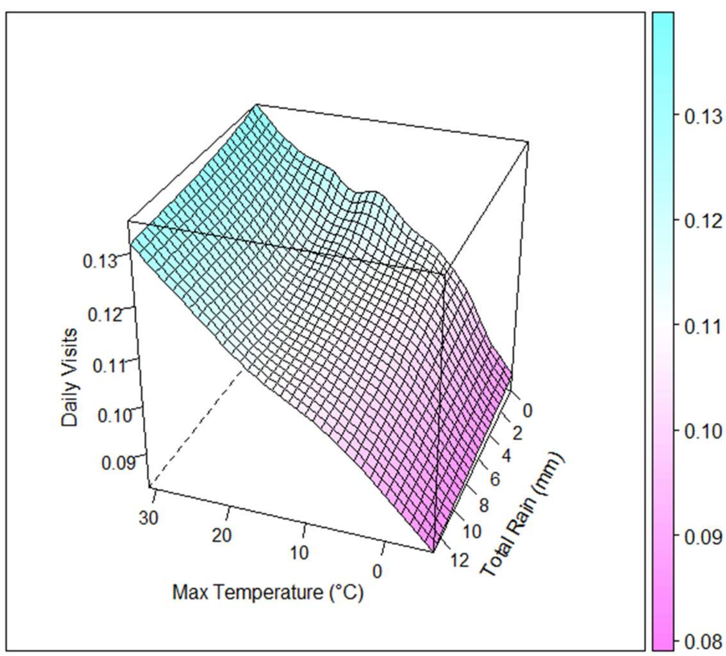
(b) Poisson

Figure 2.2: Response Surfaces from Binning

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight with non-parametric term binned. (a) follows OLS and (b) uses Poisson regression.



(a) Baseline



(b) Poisson

Figure 2.3: Response Surfaces from GAM

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM with thin plate regression spline for non-parametric term. (a) follows normal distribution and (b) uses Poisson regression.

maximum temperature and decrease of daily visits along total rain during daylight. Between baseline and Poisson results, the response surfaces also show a similar shape, with slightly lower predicted daily visits for GAM Poisson regression. The GAM results allow for smooth response surfaces compared to binning method. Since GAM allows for a smooth surface and GAM Poisson regression has lower residual sum of square compared to baseline GAM, GAM Poisson regression is the preferred specification.

This result is then taken a closer look by plotting the cross sections of this response surface along maximum temperature and total rain during daylight. Figure 2.4 shows cross section cuts at different levels of total rain during daylight. To facilitate understanding of results, the y axis is converted to index of daily visits, which converts predicted visits by linear rescaling, so that the lowest predicted visit has index of 0 and highest predicted visit has index of 1. When there is no rain, the predicted daily visits firstly increase sharply over 0 ℃, then there is generally gentler increase trend with local peaks, lastly when it is closer to 25 ℃, there is slightly sharper increase in temperature. There is a local maximum of predicted visits at around 15 ℃, followed with a smaller peak above 20 ℃. These peaks are likely to reflect temperature being appropriate for certain types of recreation activities. The sharp increase in lower temperature may represent people's preference over warm weather. When there is rainfall during daylight, as amount of rain increase up until about 5 mm, the local peak fades off, suggesting that people won't be favoring the recreation trips around certain temperature any more if rain gets heavier. However, the shape of sharp increase with temperature at lower temperature, followed by gentler increase, then shaper increase again at higher temperature remains. When rain increases beyond 5mm, the non-linear relation becomes relatively linear. This is partly due to fewer number of observations at higher level of rainfall. Overall, the relation between outdoor recreation trips and temperature

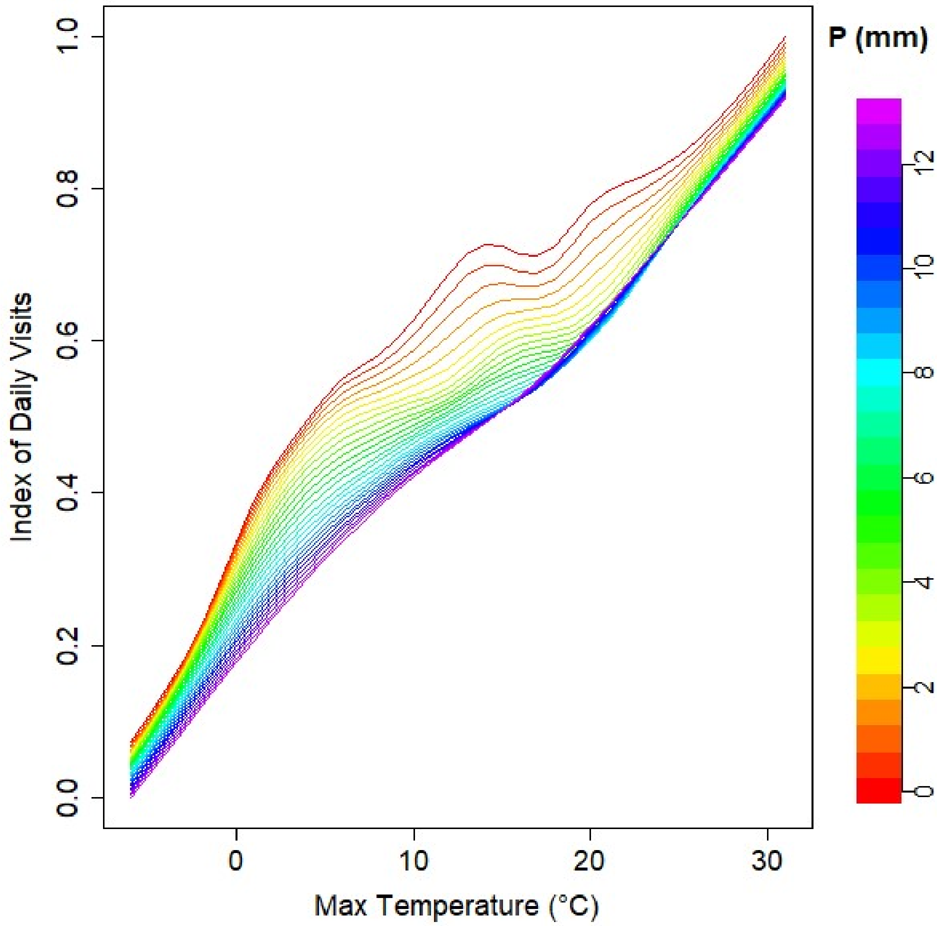


Figure 2.4: Cross Sections of Response Surface at Different Levels of Total Rain

Note: This figure shows cross section cuts of response surface using GAM Poisson regression. Each line represents index of daily visits vs. maximum temperature at different levels of total rain during daylight. Predicted daily visits are linearly rescaled as index so that it is 0 for the lowest predicted daily visits of the response surface and 1 for the highest. The color of each line represents the amount of total rain during daylight, with axis to the right.

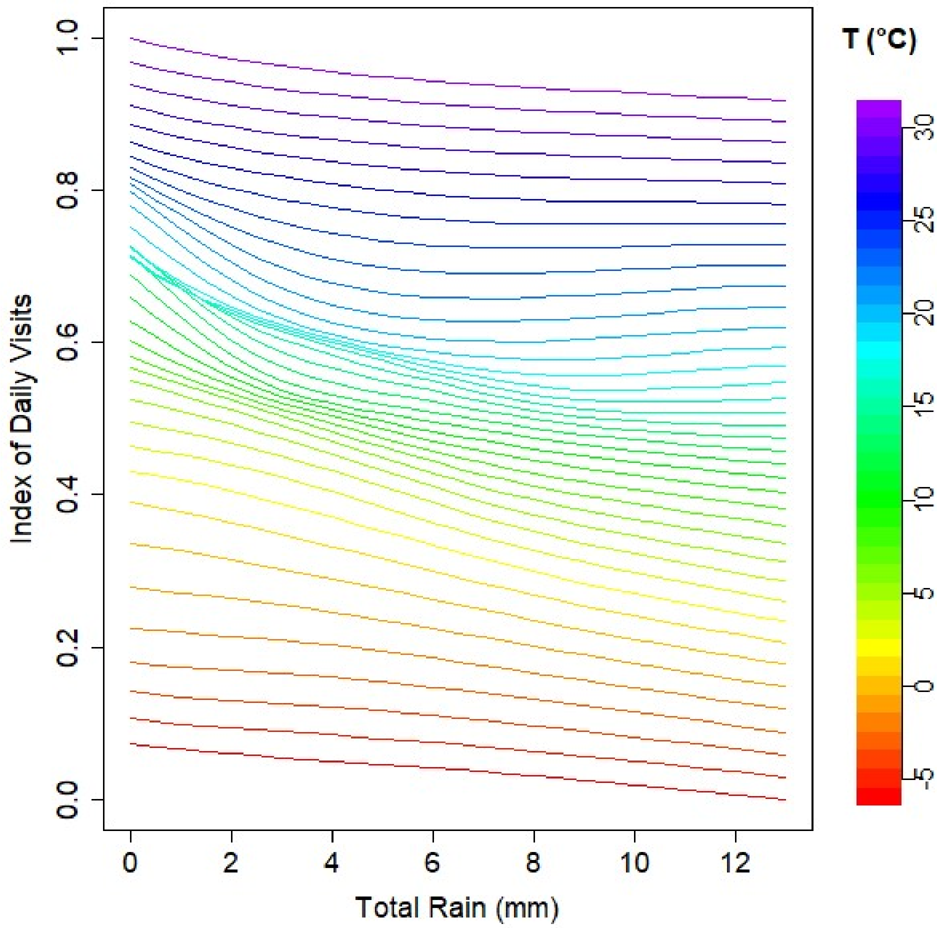


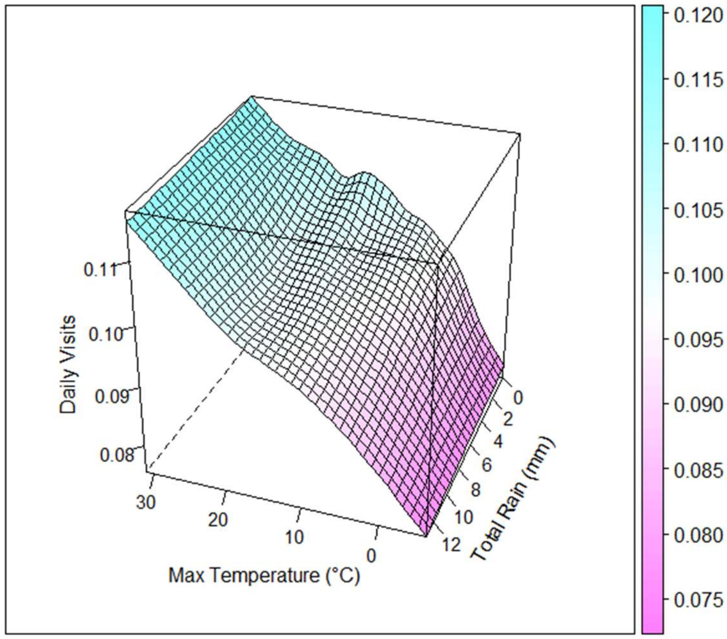
Figure 2.5: Cross Sections of Response Surface at Different Maximum Temperature

Note: This figure shows cross section cuts of response surface using GAM Poisson regression. Each line represents index of daily visits vs. total rain during daylight at different levels of maximum temperature. Predicted daily visits are linearly rescaled as index so that it is 0 for the lowest predicted daily visits of the response surface and 1 for the highest. The color of each line represents the levels of maximum temperature, with axis to the right.

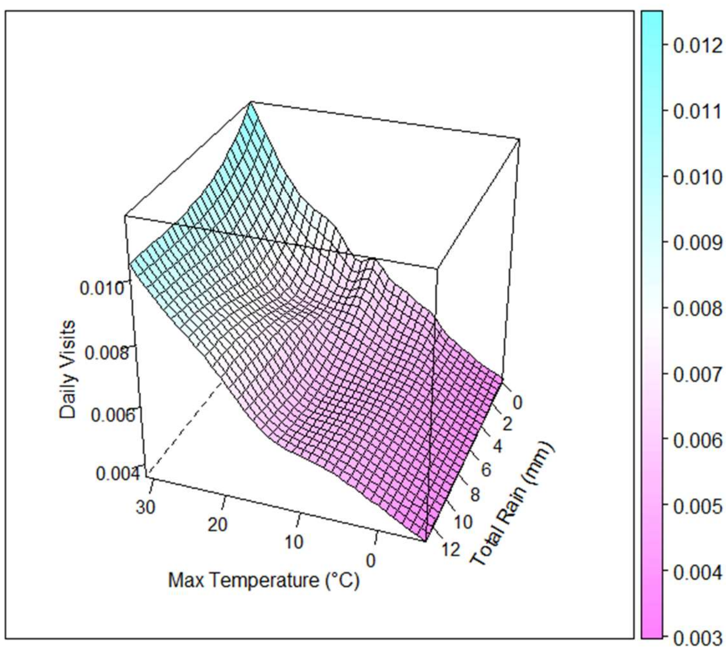
across precipitation is relatively consistent and there is only small variation in local peaks. Figure 2.5 shows the cross sections cuts at different level of maximum temperature. At lower and higher level of temperature, predicted daily visits decrease with rain relatively linearly. At middle range of temperature from about 8 ℃ to 28 ℃, there is a sharper decrease of visits at lower level of rain, then followed by gentler decrease of visits. Similarly, relation between recreation trips and rain across temperature is relatively similar with small differences at middle temperature and lower level of rain. These results suggest that interaction of temperature and precipitation only plays a small role in outdoor recreation visits.

2.5.2 Results for Different Trip Types and Individual Groups

Next, the paper shows results for different types of trips and results for different individual groups. The MENE data separates the outdoor recreation trips taken by four types: trips in town or city, trips in countryside, trips in a seaside resort or town, and trips at other seaside coastline. I further group trips into two types: trips in town or country side and trips in seaside resort or other coastline. I then rerun the GAM Poisson regression for each type of trips, i.e., the outcome variable becomes number of trips in town or country side or number of trips in seaside resort or other coastline. Figure 2.6 shows the results for two types of trips. There are much fewer number of trips to seaside resort or other coastline and predicted visits in (b) are much lower than (a). The two graphs use different scales of z-axis to show the shape of response surface. Figure 2.6 (a) follows a similar shape as main regression results, also showing small variations for interaction with local peaks at small amount of rain. Figure 2.6 (b) also indicates that there are local peaks at lower level of rain. Overall, the gradients of daily visits along temperature across rain and gradients of daily visits along rain across temperature are relatively consistent.



(a) Town or Countryside



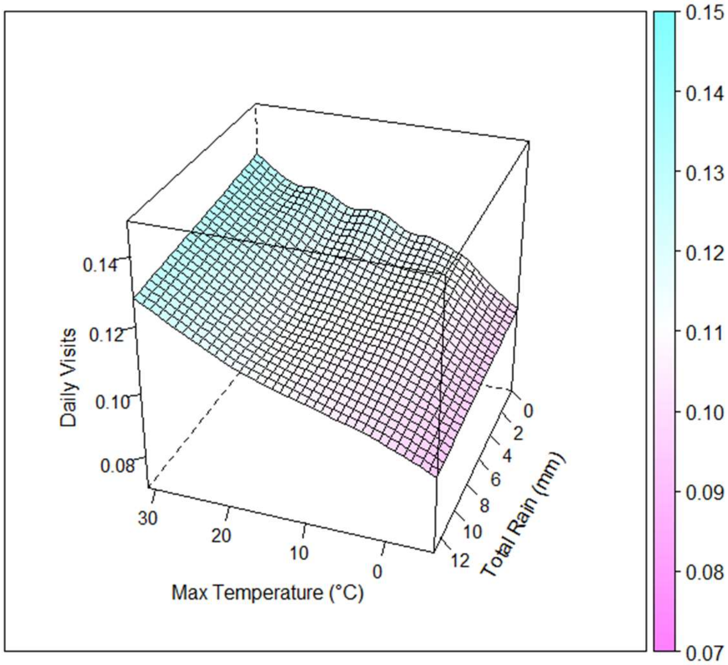
(b) Seaside resort or other coastline

Figure 2.6: Response Surfaces for Different Types of Trips

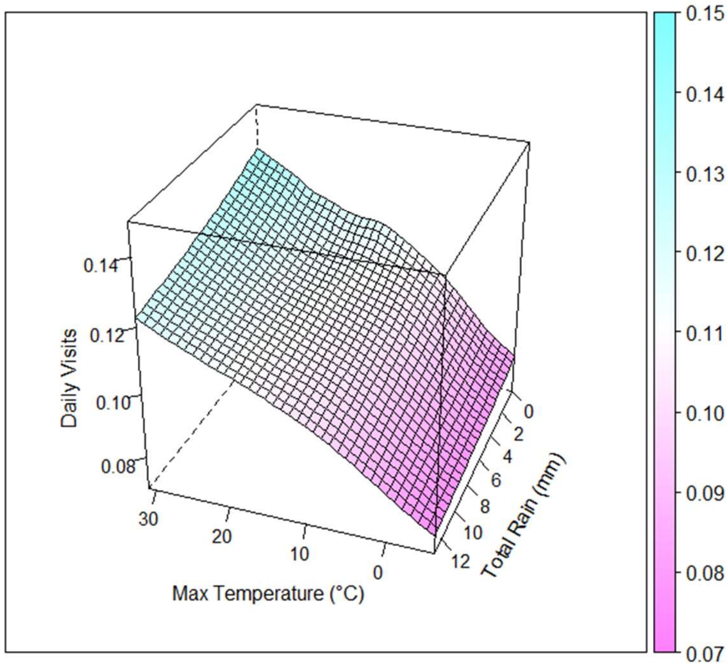
Note: This figure shows different types of predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. (a) uses daily visits in town or countryside as outcome variable and (b) uses daily visits in seaside resort or other coastline as outcome variable.

Using total number of outdoor recreation daily visits, GAM Poisson regressions are also run separately divided by individual groups to examine whether different characteristics of individuals generate different response surfaces. The first comparison is between male and female, shown in Figure 2.7. To allow for easier comparison for groups, z-axis and color key of response surfaces are scaled to be the same within each comparison. Both male and female have similar number of outdoor recreational trips. However, male is less sensitive to temperature or rain. The increase of maximum temperature and decrease in rain during daylight is associated with less change of daily visits for male than female. Similar to the main results, interaction plays a small role for both male and female. Both show a local peak at around 15 ℃ and fades away as amount of rain increases. In general, there are small differences for relationship between predicted visits and temperature across rain or predicted visits and rain across temperature.

The next feature examined is life stage. Regression is run separately by dividing individuals to four life stage: empty nester, family, older independent, young independent. Results are shown in Figure 2.8. The four life stage groups also have similar number of outdoor recreational trips in general, with young independents having slightly fewer number of trips, but they vary in shapes of response surfaces. Empty nesters have relatively sharp increase in lower temperature and milder increase in higher temperature, suggesting their dislike of cold weather. Similarly, as main results, at lower amount of rain, relation between predicted daily visits vs. temperature show local peaks, but fades off as rain increases. Interaction plays a bigger role for the family group. There is sharp increase in visits with temperature and a large peak around 25 ℃ at lower level of rain during daylight, this peak becomes smaller as the amount of rain increases and at higher level of rain, the increase in visits with temperature is gentler with much smaller peak. The response surface shows that interaction affects predicted visits for families and



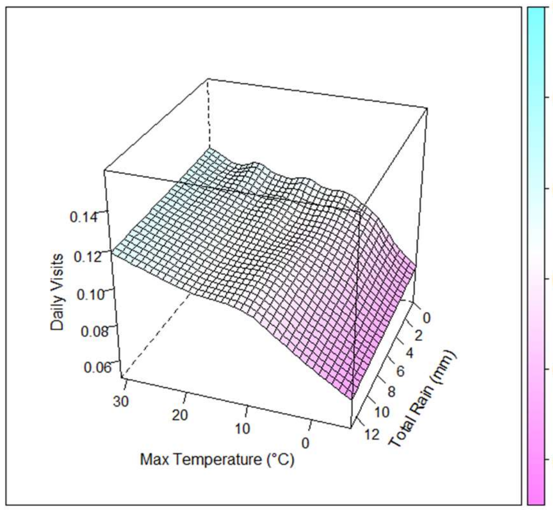
(a) Male



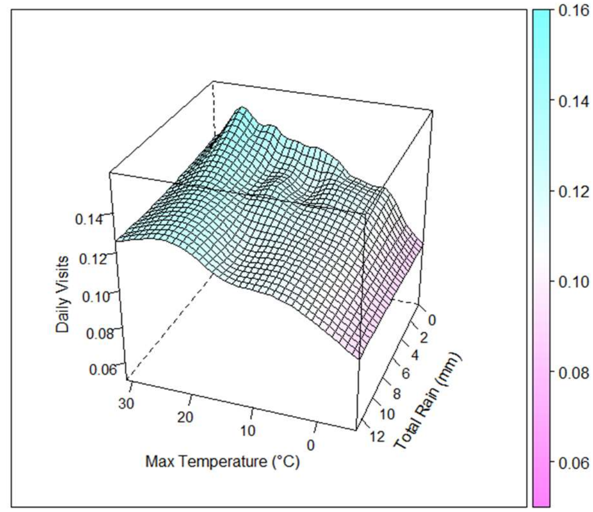
(b) Female

Figure 2.7: Response Surfaces for Male and Female

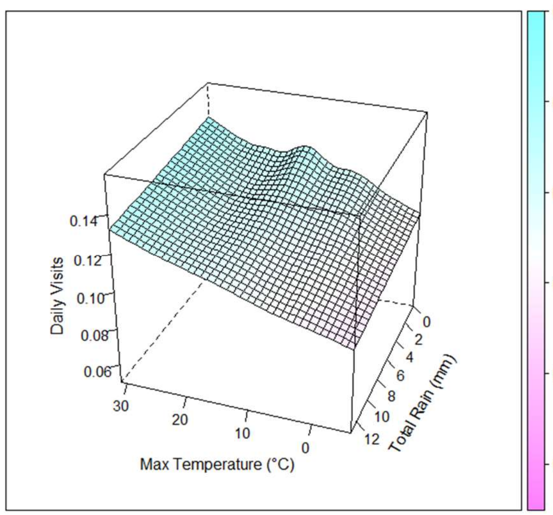
Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. (a) only uses data for male and (b) only uses data for female.



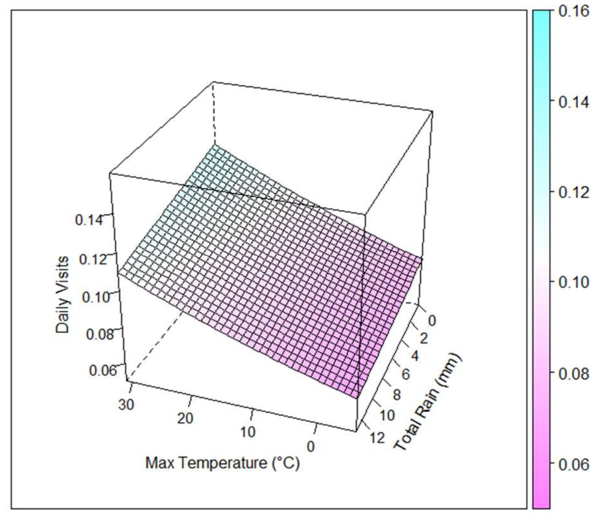
(a) Empty Nester



(b) Family



(c) Older Independent



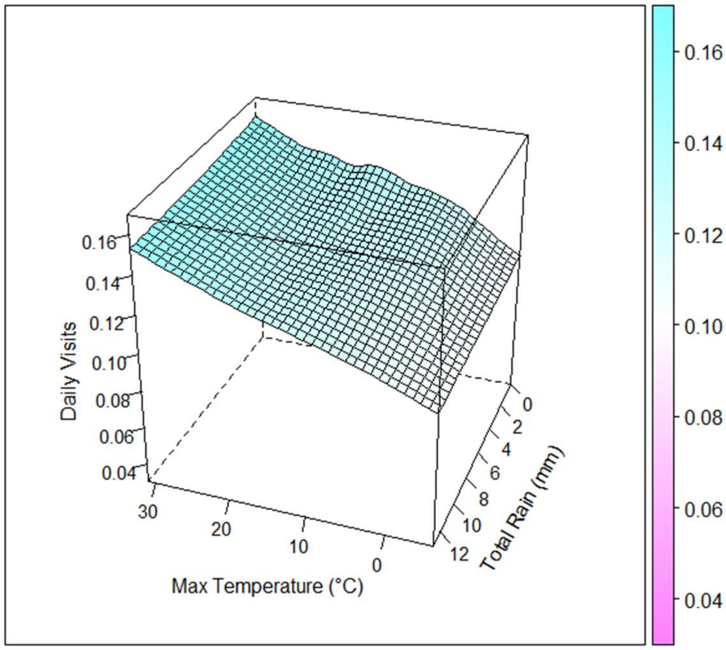
(d) Young Independent

Figure 2.8: Response Surfaces for Different Life Stages

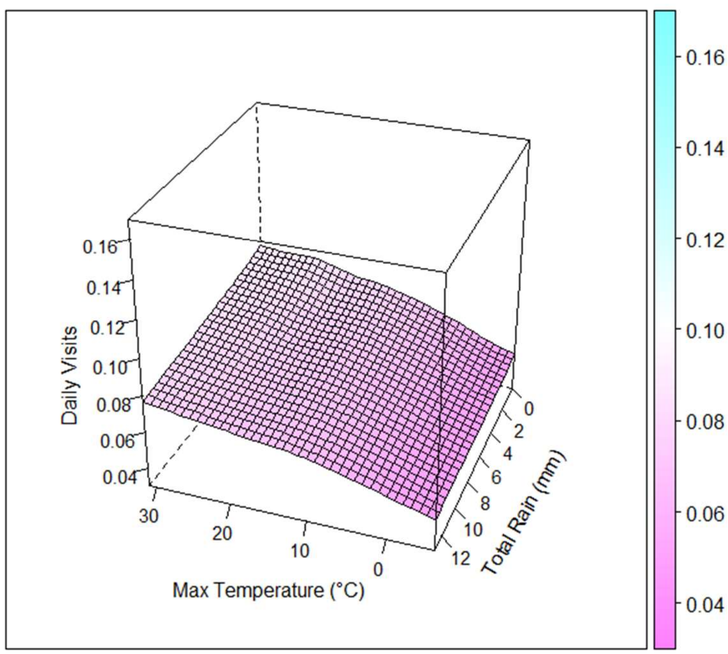
Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. (a) only uses data for individuals who are empty nesters, (b) only uses data for the family group, (c) only uses data for individuals who are older independents, and (d) only uses data for individuals who are young independent.

suggests that they strongly favor a temperature at around 25 °C at lower level of rain. Interaction plays a small role for both older and young independents. Older independents are relatively less sensitive to temperature and rain. Again, there is local peak of visits at around 15 °C at lower level of rain and peak fades off as rain increases. Young independents do not show a local peak and change in visits along temperature or rain are mostly linear.

Individuals with car(s) or with dog(s) are examined as well. Figure 2.9 shows results for individuals with car(s) or not. Number of visits predicted for individuals with car(s) are slightly higher than individuals without car. Ownership of car(s) may provide greater mobility to go out and people who own car(s) may also be the group who prefer outdoor trips more. Similar to the main results, interaction plays a small role here. For individual with car(s), there is again a small local peak of visits around 15 °C at lower level of rain, and the peak fades with greater amount of rain. After the peak, at higher temperature and lower level of rain, the increase in daily visits along temperature is steeper than people without car. Figure 2.10 shows results for individuals with dog(s) or not. The case for individuals with dog(s) indicate that the number of predicted trips is much higher than the case for individuals without dog, consistent with the need to walk dogs. For individuals with dog(s), the increase in temperature associates with a slightly sharper increase of daily visits at higher temperature and at lower amount of rain. Compared to individuals without dog, the increase of temperature leads to a milder increase of visits at higher level of temperature and lower amount of rain. Similar to most of the groups, interaction plays a small role for both cases.



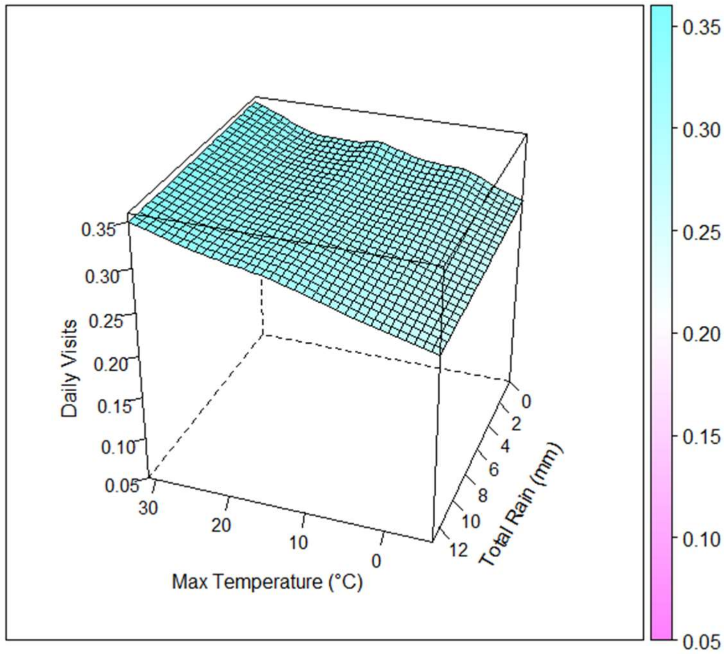
(a) With Car(s)



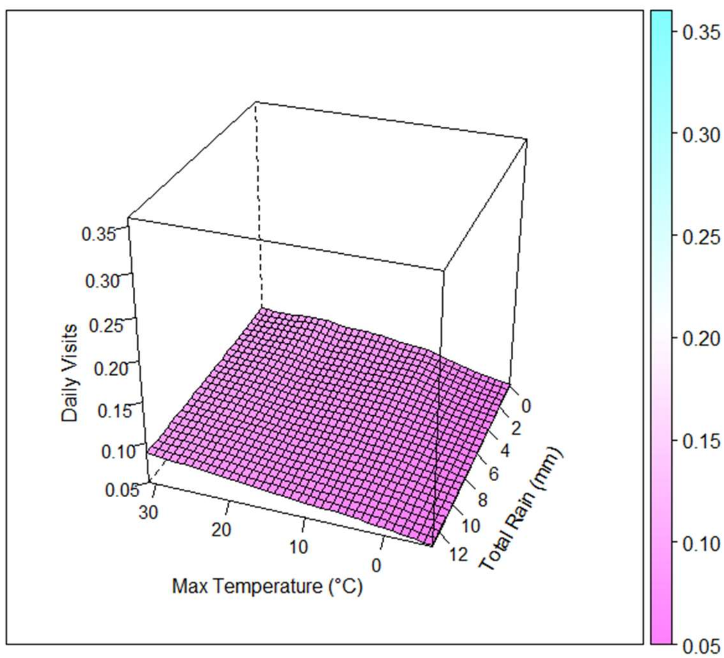
(b) Without Car

Figure 2.9: Response Surfaces for Individuals with Car(s) or Not

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. (a) only uses data for individuals with car(s), (b) only uses data for individuals without car.



(a) With Dog(s)



(b) Without Dog

Figure 2.10: Response Surfaces for Individuals with Dog(s) or Not

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. (a) only uses data for individuals with dog(s), (b) only uses data for individuals without dog.

2.5.3 Robustness Check

Various robustness checks are conducted. This study mainly focuses on rain as the precipitation variable. Another type of precipitation to be considered is snow. Snow is solid instead of liquid and thus is likely to lead to different response in visits and different interaction with temperature. I examine the interaction relation using combined precipitation of both rain and snow, converting snow into a water equivalence as described in data section and add to total rain, and see whether response surface changes. Regression is run using this total rain and snow variable to replace total rain in the main specification. The control variable of total snow is then switched to indicator of whether there is snow on a particular day. Results are shown in Figure 2.12 in Appendix and the response surface is similar to that of the main specification. Though the relation between daily visits and total rain and snow is slightly flatter compared to the main specification, combining rain and snow together does not largely change the results. In general, the interaction still plays a small role. In addition, regression is also restricted to a sample where all the days with snow are excluded, with results in Figure 2.13 in Appendix. The response surface also has a similar shape as main specification, with decrease of predicted number of trips with rain slightly steeper than main results. While this study does not specifically explore the difference between interaction effect by rain and snow, this could be explored in future research.

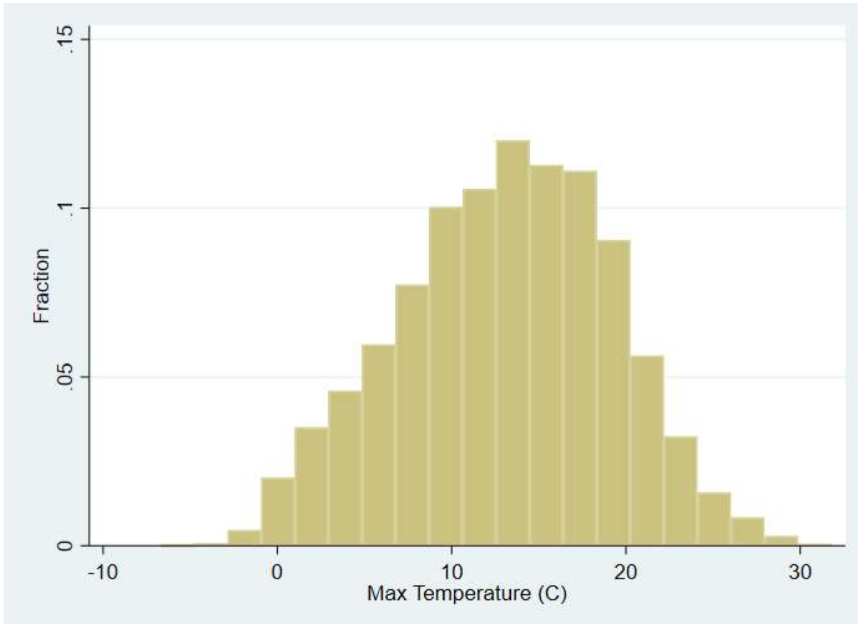
This paper focuses on number trips as the outcome variable. I also examine the results of using indicator of whether individuals have outdoor recreation trips to examine whether there is different response to this binary outcome. Results of Figure 2.14 in Appendix now show z-axis to be has visits, which can be interpreted as predicted probability of having an outdoor recreation trip. This response surface is similar to the main results, suggesting a small interaction effect also to individual's decision of whether to go for an outdoor recreation trip or not. The preferred specification follows GAM Poisson regression, which assumes variance equals to the mean. GAM is also run with Negative Binomial regression to relax this assumption and allow for variance to exceed the mean. Results are shown in Figure 2.15 in Appendix, suggesting a similar response surface as Poisson regression. Lastly, I also explore robustness in terms of changing time fixed effects. The main specification uses year and quarter fixed effects. Regressions are also run using year-quarter fixed effects, as well as year and month fixed effects. Figure 2.16 and Figure 2.17 in Appendix also indicate consistent results as the main specification.

2.6 Conclusion

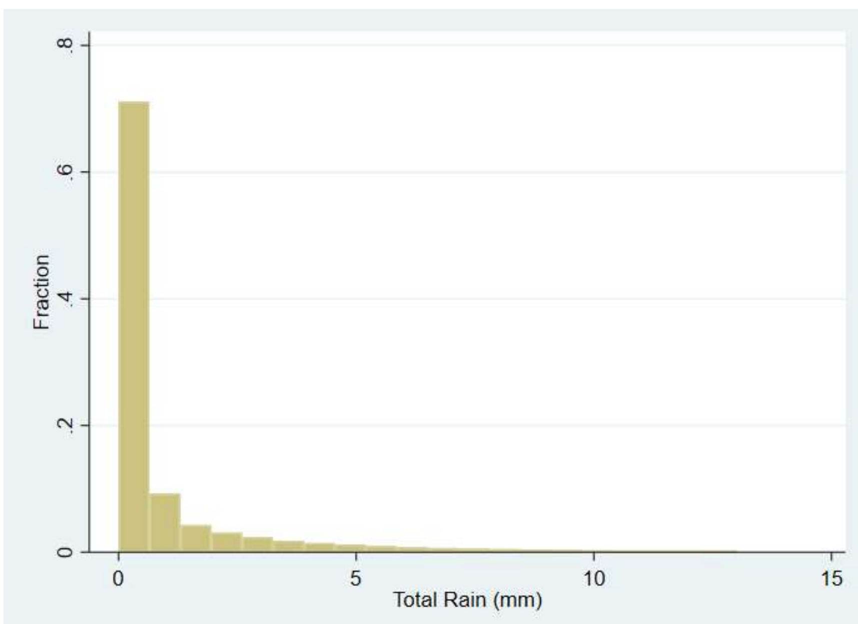
This paper studies the interaction relation of temperature and precipitation with outdoor recreation trips. Using detailed information for trips in England and a semi-parametric approach, I show response surfaces of predicted daily visits corresponding to maximum temperature and total rain during daylight. The results suggest that while there is increase of visits in temperature and decrease of visits in rain, the increasing gradient along temperature has only small variations across rain and the decreasing gradient with rain is similar across temperature. The interaction plays a small role for outdoor recreation. This paper also shows results for different types of trips, where the number of visits for town and countryside is greater than those for seaside and coastline. Gradients along temperature and precipitation are relatively consistent. Looking at different groups of individuals, interaction also plays a small role in most cases, though shapes of response surfaces vary. Female is more sensitive to both temperature and rain. Individuals with car(s) and with dog(s) are more sensitive to temperature at higher temperature and lower level of rain. For individuals in different life stage, interaction plays a greater role for people in the family group. The family group strongly favors outdoor recreation around 25 ℃ at lower level of rain during daylight, and leads to a peak in the response surface. This peak becomes much smaller as the amount of rain increases to higher level.

The results in this paper suggest that interaction of temperature and precipitation plays a small role in outdoor recreation in England. The method can be applied to other sectors, such as agriculture, labor market, etc., as well as other areas, and results may differ depending on specific sector and geographic region. In this paper, precipitation mostly refers to total rain during daylight in the main specification. How outdoor recreation reacts to interaction of temperature and snow could be further explored in future studies. Furthermore, impact from interaction of other weather variables, such as interaction of temperature and humidity, may play an important role depending on sectors and could also be further studied in future research.

2.7 Appendix



(a) Histogram for Maximum Temperature



(b) Histogram for Total Rain During Daylight

Figure 2.11: Histogram for Maximum Temperature and Total Rain During Daylight

Note: This figure shows histogram for maximum temperature and total rain during daylight separately in (a) and (b).

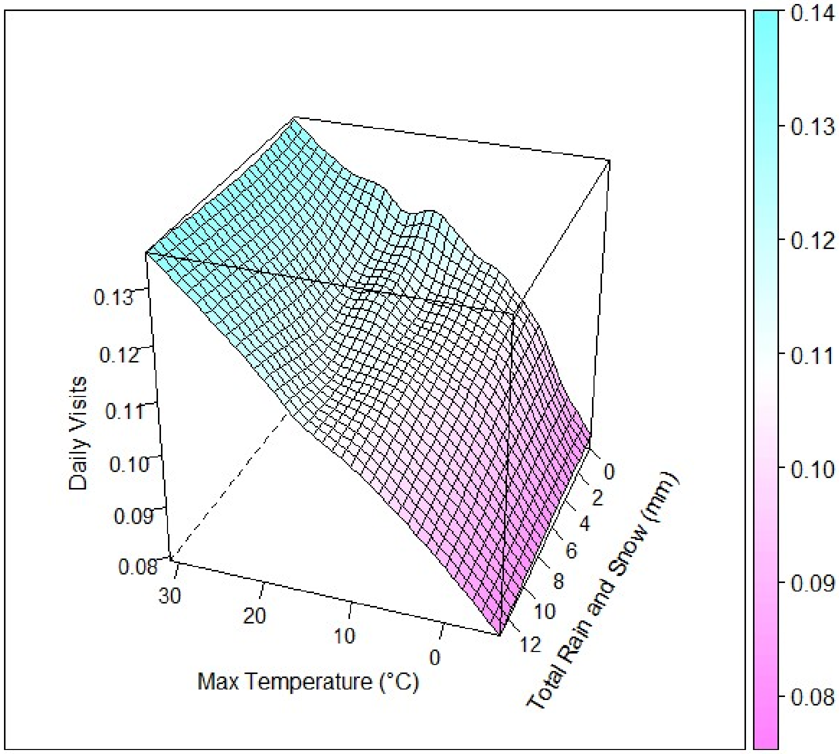


Figure 2.12: Response Surface with Total Rain and Snow from GAM

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain and snow during daylight from GAM Poisson regression.

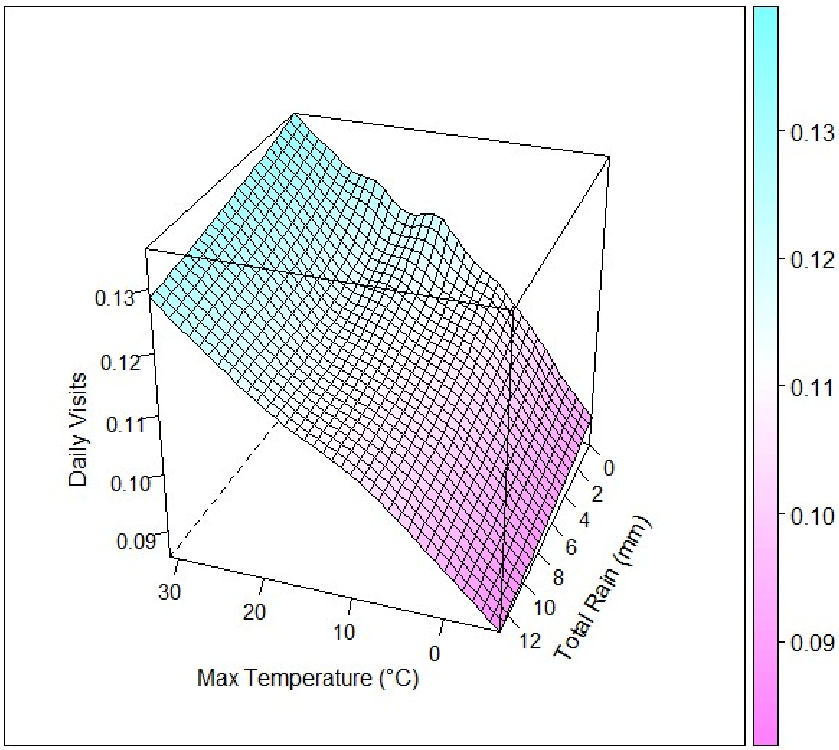


Figure 2.13: Response Surface for Observations Without Snow

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. Only observations without snow are included.

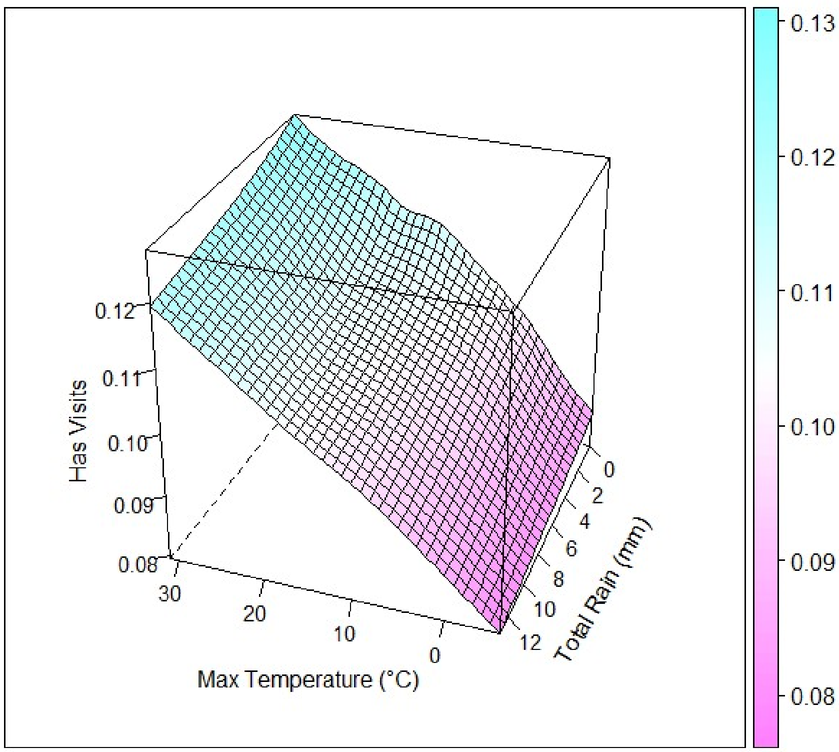


Figure 2.14: Response Surface with Has Visits as Outcome

Note: This figure represents predicted probability of visit as a response surface of maximum temperature and total rain and snow during daylight from GAM Poisson regression.

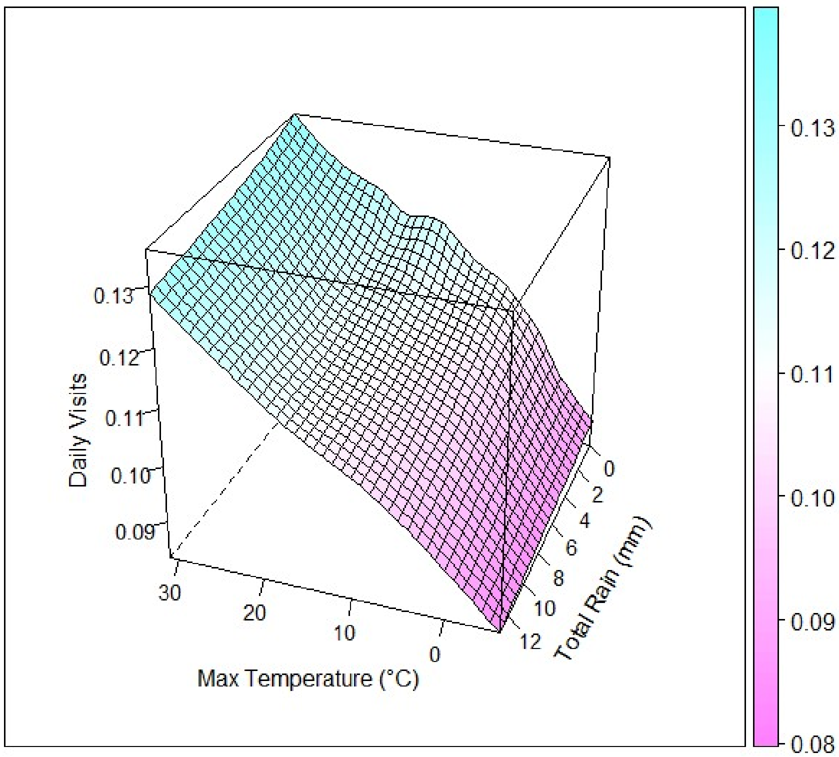


Figure 2.15: Response Surface following Negative Binomial Regression

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM following Negative Binomial regression.

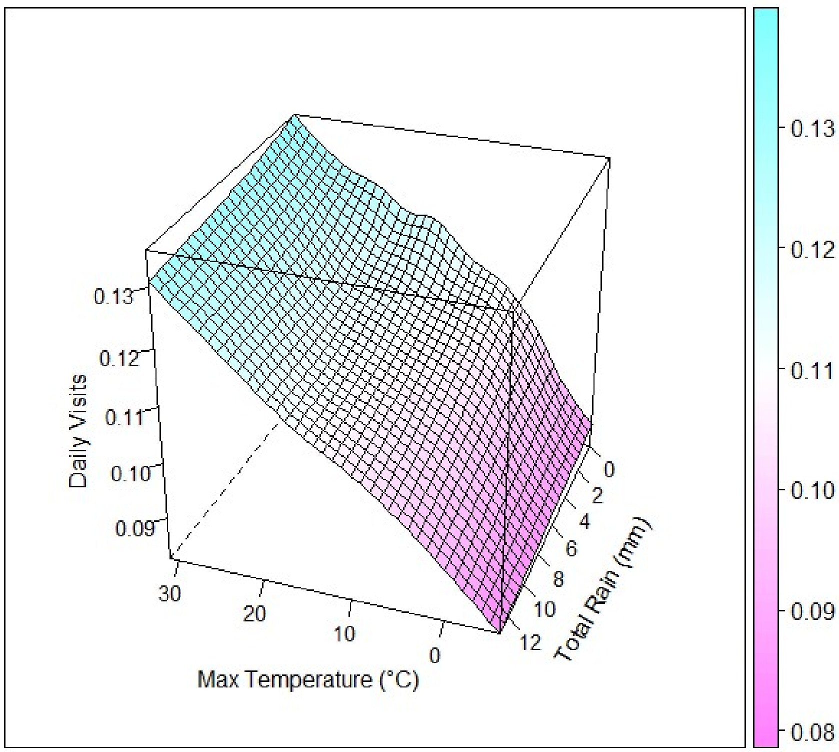


Figure 2.16: Response Surface Using Year-Quarter Fixed Effects

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. Year-quarter fixed effects are used instead of the year and quarter fixed effects in main specification.

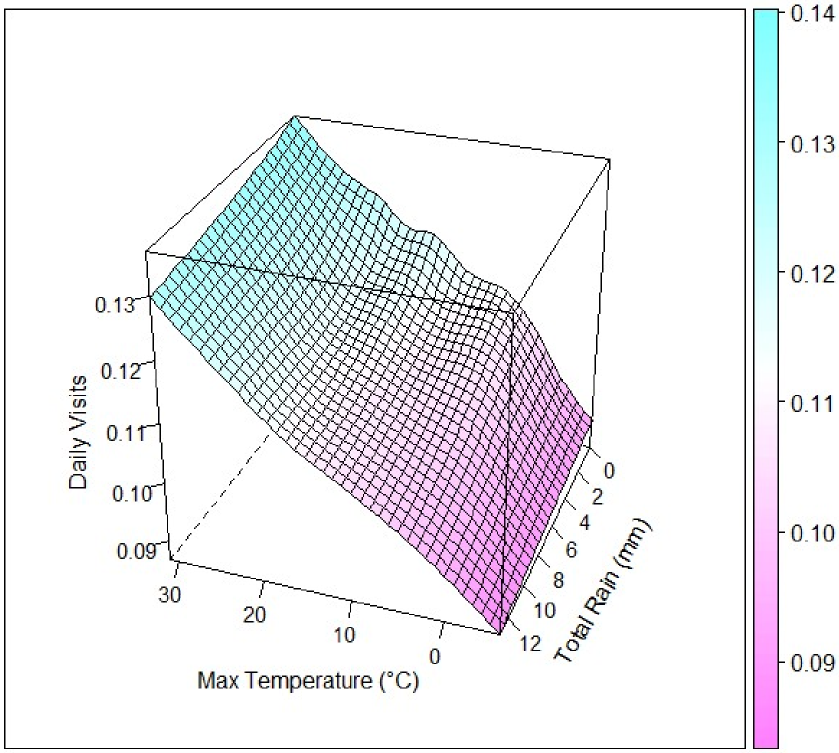


Figure 2.17: Response Surface Using Year and Month Fixed Effects

Note: This figure shows predicted daily visits as a response surface of maximum temperature and total rain during daylight from GAM Poisson regression. Year and month fixed effects are used instead of the year and quarter fixed effects in main specification.

Table 2.2: Summary Statistics of Individual Control Variables

Variable	Mean	Std. Dev.	Min.	Max.
Age				
20	0.135	0.342	0	1
30	0.159	0.366	0	1
40	0.164	0.370	0	1
50	0.152	0.359	0	1
60	0.144	0.351	0	1
75	0.246	0.430	0	1
Male	0.464	0.499	0	1
Race				
White	0.867	0.340	0	1
Black	0.039	0.194	0	1
Asian	0.076	0.265	0	1
Others	0.018	0.133	0	1
Marital Status				
Single	0.252	0.434	0	1
Married	0.569	0.495	0	1
Others	0.179	0.383	0	1
Work Status				
At School	0.007	0.086	0	1
Full Time 30+ hrs	0.352	0.477	0	1
Full Time Higher Education	0.053	0.224	0	1
Not Seeking	0.109	0.312	0	1
Part Time 8-29 hrs	0.119	0.324	0	1
Part Time < 8 hrs	0.006	0.077	0	1
Retired	0.288	0.453	0	1
Unemployed	0.066	0.248	0	1
Social Economic Groups				
Senior Manager or Professional	0.183	0.387	0	1
Clerical, Administrative	0.268	0.443	0	1
Skilled Worker	0.204	0.403	0	1
Unskilled Worker, etc.	0.345	0.475	0	1
Life Stage				
Empty Nester	0.381	0.486	0	1
Family	0.303	0.459	0	1
Older Independent	0.158	0.364	0	1
Young Independent	0.159	0.366	0	1
Physical	2.279	2.558	0	7
Tenure				
Mortgage	0.272	0.445	0	1
Owned Outright	0.322	0.467	0	1
Rent Local Authority	0.177	0.381	0	1
Rent Private	0.184	0.388	0	1

Table 2.2: Summary Statistics of Individual Control Variables (Continued)

Other	0.045	0.207	0	1
Disability	0.210	0.408	0	1
Car	0.708	0.455	0	1
Dog	0.226	0.418	0	1
Average Income	32495.33	11624.85	9168.5	127069.5

Note: This table shows the summary statistics for individual level control variables representing individual characteristics. 1,272,625 observations are included in the sample.

2.8 References

Albouy, D., Graf, W., Kellogg, R., and Wolff, H. (2016). Climate amenities, climate change, and American quality of life. *Journal of the Association of Environmental and Resource Economists*, 3(1):205–246.

Connolly, M. (2008). Here comes the rain again: Weather and the intertemporal substitution of leisure. *Journal of Labor Economics*, 26(1):73–100.

Deschênes, O. and Greenstone, M. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97:354–385.

Deschênes, O. and Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics*, 3(4):152–185.

Dundas, S. J. and von Haefen, R. H. (2020). The effects of weather on recreational fishing demand and adaptation: Implications for a changing climate. *Journal of the Association of Environmental and Resource Economists*, 7(2):209–242.

Fezzi, C., Carson, R., Ferrini, S., and Harwood, A. (2018). *The influence of weather on outdoor recreation: Forecasting the implications of climate change*. Working Paper. Department of Economics and Management, University of Trento.

Hanlon, H. M., Bernie, D., Carigi, G., and Lowe, J. A. (2021). Future changes to high impact weather in the UK. *Climatic Change*, 166(50):1–23.

Hewer, M. J. and Gough, W. A. (2016). Weather sensitivity for zoo visitation in Toronto, Canada: a quantitative analysis of historical data. *International Journal of Biometeorology*, 60:1645–1660.

Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., Sparks, T., and Garforth, J. (2021). State of the UK climate 2020. *International Journal of Climatology*, 41(2):1–76.

Meier, H. and Rehdanz, K. (2017). The amenity value of the British climate. *Urban Studies*, 54(5):1235–1262.

Moreno, A., Amelung, B., and Santamarta, L. (2008). Linking beach recreation to weather conditions: A case study in Zandvoort, Netherlands. *Tourism in Marine Environments*, 5(2-3):111–119.

Natural England (2013). *Monitor of engagement with the natural environment: The national survey on people and the natural environment: Technical report (2012-13 survey)*. Natural England Commissioned Report NECR123.

Office for National Statistics (2021a). *Family spending in the UK: April 2019 to March 2020*. https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/.

Office for National Statistics (2021b). *Tourism and outdoor leisure accounts, natural capital, UK: 2021*. https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/tourismandoutdoorleisureaccountsnaturalcapitaluk/2021.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37):15594–15598.

Scott, D. and Jones, B. (2006). The impact of climate change on golf participation in the Greater Toronto Area (GTA): A case study. *Journal of Leisure Research*, 38(3):363–380.

Scott, D., Jones, B., and Konopek, J. (2007). Implications of climate and environmental change for nature-based tourism in the Canadian Rocky Mountains: A case study of Waterton Lakes National Park. *Tourism Management*, 28:570–579.

Scott, D., McBoyle, G., Minogue, A., and Mills, B. (2006). Climate change and the sustainability of ski-based tourism in Eastern North America: A reassessment. *Journal of Sustainable Tourism*, 14(4):376–398.

U.S. Bureau of Economic Analysis (2020). *Outdoor recreation satellite account, U.S. and states, 2019*. https://www.bea.gov/news/2020/outdoorrecreation-satellite-account-us-and-states-2019.

U.S. National Oceanic and Atmospheric Administration. *Estimating the water equivalence of snow*. https://www.ncdc.noaa.gov/sites/default/files/attachments/Estimating_the_Water_Equivalent_of_Snow.pdf.

Wake, C., Burakowski, E., and Goss, L. (2006). *Winter recreation and climate variability in New Hampshire: 1984 - 2006*. The Carbon Coalition and Clean Air – Cool Planet, Portsmouth, NH.

Graff Zivin, J. and Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26.

Chapter 3

The Role of Ridesharing in Changing Urban Trip Patterns

Abstract: Ridesharing is becoming an important consideration in many discussions involving urban planning and transportation initiatives. This paper looks at how the introduction of ridesharing services such as Uber and Lyft has influenced trip choice decisions. Using data from the 2009 and 2017 National Household Travel Surveys, I show that the longer Uber and Lyft have been in an urban market, the greater the increase in the 2017 survey trips that were made using taxi/rideshare services relative to the 2009 survey benchmark. This increase is driven by an upward shift in the utilization of taxi/rideshare services that is more pronounced for short and longer distance trips than for middle distance trips. This upward shift is also more pronounced on weekdays relative to weekends. Ridesharing services are shown to be a substitute for short haul bus trips, but a complement with longer rail trips.

3.1 Introduction

Policies on urban planning related to transportation and its environmental impacts have long received considerable attention. Traditionally, mass public transit has been seen as one of the major tools for decreasing driving to reduce traffic congestion and air pollution. In recent years, a new transportation choice has been added to the traditional mix of buses, cars, taxi cabs and rails by ridesharing companies such as Uber and Lyft. Ridesharing companies allow people to easily arrange a trip through mobile-app and then be picked up by a driver and taken to their destination. These services have been widely adopted by many users and are now in heavy use in some urban areas. Calculating using the 2017 National Household Travel Survey (NHTS) from U.S. Department of Transportation, Federal Highway Administration, nearly 10% of people across the United States used ride-sharing services in the previous month. In metropolitan statistical areas (MSA) with populations above one million, this fraction, about 15%, was much bigger. The numbers are likely to continue to grow after the survey.

The availability of ridesharing services is changing how people make trip decisions and this brings into sharp perspective the link between driving and mass transit. Are ridesharing services and mass transit substitutes or complements? The case for substitution is easy to make. Ridesharing services represent a new option to personal driving that is more expensive than traditional mass transit but less expensive than taxi cabs. Ridesharing services offer door to door service on the individual’s schedule and avoids issues involved with parking, thereby generally reducing commute times relative to mass transit. Vehicles used for ridesharing are typically more comfortable than mass transit and taxi cabs, in most places, and are more conducive to activities such as making phone calls. As such they would be expected to draw market share from both driving and mass transit.

Ridesharing services, however, might be a complement to mass transit if they are used to make a relatively short low-cost trip to and from a transit hub such as a light rail station and this tips the decision to drive or take mass transit in favor of mass transit. Other more extreme variants are also possible. A household may decide to give up one or more of its current vehicles now that ridesharing is ubiquitously available, shifting that demand to some combination of ridesharing and mass transit. The greater flexibility of ridesharing can also potentially increase the overall amount of driving, if it garners market share from trips that would have been previously undertaken by biking, walking, or carpooling. As such the widespread availability of Uber and Lyft and other similar ridesharing services has the potential to substantive change views on urban planning policies and city structures.

This paper addresses some of the same issues as the seminal paper by Hall et al. (2018), on the implications of introducing modern ridesharing services like Uber and/or Lyft. I do this using the shift observed in the last two waves (2009 and 2017) of the National Household Travel Survey (NHTS), the large scale nationally representative survey periodically undertaken by the U.S. Department of Transportation. I look at how long Uber and/or Lyft have been serving a metropolitan area influences different aspects of trip choice behavior such as mode, trip start time, and trip distance. By using the travel diary component of the 2009 and 2017 NHTS, I am able to implement a difference-in-differences framework with large nationally representative samples. I find number of daily rideshare trips (which includes traditional taxi trips in the NHTS survey) increases substantially with these increases being concentrated in both short and longer distant trips (relative to middle distant trips) and in weekday (as opposed to weekend) trips.

I also find modern ridesharing services to be a clear substitute for short haul bus trips while a complement for longer distant rail trips. In contrast, Hall et al. (2018) found the entry of

Uber results in increased bus ridership and decreased rail ridership. The main reason for the divergence between my bus and rail ridership results and those of Hall et al. (2018) is due to their study focus on transit agency level bus and rail ridership, weighting agencies equally, instead of using a random sample of individual level trip data. When Hall et al. (2018) uses a different weighting scheme to estimate the effect of Uber entry on national transit ridership, my results for bus trips and rail trips have consistent signs with Hall et al. (2018).

The ability to draw on the individual trip level diary data from the 2009 and 2017 waves of the very large nationally representative NHTS survey allows me to paint a rich, representative picture of how ridesharing services has thus far changed the nature of travel decisions in American cities. In addition to allowing me to examine how the presence of ridesharing services like Uber and Lyft influence public transit use, I am also able to examine how it impacts other trip modes including private vehicles, biking and walking., as well as the interaction of trip mode choice with distance and when the trip was taken.

3.2 Literature Review

The question of how ride sharing services would change the nature of mode choice in making trips in urban areas has received considerable media attention. Jaffe (2013), for instance, argues that ridesharing services may actually increase public transit ridership and reduce single-occupancy drivers. Walker (2018) believes that, even with ridesharing, buses will still play an important role in transporting most people in urban areas as the cost of doing so using ridesharing services is still too high. Lee (2019), on the other hand, contends that pooling services provided by ridesharing companies reduce cost for riders while providing convenient services, and have the potential to reduce bus usage. McFarland (2019) notes that if ridesharing services attract ridership away from buses and subways, congestion will result. The likelihood of ridesharing reducing congestion has also been questioned if people take rideshare trips alone and if they replace trip segments now undertaken by walking (Walker, 2018). Jaffe (2015) sees opportunities for urban agencies to collaborate with ridesharing companies to better integrate ridesharing and public transit. My paper contributes to this debate by providing empirical evidence on the effect of ridesharing on the number and nature of trips taken via public transit.

There has been limited evidence in the economics literature that examines how ridesharing services affect trip choice decisions. A major empirical paper in the literature is by Hall et al. (2018). They look at how Uber impacts public transit. To do so, they use aggregate monthly ridership data at transit agency level from 2004 to 2015, and their main analysis is based on a difference-in-differences design. Hall et al. (2018) show that for average transit agency, Uber is a complement. Specifically, they find Uber increases public transit ridership by 5 percent in the two years after Uber's entry into the market. They also show entry of Uber increases bus ridership and decreases rail ridership. Another paper by Nelson and Sadowsky (2018) shows increase in public transit use after entry of first ridesharing company in 28 major U.S. urbanized areas, also using aggregate ridership data. A very different approach was taken in a recent paper by Zhao (2019). That paper looks at the general equilibrium effects of Uber on urban area in the long run using numerical simulation approach. Zhao's results show that with a high-quality transit system, Uber enhance public transit and with low quality transit system, it reduces public transit. Parameters in Zhao's model follows existing literature or are calibrated based on Chicago's characteristics. While a simulation study, it points to the possibility of heterogeneous outcomes across urban areas, indicating that average effects potentially depend on the aggregation scheme used.

While the literature on the national impacts of ridesharing on urban trip mode choice is sparse, there have been many studies in individual locales and states, or combination of large urban areas. These papers also tend to produce a range of results that are sometimes contradictory. Babar and Burtch (2020) show that ridesharing largely decreases utilization of bus services and increases commuter rail services in U.S. urbanized areas. Alemi (2018) finds in a survey of young and middle-aged California adults that, in the absence of Uber and Lyft, about 35% of frequent rideshare users would have driven a car, a bit more than 30% would have used public transportation, and somewhat under 20% would have walked or biked. Another paper using data from San Francisco, shows 33% of people using ridesharing services would use public transit as alternative while 6% will drive on their own (Rayle et al., 2016). Smith (2016) finds that people who use ride-sharing services on a daily or weekly basis are more likely to walk, bike or use public transit than non-rideshare users. In a study sampled from seven metropolitan areas, Clewlow and Mishra (2017) find that ridesharing services reduce use of public buses by 6%, light rail by 3%, while increasing use of heavy rail by 3%. Murphy and Feigon (2016) find in seven large U.S. cities that ridesharing tends to complement public transit and substitute for vehicle trips. In another paper, this time using data from 50 largest transit agencies in the U.S., Malalgoda and Lim (2019) show an increase in ridesharing was associated with increased rail ridership in 2015, while having an insignificant influence on bus ridership. Erhardt et al. (2021) use data in San Francisco to show that ridesharing decreases the net bus ridership while insignificantly affect light rail ridership.

The work in this paper is also related to the growing literature that looks at a range of other issues related to the advent of ridesharing services. On strand of the literature estimates the consumer surplus generated by the introduction of ridesharing services, Cohen, et al. (2016), for instance, estimates that UberX service provided $2.9 billion in consumer surplus in four U.S. cities in 2015. Lam et al. (2017) shows that the magnitude of the gain in consumer surplus differs substantially across neighborhoods with different accessibility in New York City. Another strand of the literature, looks into labor market for Uber drivers. Chen, et al. (2019) show Uber drivers earn more surplus than they would in less-flexible environment. Uber drivers are attracted largely due to the flexibility, compensation, and invariant hourly earnings from hours worked (Hall and Krueger, 2018). Still another branch of the literature examines Uber's surge pricing scheme. Chen and Sheldon (2015) detail how the surge pricing mechanism helps the Uber ecosystem generate more supply of rides. Similarly, Uber's pricing mechanism allows fare changes to influence both driver utilization and passenger wait times (Hall et al., 2020). Surge pricing also allows low prices to be charged during low passenger demand hours (Castillo et al., 2017). Work comparing the ridesharing and taxi industry reveals UberX drivers have higher capacity utilization than taxi drivers (Cramer and Krueger, 2016). Uber has been shown to reduce earnings of taxi drivers (Berger et al., 2018; Brodeur and Nield, 2018). Uber also reduces drunk driving and accidents by making rides easier and less expensive to obtain (Peck, 2017; Dills and Mulholland, 2018). Finally, Uber has been shown to lead to better air quality (Sarmiento and Kim, 2021).

3.3 Conceptual Framework and Data

When ridesharing enters a market, it does not immediately grab a large number of users and achieve a high market share of trips. A potential customer first needs to download the relevant app and get familiar with how to use. The service is initially awkward because a large pool of users is needed to attract a large pool of drivers. Like many new technologies, as early adapters gain both experience with the services and enjoy favorable outcomes relative to their prior pattern of trip mode choice, more and more people start to hear about the ridesharing via word of mouth, news reports, and social media or newspaper. The user base responds by growing as does the pool of drivers, who are being motivated by some of the same information diffusion but also learning about opportunities to drive for Uber from its initial driver pool in the urban area. Overtime this process has caused ridesharing services to grow and, in turn, to have more influence on trip mode choice decisions. The import of this is that the specific date when Uber/Lyft entered a particular market can, with some important caveats, be used as variable of interest to estimate the effects of ridesharing on individual trip choices. Following Hall et al. (2018), I use a difference-in-differences framework that compares trip patterns before the first launch of ride-sharing services and trip patterns years later, considering whether and when Uber and/or Lyft entered the market.

To examine individual trip decisions, detailed information concerning those decisions including trip mode(s) choice, trip distance, the day of week and start time of the trip, and the traveler's demographic characteristics are needed. The main data sources I use are the 2009 and 2017 NHTS, which contain this information for a very large representative sample of American households. All trips made by each household member above 5 years old on a particular travel day are recorded in the diary provided as part of the NHTS surveys. Travel days are assigned over the course of a year, from March, 2008 to April, 2009 for 2009 NHTS and from April, 2016 to April, 2017 for 2017 NHTS.

For this paper, modes of taxi/rideshare, private vehicle[1], bus[2], rail[3], walk and bike are included in the analysis. Taxi, hired car (limo) and rideshare trips are recorded as a combined

[1] Private vehicle mode includes car, SUV, van, pickup truck, golf cart, motorcycle and recreation vehicles, following the NHTS definition of privately operated vehicle.
[2] Bus mode includes public or commuter bus.
[3] Rail mode includes subway, elevated rail, light rail, street car, Amtrak, and commuter rail.

mode in the 2017 NHTS. Therefore, it is important to note that whether a trip is taxi or Uber/Lyft is not differentiated in the travel day trip records. In general, taxi usage has declined with the introduction of Uber and Lyft (Brodeur and Nield, 2018). As such substitution of rideshare services for taxi and hired car services is not captured here. I restrict my analysis, to households located in 51 MSAs with over one million in population in both surveys. This is largely done to facilitate matching the household MSA codes (which are not provided for smaller places) with the date that Uber or Lyft first began providing rideshare services in the MSA.

The NHTS provides a set of individual and household demographic characteristics, including gender, age, education level and worker status of each individual, as well as the life cycle classification for each household. The main role of the 2009 NHTS is to provide trip patterns before the first launch of ridesharing, while the NHTS 2017 provides information on later trip patterns. People living in these MSAs experienced substantially different ridesharing entry times and, for two of my MSAs, neither Uber or Lyft had entered that market by the 2017 NHTS survey period.

The month and year when Uber and Lyft were firstly launched in an MSA was collected through various online sources, including Uber Blog, Lyft Blog, as well as online news websites. By collecting the entry time, ridesharing entry length faced by each individual can be calculated as number of months between the month that Uber/Lyft enters the MSA that the individual lives in and the month of travel day the individual is surveyed. If Uber or Lyft enters a city in different month, the earlier month is considered as the entry month. Entry length thus represents how long Uber/Lyft has been available to the individual. Entry length is 0 if Uber/Lyft hasn’t entered where a person lives in till the survey day. Table 3.1 shows the summary statistics for number of trips in different modes as well as entry length calculated from sample.

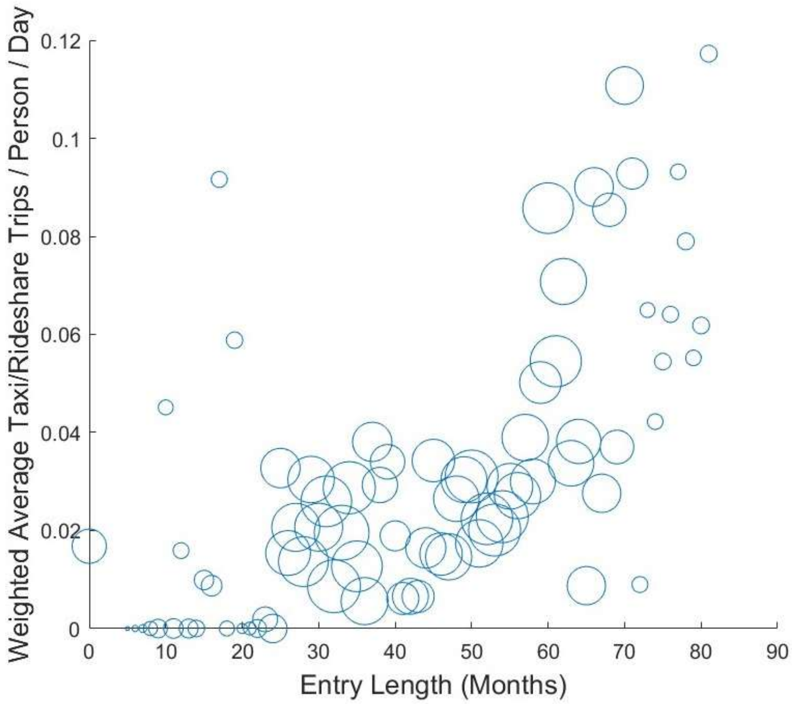
Figure 3.1 a) shows the number of taxi/rideshare trips over entry length using data in NHTS 2017. Each bubble comes from an entry length shown in the sample, the vertical axis represents the weighted average number of trips using NHTS person weights traveled on the survey day from individuals with same entry length. The bubble size represents the person weights at that entry length. It can be seen that in general, when Uber/Lyft becomes available longer to a person, the number of taxi/rideshare trips increase. Although in the travel diary data, taxi and rideshare trips cannot be separated, in the 2017 NHTS, a specific question was asked to collect data on number of rides purchased through a ridesharing app in the previous 30 days of travel day surveyed for each individual. Figure 3.1 b) shows a similar graph as Figure 3.1 a) using the number of rideshare trips in 30 days without taxi trips in the vertical axis. Figure 3.1 b) also shows that when Uber/Lyft enter longer in the market, more rideshare trips are used. These graphs support that it takes time for ridesharing services to be adopted and support the use of how long Uber/Lyft enter as a measure to estimate the impacts on trip decision.

Table 3.1: Summary Statistics for Number of Trips and Entry Length

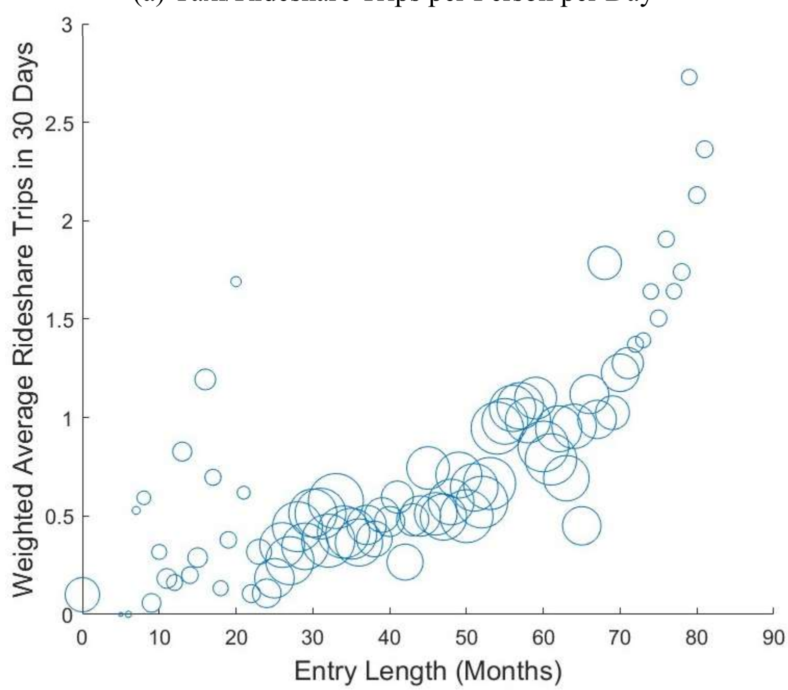
Variable	Mean	Std. Dev.	Min.	Max.
Number of Trips in Taxi/Rideshare Mode	0.013	0.165	0	8
Number of Trips in Private Vehicle Mode	3.775	2.428	0	16
Number of Trips in Bus Mode	0.039	0.299	0	7
Number of Trips in Rail Mode	0.031	0.251	0	8
Number of Trips in Walk Mode	0.433	1.071	0	16
Number of Trips in Bike Mode	0.031	0.298	0	12
Entry Length	21.087	24.932	0	81

Notes: This table shows the summary statistics for number of trips under the specified mode per person per day, as well as the summary statistics for entry length in the sample.

The control variables are collected through several different sources. The MSA population and land area each year is collected from U.S. Census Bureau. MSA level per capita income is obtained through U.S. Bureau of Economic Analysis. The MSA level demographic



(a) Taxi/Rideshare Trips per Person per Day



(b) Number of Rideshare Trips in 30 Days

Figure 3.1: Number of Trips and Entry Length from NHTS 2017

Note: This figure shows weighted average number of trips against entry length, with bubble size representing person weights at corresponding entry length, calculated based on 2017 NHTS.

controls are aggregated using data in Current Population Survey (CPS), from the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS). Gasoline prices are from U.S. Energy Information Administration and are available weekly at each Petroleum Administration for Defense Districts (PADD). The data is aggregated to monthly level and then matched to households according to the state they reside in. Various transit capacity variables are collected from National Transit Database provided by U.S. Department of Transportation, Federal Transit Administration. These data are available at transit agency level and are aggregated and matched to each MSA. Table 3.7 provides summary statistics for control variables in Appendix.

3.4 Empirical Strategy

Difference-in-differences framework is used as the main empirical strategy with following specification:

$$Y_{ict} = \alpha + \beta \text{Entry Length}_{ict} + \gamma X_{ct} + \delta Z_{ict} + \mu_c + \theta S_t + \varepsilon_{ict} \tag{3.1}$$

The dependent variable Y_{ict} is number of trips for a particular type per person per day, where i is each surveyed individual, c is each MSA, t is survey time. This is calculated by counting all trips in a particular type made by each person on the surveyed travel day. In the NHTS, a trip is defined as going from one place to another. In particular, bus and rail stations are not counted as a separate place. Therefore, in the case where people mainly use public transit mode but using walk, Uber, Lyft or other mode to connect origin or destination to the stations, it will be counted as one trip and public transit mode. Entry Length_{ict} is the main variable of interest, calculated as mentioned in the previous section. X_{ct} are MSA level control variables[4]. Z_{ict} are individual and household

[4] MSA level control variables include log(populaiton), log(population density), log(per capita income), gas price, unemployment rate, percentage population above 65 years old, percent population with bachelor degree, as well as $\mathbb{1}$ (revenue miles>0) * log(revenue miles), $\mathbb{1}$ (revenue miles>0) * fare per trip, for each public transit category. Monetary variables are calculated to be in 2008 dollars.

control variables[5]. μ_c is MSA fixed effects and S_t includes survey fixed effects, month fixed effects, and day of week fixed effects. The regression is weighted using NHTS person weights. Standard errors are clustered at the MSA level.

Variation of entry length comes from two sources. Firstly, entry length varies with Uber and Lyft entering different MSA at different time. Secondly, within an MSA, households face different entry length since they are randomly surveyed for a travel day during the survey year and lengths of time from survey day to rideshare entry are different. This creates randomness of entry length across about a year for each MSA. One concern to the identification is that the entry of Uber and Lyft not being random. In this paper, only MSAs with population above one million in both surveys are included. These are large MSAs where Uber and Lyft would want to launch in all quickly, making less of concern that may be raised if Uber/Lyft chose to enter some MSAs later when they expect there will be less growth. The change in percent of taxi drivers out of total employment in each MSA is used as a proxy for taxi growth before ridesharing to compare with entry time. Figure 3.2 in Appendix shows the change in percent taxi drivers from 2006 to 2009 and 2008 to 2009 vs. entry time. No clear pattern is shown for taxi growth and entry time. Then, percentage changes in ridership are used as proxies to examine growth in public transit before ridesharing. Figure 3.3 and Figure 3.4 in Appendix show percentage changes in bus ridership and rail ridership from 2006 to 2009 and 2008 to 2009 for each MSA against entry time. No clear relation is found either. According to Hall et al. (2018), Uber's entry decision into city is largely based on population rank and other than population, education also well predicts Uber's entry. These variables, as well as other MSA level variables that are likely to affect entry decisions of Uber and Lyft are added as control variables. For each mode, average number of trips weighted

[5] Individual and household control variables include gender, indicator of age group, indicator of having bachelor degree, indicator of being a worker and indicator of category in the life cycle classification.

by person weights for MSAs with different categories of entry time is also plotted over each survey including earlier surveys of 2001 NHTS, 1995 Nationwide Personal Transportation Survey (NPTS) and 1990 NPTS to show trends of trips over time. The plot is shown in Figure 3.5 in Appendix and trends are relatively consistent. Although earlier surveys are available, new MSA definitions were made since 2003 (U.S. Census Bureau, 2016). This makes earlier surveys with less similar MSA geographic area than later surveys. Only 2009 and 2017 NHTS are used for regression and most MSAs in the study sample provide similar geographic area.

3.5 Results

Regressions are run following Equation 3.1 in each transit mode and results are shown in Table 3.2. Each column represents a separate regression where the dependent variable is total number of trips per person per day in the mode described. In column (1), it shows that when Uber/Lyft enter the market for one more month, the number of trips in taxi or rideshare mode increases by 0.000418 per person per day in large MSAs. This can be translated to when Uber/Lyft enter the market for one more year, each person takes 0.15 more taxi or rideshare trips each month. This is relatively small in magnitude and increase is significant at 5%. For trips using private vehicles, while number of trips decrease when Uber/Lyft are available longer in the market, the decrease is insignificant. The change in number of trips in public transportation show statistically significant results, with a decrease in bus trips and increase in rail trips. As mentioned in the previous section, when people mainly use public transit to travel while using other modes such as Uber/Lyft to connect origin and destination with stations, it will only be counted as one trip as public transit mode. Therefore, the increase in rail trips may include the ones using Uber/Lyft to connect. There can also be other channels that Uber/Lyft entry can increase number of rail trips. For example, when Uber/Lyft allows for access to public transportation at one station, there can be further increase in rail trips linked to that trip. Both walk and bike trips decrease insignificantly. The coefficients showing change in number of trips are computed as percentage of mean trips weighted by person weights in the sample for each mode, shown in Table 3.8 in Appendix.

Table 3.2: Impacts of Entry Length on Number of Trips in Different Modes

Variables	(1) Taxi/ Rideshare	(2) Private Vehicles	(3) Bus	(4) Rail	(5) Walk	(6) Bike
Entry Length	0.000418**	-0.00167	-0.000969***	0.00155***	-0.00150	-0.000157
	(0.000200)	(0.00166)	(0.000328)	(0.000267)	(0.00112)	(0.000286)
N	184,421	184,421	184,421	184,421	184,421	184,421

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows coefficients of entry length following regression specification of Equation 3.1. Dependent variable for each column is number of trips per person per day under specified mode. Standard errors are clustered at the MSA level.

When Uber/Lyft are available longer in the market, there are significant impacts on public transit for both bus and rail. Regression is also run by combining the total number of trips for bus and rail as dependent variable to examine the effect on overall public transit. Results are shown in Table 3.3. When Uber/Lyft enter one more month in the market, total public transit trips per person per day increase by 0.000585 in large MSAs. However, the increase is small in magnitude and not significant. Compared to Hall et al. (2018), their results show that Uber entry increase bus ridership and decrease rail ridership. These results differ from my results mainly due to their study focus on transit agency level bus and rail ridership and they weigh transit agencies equally. When they use a different weighting scheme of pre-Uber average ridership to reflect national transit ridership, their results show that Uber entry decrease bus ridership and increase rail ridership. In this case, my results have consistent signs of bus trips and rail trips. Although when weighting by pre-Uber average ridership, Hall et al. (2018) show decrease of total public transit and my results show an increase, the effect is insignificant in both cases.

Table 3.3: Impacts of Entry Length on Public Transit

Variables	(1) Bus and Rail Total
Entry Length	0.000585 (0.000362)
N	184,421

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Note: This table shows coefficients of entry length following Equation 3.1. Dependent variable is number of total bus and rail trips per person per day. Standard errors are clustered at the MSA level.

Although in NHTS trip diary data, taxi/rideshare are recorded as a combined category of trips, 2017 NHTS also separately ask each individual for number of rideshare purchased in the previous 30 days. The 30 day rideshare trips are averaged to daily rideshare trips and used as a dependent variable for regression following Equation 3.1. This variable is set to 0 for 2009 NHTS survey. The results are shown in Table 3.4. When Uber/Lyft becomes available for one more month, it corresponds to 0.000565 more rideshare trips per person per day. This translates to 0.2 more trips per month for one more year of entry. This change is of 2.53% weighted mean of rideshare trips in NHTS 2017. The magnitude is greater than the change in taxi/rideshare trips and the coefficient is statistically significant. The greater magnitude may come from the increase in rail trips connecting to origin or destination using ridesharing services, as well as increase in rideshare trips substituting from taxi trips.

Table 3.4: Impacts of Entry Length on Average Daily Rideshare

Variables	(1) Average Daily Rideshare
Entry Length	0.000565***
	(9.60e-05)
N	184,278

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows coefficients of entry length following Equation 3.1. Dependent variable is average daily rideshare trips per person. Standard errors are clustered at the MSA level.

The number of trips is then further decomposed into different type of distances, with results shown in Table 3.5. Each coefficient in Table 3.5 is from a separate regression, with number of trips in a particular range of distances and particular mode as the dependent variable. Some coefficients are left blank since some modes are naturally shorter, thus not having a lot of trips for longer distances. From Table 3.5, it can be seen that the most significant results are in taxi/rideshare mode and public transit. The overall significant increase in taxi/rideshare trips are driven by trips less than 4 miles and trips greater than 10 miles. On the other hand, the significant decrease in bus trips is driven by shorter trips that are less than 4 miles. For rail trips, the significant increase is mainly driven by trips greater than 4 miles. For private vehicle trips, results are insignificant and number of trips decrease in all ranges except for trips from 4 to 10 miles. There is increase in trips from 4 to 10 miles, indicating there may be shift from self-driving in shorter or longer trips to median distance trips when Uber/Lyft enter longer in the market. This result is significant, but only at 10%. Most of the walk trips are within 2 miles and thus the result is very similar to overall walk trips. The effect on bike trips is small in magnitude and insignificant.

Table 3.5: Impacts of Entry Length on Number of Trips in Different Modes and Distances

Variables	(1) Taxi/ Rideshare	(2) Private Vehicles	(3) Bus	(4) Rail	(5) Walk	(6) Bike
Panel A: Number of trips < 2 miles in a particular mode						
Entry Length	0.000150**	-0.000951	-0.000441**	-5.74e-05	-0.00145	-9.87e-05
	(6.88e-05)	(0.00132)	(0.000203)	(4.28e-05)	(0.00111)	(0.000226)
Panel B: Number of trips from 2 to 4 miles in a particular mode						
Entry Length	0.000112*	-0.00119	-0.000435***	0.000100		4.50e-05
	(6.59e-05)	(0.000896)	(0.000106)	(7.62e-05)		(7.87e-05)
Panel C: Number of trips from 4 to 10 miles in a particular mode						
Entry Length	1.53e-05	0.00175*	-5.92e-05	0.000341**		
	(9.56e-05)	(0.00104)	(0.000174)	(0.000166)		
Panel D: Number of trips > 10 miles in a particular mode						
Entry Length	0.000140***	-0.00128	-3.40e-05	0.00117***		
	(5.22e-05)	(0.000921)	(0.000133)	(0.000207)		

Standard errors in parentheses

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

Note: This table shows coefficients of entry length following regression specification of Equation 3.1. Dependent variable for each column under each panel is number of trips per person per day under specified mode and specified distance category. Observations used for each regression are 184,421. Standard errors are clustered at the MSA level.

Other than decomposing trips into different distances, trips are also decomposed into different starting time. In Table 3.6, each coefficient is from a separate regression with number of trips in different starting time in a particular mode as the dependent variable. The weekday peak hours are defined as 6:00 to 10:00 AM and 4:00 to 8:00 PM. It can be seen that the significant increase of taxi/rideshare trips are mainly driven by weekday trips, especially during peak hours. There is also significant decrease in number of trips in private vehicles during weekday peak hours, while there is insignificant increase in trips during weekday non peak hours and weekends. The different effects lead to the insignificant decrease of overall private vehicle trips as seen in Table 3.2. Bus trips starting from all times decrease and the overall significant decrease is driven by weekend trips. Rail trips starting from all times increase and the result is significant during weekday peak hours as well as weekends. During weekends, people also take significantly fewer walk trips when Uber/Lyft enter market longer.

Table 3.6: Impacts of Entry Length on Number of Trips in Different Modes and Starting Time

Variables	(1) Taxi/ Rideshare	(2) Private Vehicles	(3) Bus	(4) Rail	(5) Walk	(6) Bike
Panel A: Number of trips starting in weekday peak hours by a particular mode						
Entry Length	0.000287***	-0.00493***	-0.000396	0.00150***	-0.000113	8.68e-05
	(9.77e-05)	(0.00144)	(0.000260)	(0.000274)	(0.000690)	(0.000192)
Panel B: Number of trips starting in weekday non-peak hours by a particular mode						
Entry Length	0.000199*	0.000358	-0.000307	0.000204	6.03e-05	-9.92e-05
	(0.000106)	(0.00122)	(0.000265)	(0.000127)	(0.000745)	(0.000152)
Panel C: Number of trips starting in weekend by a particular mode						
Entry Length	7.06e-05	0.00573	-0.00157***	0.00116***	-0.00495***	-0.000598
	(0.000413)	(0.00496)	(0.000427)	(0.000292)	(0.00181)	(0.000449)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows coefficients of entry length following regression specification of Equation 3.1. Dependent variable for each column under each panel is number of trips per person per day under specified mode and specified starting time category. Observations used for Panel A and Panel B are 139,359. Observations used for Panel C are 45,044. Standard errors are clustered at the MSA level.

In the main specification, variable of interest is entry length to estimate average impact of how long ridesharing has entered market affect number of trips. For taxi/rideshare mode of trips, different function forms of entry length, including adding higher order terms of entry length as well as using indicators of categories of entry length, are further explored to examine whether the rate of increase of taxi/rideshare trips changes over time. Results are shown in Table 3.9 in Appendix. According to the F-test for whether coefficients for higher order terms of entry length are 0, tests are all rejected. Then, predicted number of taxi/rideshare trips over entry length in these specifications with other variables held at mean are shown in Figure 3.6 in Appendix. The results indicate that as Uber/Lyft enter longer in the market, the increasing rate of taxi/rideshare trips also increases. These results are based on current data where longest entry length is less than seven years. How the increasing rate change with more years of entry could be studied in future research.

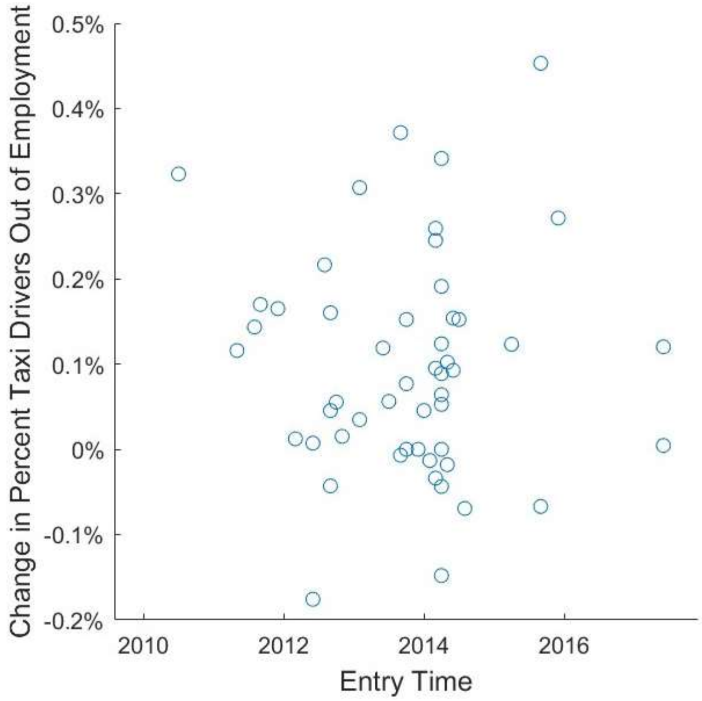
Lastly, various robustness checks are conducted. The dependent variables in this paper are number of trips and thus are non-negative discrete values. Count data regression models of Poisson and Negative Binomial are also used to examine the results. Equation 3.1 are run under Poisson regression and Negative Binomial regression. Results are shown in Table 3.10 in Appendix. Marginal effects on entry length show similar sign for most of these regressions, although magnitudes for some are smaller than linear regression. Robustness check is also conducted using year-month fixed effects instead of survey and month fixed effects. Results are shown in Table 3.11 in Appendix. Sign and magnitude of the results are similar to the main specification.

3.6 Conclusion

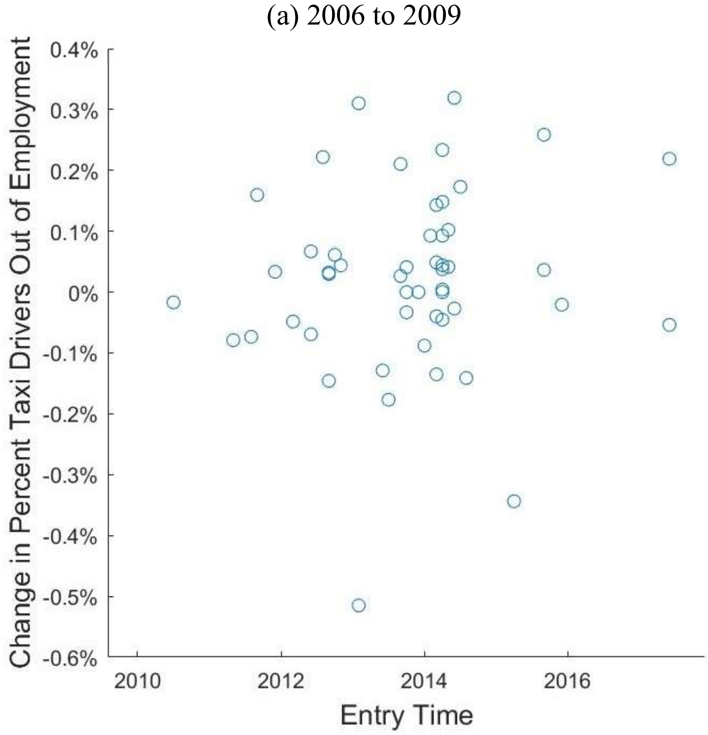
Ridesharing companies such as Uber and Lyft have been increasing in use in urban areas. Such services affect how people make trip decisions, and are linked to larger discussions of urban planning policies. This paper uses 2009 NHTS and 2017 NHTS individual travel diary data to explore how Uber/Lyft enter longer in the market affect people's trip choices. I find that when Uber/Lyft has been longer in the market, number of taxi/rideshare trips largely increase. Individuals take more taxi/rideshare trips mainly for shorter trips below 4 miles or longer trips above 10 miles. While people substitute away from short bus trips, they take more long rail trips.

When looking into the trip starting time, the increase in taxi/rideshare trips are mainly driven by trips that are starting from weekday. People significantly reduce private vehicle trips during weekday peak hours. While bus trips mostly decrease on weekends, increase in rail trips are mostly seen in weekday peak hours and weekends. Providing empirical evidence on how ridesharing services influence trip choice decisions, this paper adds knowledge to current literature looking into impacts of ridesharing on mass transit and other modes of transport, as well as provides insights for the many discussions involving urban planning policies and transportation initiatives.

3.7 Appendix



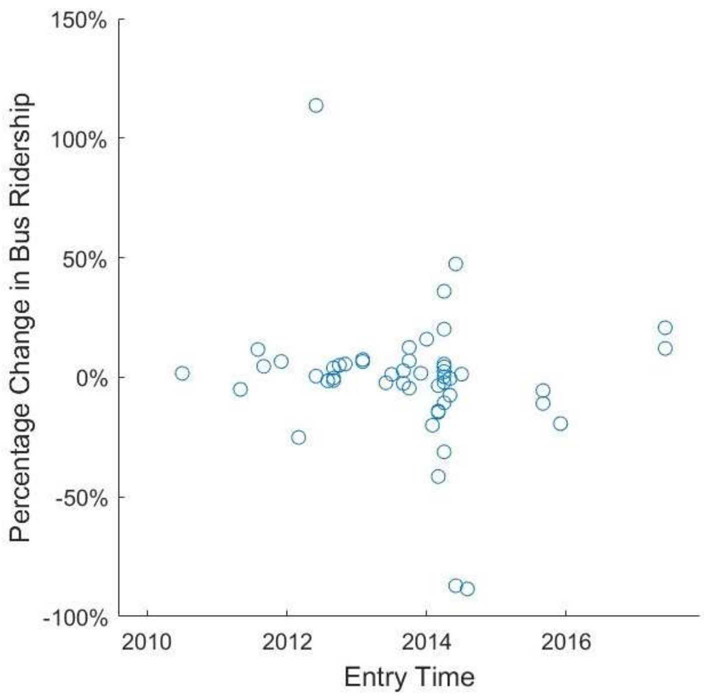
(a) 2006 to 2009



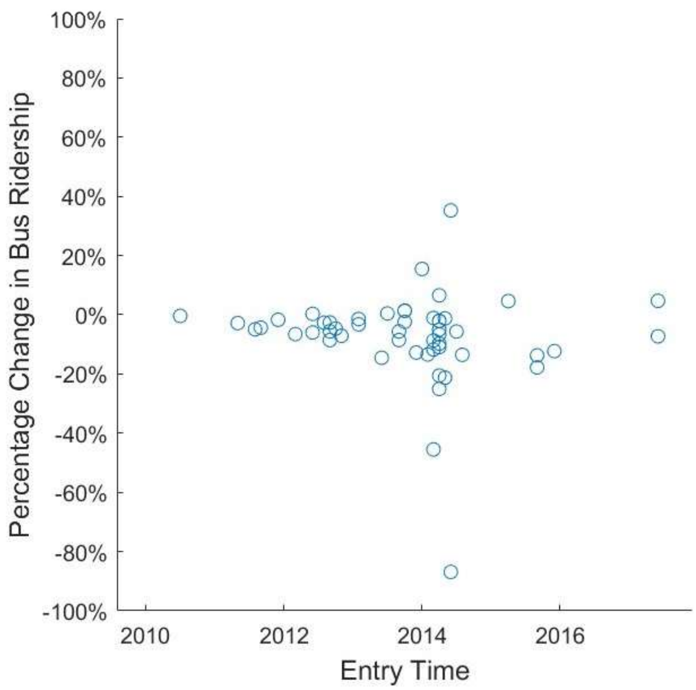
(b) 2008 to 2009

Figure 3.2: Change in Rate of Taxi Driver Out of Total Employment vs. Entry Time

Note: This figure shows change in percent of taxi drivers out of total employment from 2006 to 2009 and 2008 to 2009 for each MSA against entry time of Uber/Lyft.



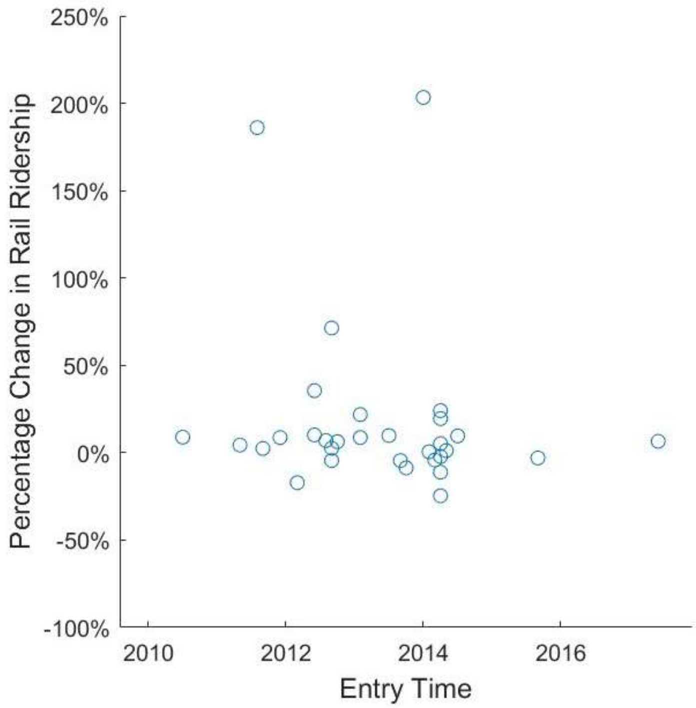
(a) 2006 to 2009



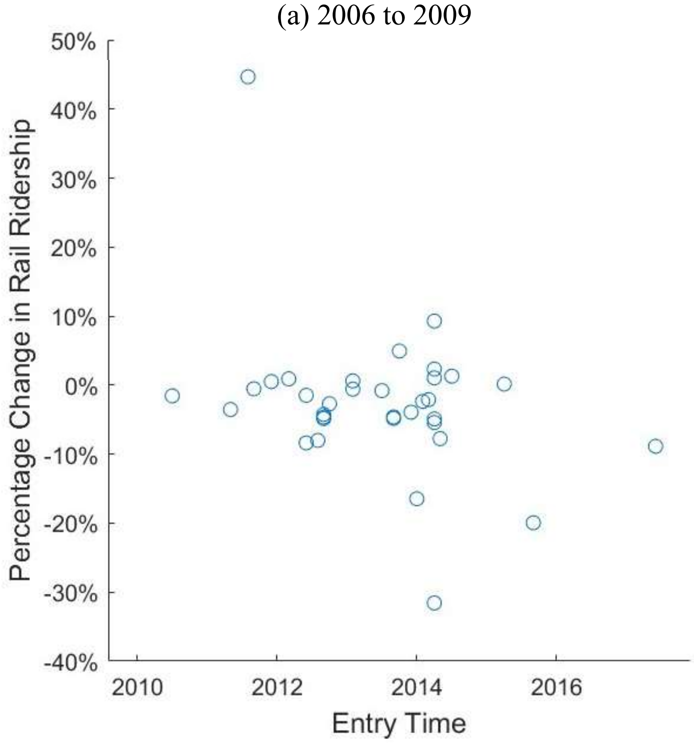
(b) 2008 to 2009

Figure 3.3: Percentage Change in Bus Ridership against Entry Time

Note: This figure shows percentage change in bus ridership (unlinked passenger trips) from 2006 to 2009 and 2008 to 2009 for each MSA against entry time of Uber/Lyft. One outlier with large percentage change is not shown in (a).



(a) 2006 to 2009



(b) 2008 to 2009

Figure 3.4: Percentage Change in Rail Ridership against Entry Time

Note: This figure shows percentage change in rail ridership (unlinked passenger trips) from 2006 to 2009 and 2008 to 2009 for each MSA against entry time of Uber/Lyft. Only MSAs with positive ridership in 2006 are included in (a) and an outlier with large percentage change is not shown in (a). Only MSAs with positive ridership in 2008 are included in (b).

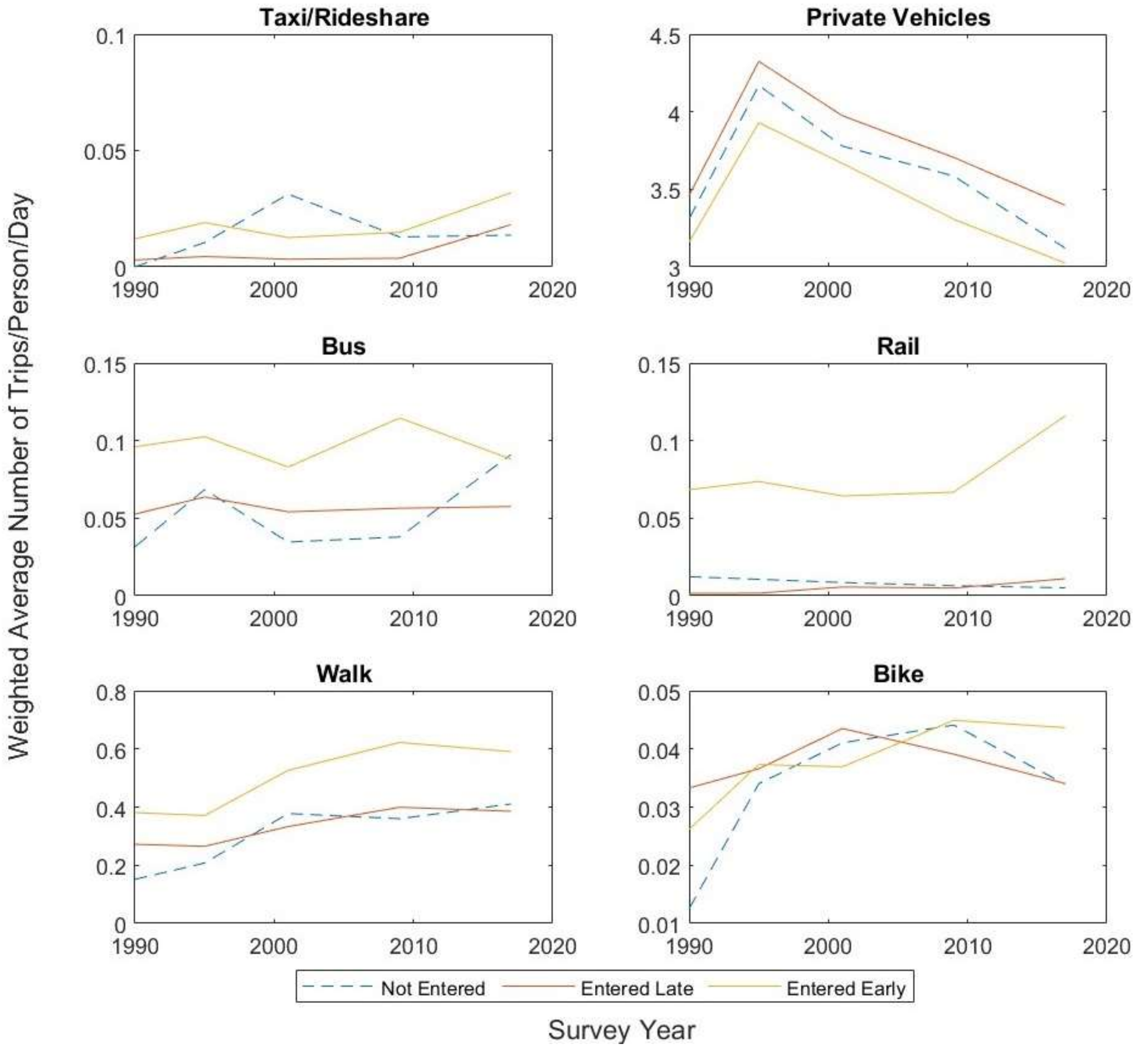


Figure 3.5: Number of Trips Over Surveys

Note: This graph represents number of trips per day weighted by NPTS or NHTS person weights over 1990, 1995, 2001, 2009, and 2017 survey for three groups of MSAs. Dashed line represents MSAs that Uber/Lyft did not enter by 2017 survey, red line represents MSAs that Uber/Lyft entered after January of 2014 and yellow line represents MSAs that Uber/Lyft entered on or before January of 2014. Only MSAs included in all or four of the surveys are included in the calculation.

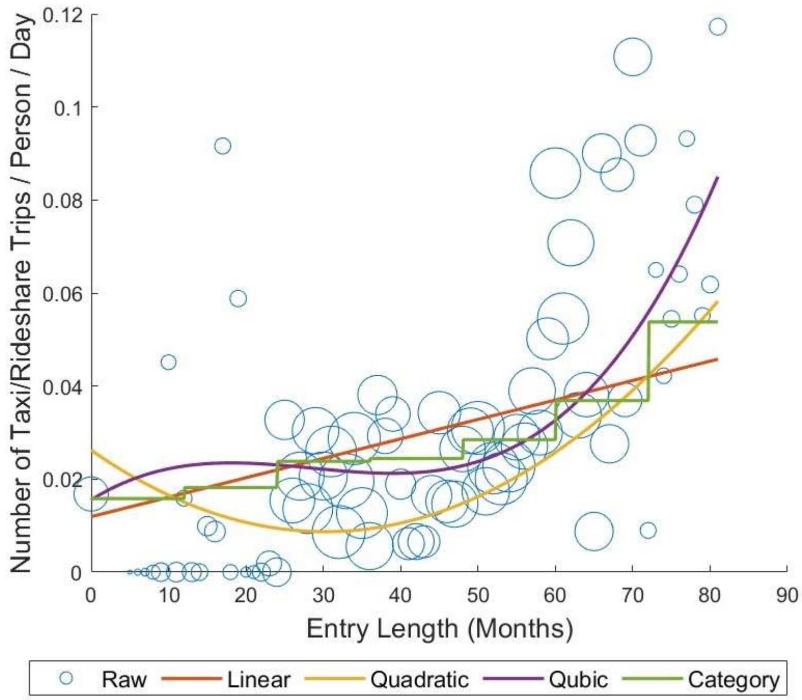


Figure 3.6: Number of Taxi/Rideshare Trips Predicted by Different Functional Forms of Entry Length

Note: The four lines in the figure show predicted number of taxi/rideshare trips from different entry length functional forms with other variables held at mean. Bubbles are weighted average number of trips per person per day from individuals with same entry length, with bubble size representing the person weights at that entry length.

Table 3.7: Summary Statistics for Control Variables

Variable	Mean	Std. Dev.	Min.	Max.
MSA Level Variables				
Population (1,000,000)	5.95	5.16	1.05	20.32
Population Density	835.10	588.98	109.34	2343.15
Per Cap Income (10,000)	4.6	0.87	2.87	8.34
Unemployment	0.06	0.02	0.01	0.14
Gas Price	2.52	0.82	1.55	4.45
Percent Population above 65	0.13	0.03	0.06	0.25
Percent Population with Bachelor Degree	0.32	0.07	0.11	0.57
Bus Revenue Miles (1,000,000)	4.96	6.48	0.04	27.35
Bus Average Trip Fare	0.84	0.22	0.33	1.43
Rail Type 1 Revenue Miles (1,000,000)	1.50	4.24	0	16.00
Rail Type 1 Rail Average Trip Fare	2.39	2.46	0	6.38
Rail Type 2 Revenue Miles (1,000,000)	3.26	7.85	0	29.40
Rail Type 2 Average Trip Fare	0.79	0.62	0	3.80
Individual and Household Level Variables				
Female	0.527	0.499	0	1
Age Group:				
5 to 20	0.036	0.186	0	1
21 to 64	0.701	0.458	0	1
over 65	0.263	0.440	0	1
Has Bachelor Degree	0.489	0.500	0	1
Is Worker	0.608	0.488	0	1
Life Cycle Classification				
one adult, no children	0.084	0.278	0	1
2+ adults, no children	0.256	0.436	0	1
one adult, youngest child 0-5	0.004	0.059	0	1
2+ adults, youngest child 0-5	0.104	0.305	0	1
one adult, youngest child 6-15	0.011	0.106	0	1
2+ adults, youngest child 6-15	0.138	0.345	0	1
one adult, youngest child 16-21	0.008	0.091	0	1
2+ adults, youngest child 16-21	0.072	0.259	0	1
one adult, retired, no children	0.064	0.245	0	1
2+ adults, retired, no children	0.258	0.438	0	1

Note: This table shows the summary statistics for MSA level control variables, as well as individual and household level control variables. Rail type 1 includes commuter rail. Rail type 2 includes monorail, light/heavy rail and streetcar. For MSA level variables, population, population density, per cap income, and average trip fare varies yearly while other variables vary monthly. For individual and household level variables, life cycle classification is household characteristic while other variables are individual characteristics.

Table 3.8: Change in Number of Trips as % of Weighted Mean

Variables	(1) Taxi/ Rideshare	(2) Private Vehicles	(3) Bus	(4) Rail	(5) Walk	(6) Bike
Change as % of weighted mean	1.90%	-0.05%	-1.17%	2.30%	-0.29%	-0.44%

Note: This table shows change as % of weighted mean, calculated using coefficients from Table 3.2 divided by average number of trips weighted by person weights under specified mode times 100%.

Table 3.9: Impacts of Entry Length on Number of Trips in Different Functional Forms

Variables	(1) Taxi/ Rideshare	(2) Taxi/ Rideshare	(3) Taxi/ Rideshare	(4) Taxi/ Rideshare
Entry Length	0.000418**	-0.00116**	0.00104	
	(0.000200)	(0.000544)	(0.00127)	
Entry Length^2		1.92e-05**	-4.21e-05	
		(7.59e-06)	(3.36e-05)	
Entry Length^3			4.92e-07*	
			(2.49e-07)	
$\mathbb{1}$(Entered < 1 yr)				0.000338
				(0.0153)
$\mathbb{1}$(Entered 1 to 2 yrs)				0.00272
				(0.0154)
$\mathbb{1}$(Entered 2 to 3 yrs)				0.00835
				(0.0135)
$\mathbb{1}$(Entered 3 to 4 yrs)				0.00895
				(0.0129)
$\mathbb{1}$(Entered 4 to 5 yrs)				0.0130
				(0.0116)
$\mathbb{1}$(Entered 5 to 6 yrs)				0.0214*
				(0.0127)
$\mathbb{1}$(Entered > 6 yrs)				0.0383***
				(0.0123)
P-Values for F-Test				
$\beta_{Entry\ Length^2}$=0		0.0145		
$\beta_{Entry\ Length^3}$=0			0.0532	
$\beta_{Entry\ Length^2}=\beta_{Entry\ Length^3}$=0			0.00200	
Observations	184,421	184,421	184,421	184,421

Standard errors in parentheses

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

Note: Column (1) shows regression following Equation 3.1 using number of taxi/rideshare trips per person per day as dependent variable. Column (2) and (3) adds quadratic and cubic entry length. Column (4) switches entry length with indicators of entry length categories. Standard errors are clustered at the MSA level.

Table 3.10: Impacts of Entry Length on Number of Trips Using Count Data Model

Variables	(1) Taxi/ Rideshare	(2) Private Vehicles	(3) Bus	(4) Rail	(5) Walk	(6) Bike
Panel A: Poisson Regression						
Marginal Effects of Entry Length	0.000103	-0.00332*	-0.00126***	0.0000930	-0.00157	-0.000109
	(0.000260)	(0.00173)	(0.000339)	(0.000507)	(0.00122)	(0.000263)
N	184,421	184,421	184,421	184,421	184,421	184,421
Panel B: Negative Binomial Regression						
Marginal Effects of Entry Length	8.64e-05	-0.00336*	-0.00168***	0.000371	-0.00199	9.43e-06
	(0.000256)	(0.00175)	(0.000425)	(0.000549)	(0.00121)	(0.000343)
N	184,421	184,421	184,421	184,421	184,421	184,421

Standard errors in parentheses

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

Note: This table shows marginal effects of entry length using Poisson and Negative Binomial Regression on specification of Equation 3.1. Dependent variable for each column is number of trips per person per day under specified mode. Standard errors are clustered at the MSA level.

Table 3.11: Impacts of Entry Length on Number of Trips in Different Modes Using Year-Month Fixed Effects

Variables	(1) Taxi/ Rideshare	(2) Private Vehicles	(3) Bus	(4) Rail	(5) Walk	(6) Bike
Entry Length	0.000341	-0.000910	-0.000968***	0.00180***	-0.00177	-0.000307
	(0.000214)	(0.00180)	(0.000341)	(0.000293)	(0.00120)	(0.000297)
N	184,421	184,421	184,421	184,421	184,421	184,421

Standard errors in parentheses

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

Note: This table shows coefficients of entry length using year-month fixed effects instead of survey and month fixed effects in Equation 3.1. Dependent variable for each column is number of trips per person per day under specified mode. Standard errors are clustered at the MSA level.

3.8 References

Alemi, F. (2018). *What makes travelers use ridehailing? Exploring the latent constructs behind the adoption and frequency of use of ridehailing services, and their impacts on the use of other travel modes*. PhD dissertation, University of California, Davis.

Babar, Y. and Burtch, G. (2020). Examining the heterogeneous impact of ride-hailing services on public transit use. *Information Systems Research*, 31(2):820–834.

Berger, T., Chen, C., and Frey, C. B. (2018). Drivers of disruption? Estimating the Uber effect. *European Economic Review*, 110:197–210.

Brodeur, A. and Nield, K. (2018). An empirical analysis of taxi, Lyft and Uber rides: Evidence from weather shocks in NYC. *Journal of Economic Behavior and Organization*, 152:1–16.

Castillo, J. C., Knoepfle, D., and Weyl, E. G. (2017). Surge pricing solves the wild goose chase. *Proceedings of the 2017 ACM Conference on Economics and Computation*, page 241–242.

Chen, M. K., Chevalier, J. A., Rossi, P. E., and Oehlsen, E. (2019). The value of flexible work: Evidence from Uber drivers. *Journal of Political Economy*, 127(6):2735–2794.

Chen, M. K. and Sheldon, M. (2015). *Dynamic pricing in a labor market: Surge pricing and flexible work on the Uber platform*. Working Paper. Anderson School of Management, University of California, Los Angeles.

Clewlow, R. R. and Mishra, G. S. (2017). *Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States.* (Research Report – UCD-ITS-RR-17-07). Institute of Transportation Studies, University of California, Davis.

Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R. (2016). Using big data to estimate consumer surplus: The case of Uber. Working Paper 22627. National Bureau of Economic Research. https://www.nber.org/papers/w22627.

Cramer, J. and Krueger, A. B. (2016). Disruptive change in the taxi business: The case of Uber. *American Economic Review: Papers & Proceedings*, 106(5):177–182.

Dills, A. K. and Mulholland, S. E. (2018). Ride-sharing, fatal crashes, and crime. *Southern Economic Journal*, 84(4):965–991.

Erhardt, G. D., Mucci, R. A., Cooper, D., Sana, B., Chen, M., and Castiglione, J. (2021). Do transportation network companies increase or decrease transit ridership? Empirical evidence from San Francisco. *Transportation*, page 1–30.

Hall, J. D., Palsson, C., and Price, J. (2018). Is Uber a substitute or complement for public transit? *Journal of Urban Economics*, 108:36–50.

Hall, J. V., Horton, J. J., and Knoepfle, D. T. (2020). *Ride-sharing markets re-equilibrate*. Working Paper. Uber Technologies.

Hall, J. V. and Krueger, A. B. (2018). An analysis of the labor market for Uber driver-partners in the United States. *ILR Review*, 71(3):705–732.

Jaffe, E. (2013). No, ride-sharing is not the death of public transportation. *Bloomberg Citylab*. https://www.bloomberg.com/news/articles/2013-07-11/no-ride-sharing-isnot-the-death-of-public-transportation.

Jaffe, E. (2015). How the microtransit movement is changing urban mobility. *Bloomberg Citylab*. https://www.bloomberg.com/news/articles/2015-04-27/how-themicrotransit-movement-is-changing-urban-mobility.

Lam, C. T., Liu, M., and Hui, X. (2021). The geography of ridesharing: A case study of New York City. *Information Economics and Policy*, 57(100941):1–18.

Lee, T. B. (2019). Lyft's new carpooling service is the beginning of the end for public buses. *Vox*. https://www.vox.com/2014/8/7/5977685/lyfts-new-ridesharing-service-is-thebeginning-of-the-end-for-public.

Malalgoda, N. and Lim, S. H. (2019). Do transportation network companies reduce public transit use in the U.S.? *Transportation Research Part A: Policy Practice*, 130:351–372.

McFarland, M. (2019). Uber wants to compete with public transit. These experts are horrified. *CNN Business*. https://www.cnn.com/2019/04/25/tech/uber-publictransportation/index.html.

Murphy, C. and Feigon, S. (2016). *Shared mobility and the transformation of public transit*. American Public Transportation Association.

Nelson, E. and Sadowsky, N. (2018). Estimating the impact of ride-hailing app company entry on public transportation use in major US urban areas. *The B.E. Journal of Economic Analysis & Policy*, 19(1):1–21.

Peck, J. L. (2017). *New York City drunk driving after Uber*. Working Paper. The Graduate Center, City University of New York.

Rayle, L., Dai, D., Chan, N., Cervero, R., and Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, 45:168–178.

Sarmiento, L. and Kim, Y. J. (2021). *The air quality effects of uber*. Working Paper. RFF-CMCC European Institute of Economics and the Environment.

Smith, A. (2016). *Shared, collaborative and on demand: The new digital economy*. Pew Research Center.

U.S. Bureau of Economic Analysis. *Regional data, GDP and personal income*. Available from: https://www.bea.gov/tools.

U.S. Census Bureau. *Population*. Available from: https://www.census.gov/topics/population.html.

U.S. Census Bureau. *Tiger/line shapefiles*. Available from: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html.

U.S. Census Bureau. (2016). *Core-Based Statistical Areas*. https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html.

U.S. Census Bureau and U.S. Bureau of Labor Statistics. *Current population survey*. https://www.census.gov/programssurveys/cps.html.

U.S. Department of Transportation, Federal Highway Administration (1990). *1990 Nationwide Personal Transportation Survey*. Available from: https://nhts.ornl.gov.

U.S. Department of Transportation, Federal Highway Administration (1995). *1995 Nationwide Personal Transportation Survey*. Available from: https://nhts.ornl.gov.

U.S. Department of Transportation, Federal Highway Administration (2001). *2001 National Household Travel Survey*. Available from: https://nhts.ornl.gov.

U.S. Department of Transportation, Federal Highway Administration (2009). *2009 National Household Travel Survey*. Available from: https://nhts.ornl.gov.

U.S. Department of Transportation, Federal Highway Administration (2017). *2017 National Household Travel Survey*. Available from: https://nhts.ornl.gov.

U.S. Department of Transportation, Federal Transit Administration. *National transit database data reports*. Available from: https://www.transit.dot.gov/ntd/ntd-data.

U.S. Energy Information Administration. *Weekly U.S. regular all formulations retail gasoline prices*. Available from: https://www.eia.gov/petroleum/gasdiesel/.

Walker, J. (2018). The bus is still best. *The Atlantic*. https://www.theatlantic.com/technology/archive/2018/10/bus-best-public-transitcities/574399/.

Zhao, W. (2019). *The long run effects of Uber on public transit, congestion, sprawl, and the environment*. Working Paper. Department of Economics, University of Louisville.