2019

The Effects of Environmental Factors on Maternal Health, Infant Health and Mental Health

Li Zeng
Lehigh University, lizeng212@gmail.com

The Effects of Environmental Factors on Maternal Health, Infant Health and Mental Health

by

Li Zeng

A Dissertation

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Business and Economics

Lehigh University

August 30, 2018

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Li Zeng
The Effects of Environmental Factors on Maternal Health, Infant Health and Mental Health

Defense Date

Committee Members:

Shin-Yi Chou (Committee Co-Chair)

Muzhe Yang (Committee Co-Chair)

Seth Richards-Shubik

Cheng Chen

ACKNOWLEDGMENT

I feel so lucky to have many wonderful people in my life who have supported me to pursue my dream. My life at Lehigh University would not have been the same without the support of my advisors, my family, my teachers and my friends.

To my parents, my father Lingcai Zeng and my mother Shaofen Lai. Thank you for always believing in me and being there for me, with full of love and support. Your love and support are the source of my strength to face and overcome all the obstacles, and to keep chasing my dream. I want to especially thank my Dad, for taking care of my Mom without my help all these years to let me pursue my dream thousand miles away. I also want to especially thank my Mom, for never giving up fighting for being there for us.

I would like to express my sincere gratitude to my dissertation advisor and committee co-chair Dr. Shin-Yi Chou, who is the role model I always look up to. Among all those wonderful things I have ever experienced, having her as my advisor is one of the luckiest. She has encouraged, inspired and helped me to conquer the challenges in my research endeavors and my life. For me, she is like the brightest sunshine, brightening my days with all those warm smiles, cheerful encouragements and invaluable suggestions. There are countless times that I stepped into her office with upsets and concerns, and walked out with smiles and confidence. She has taught me how to be a good researcher, and how to become a better person. She is and will always be my advisor in my heart.

I would like to sincerely thank Dr. Muzhe Yang. It is my great fortune to have him as my committee co-chair. He is always patient and always there for me whenever I need suggestions. He has showed me how to be a passionate, organized and disciplined researcher, like him. I have learned a lot precious experience from him, about being a qualified researcher, being strong facing life challenges, and being a better myself. He has inspired me with his research attitude, suggestions, and life experience.

I am thankful to my committee members Dr. Richards-Shubik and Dr. Cheng Chen, for sharing their suggestions and experience with me, for taking time to talk with me, mentor me and read through all the drafts of my work. I also thank Dr. Thomas Hyclak and Dr. Irina Panovska. Working as teaching assistant for them provides me the opportunity to gain teaching skills and to know my passion in teaching economic courses. Thanks to all the professors who have taught me economic courses during the PhD program, for teaching me the knowledge and the method to explore more in economics, and for always being willing to help whenever I have questions.

I want to thank all my friends, especially Qiuming Wang, Lu Chen and Xiang Yan, for always taking time to talk with me whenever I need them, for reaching out to me and encouraging me during those bad days, for being on the other side of phone and listening, and the most important, for being my friends.

I also want to thank all my peers in the Ph.D. program. We have spent so much time together in the room 209 and 210, writing homework, finishing course projects, working on research papers and preparing for teaching. We shared not only study and research materials but also feelings and experience to support each other. I'm glad you are part of my Ph.D. life.

Last but not least, I want to thank our Ph.D. program in economics at Lehigh University. Thank you for the financial support and for the opportunity to let me become part of the program. My experience during the program is a remarkable experience in my life that I cherish a lot, and it makes me stronger.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation explores the effects of environmental factors on adult's health and infant's health at birth.

The first chapter of this dissertation studies the impact of maternal stress triggered by wildfire on infant birth outcomes. As a common natural event, one of the most noticeable effects of wildfires is the stress caused by their threat to people's life and property. This study estimates the impact of exposure to wildfire events during pregnancy, especially the effects of maternal stress triggered by wildfire outbreaks, on infant birth outcomes. By linking three data sources – birth records from the New Jersey State Department of Health (2004-2012), wildfire events data from the New Jersey Hazard Mitigation Plan (2003-2012), and air pollution data from the Environmental Protection Agency (EPA, 2003-2012) – I am able to disentangle the impact of maternal stress triggered by wildfire outbreaks from the impact of air pollution caused by wildfire smoke on birth outcomes, especially birth weight. The results suggest that being exposed to a significant wildfire event would reduce an infant's birth weight by approximately 39 grams on average. That effect persists after ruling out the possible impacts of air pollution from wildfire smoke. When I estimate the impact of maternal stress triggered by wildfire outbreaks at different pregnancy stages, I find that the adverse effects of prenatal exposure to wildfires are more significant and more dangerous at the early stage (especially the first trimester) of pregnancy.

The second chapter of this dissertation investigates how the access to restaurants (fast-food and full-service restaurants) affect the probability of mothers gaining excessive maternal weight during pregnancy, and how would excessive maternal weight gain affect infant's birth outcomes. Using the Linked Patient Discharge Data and Birth Cohort File (2007-2010) together with the County Business Patterns Data, I first estimate the effects of access to restaurants (both fast-food

and full-service restaurants) on infant health at birth. Based on the adverse effects of fast-food restaurants on infant birth outcomes I have observed, I further test whether excessive maternal weight gain is the channel, through which access to fast-food restaurants cause the adverse effects on infant health. The estimation results confirm that the ease of restaurant availability (measured by the number of fast-food restaurants in residential areas) is the factor that causes mothers to gain excessive maternal weight gain and therefore cause the negative effects on infant health (i.e. C-section rate, complications at delivery, and Apgar scores). In addition, I also find that increasing the establishment of full-service restaurants and stores that provide more options of healthier food might help to mitigate the effects cause by increasing fast-food restaurants.

In the third chapter, I use the data collected from tweeter accounts to estimate how air pollution could affect people's emotion. Substantial economic growth is the challenge every economy is facing, part of which requires keeping economic development without scarifying environment. Air pollution, as a side effect caused by fast economic development, is the problem waiting to be solved not only in developing countries but also in developed countries. It is widely approved that air pollution can cause negative impact on physical health in both epidemiology and economics studies. However, the effects of air pollution on mental health are much less studied in economics for the causation. Using the data collected from tweets posted through tweeter accounts, I use a relatively large and representative sample in our analysis. Applying the Linguistic Inquiry and Word Count (LIWC) method. We have constructed an emotion score based on the contains of each tweet, which can reflect people's mental health status. The estimation looks at the effects of different pollutants on people's emotion (both positive and negative emotions). We have found that increased concentration level of sulfur dioxide or particulate matters can cause significant adverse effect on positive emotion and the polarity value (i.e. the spread between positive and negative emotion scores).

CHAPTER 1:

Wildfires and Infant Health at Birth: Evidence from New Jersey

1.1. Introduction

Much has been written about the importance of health at birth and its strong link to long-term outcomes, including long-term health, educational attainment, and income in adulthood[1]. Many factors can affect birth outcomes through various mechanisms by triggering the fetal programming process. Among those, maternal stress has been suggested as one of the most important mechanisms through which mother's emotional or psychological condition during pregnancy could affect fetus' development in utero and further infant health at birth. However, this emotional or psychological condition cannot be quantified easily.

Cortisol level in humans is a relatively accurate measure of stress based on the evidence collected from different clinical surveys[2], and have been proposed as a primary factor that could affect fetal programming, a critical process through which a stimulus or insult during the fetus's vulnerable development process may cause a permanent effect on fetal development *in utero* (Davis and Sandman, 2010), leading to the changes in birth outcomes. Given the fact that cortisol levels higher than the normal average usually have been observed among people who are experiencing greater stress (Wust et al., 2000), a convincing hypothesis has proposed that maternal stress can increase the maternal cortisol level, which will further negatively affect fetal development through neuroendocrine system, immune function, and behavioral channels (Schetter, 2011).

[1] Currie (2011); Almond and Currie (2010, 2011); Royer (2009); Black, Devereux, and Salvanes (2007).

[2] Simmons and etc. (1984); Kirschbaum and etc. (1995); de Quervain, Roozendaal and McGaugh (1998); Dickerson and Kemeny (2004).

Studies using clinical data collected from mothers and infants have examined the effect of increased cortisol levels on infant health at birth and found strong link between higher cortisol levels and worse birth outcomes. However, broader applications of these studies are limited by the accessibility of data and the small sample size problem. Rather than using actual cortisol level as the relatively accurate measure of stress, another group of studies exploits certain types of natural events as exogenous shocks (for example, famine, earthquake and terrorist attack) to investigate the effects of maternal stress triggered by such shocks and comparing birth outcomes of the affected group to birth outcomes of the group that is not affected. And this method can be applied broadly to investigate the effects of maternal stress.

This paper aims to investigate the impact of maternal stress on infant birth outcomes using a common nature event -a wildfire outbreak- as the natural experiment, which generates a source of unpredictable and unexpected stress during pregnancy and allows for the reasonable casual inference. We use a confidential version of the birth records from the New Jersey State Department of Health for 2004 to 2012 to estimate the impacts of prenatal exposure to wildfire outbreaks on infant health at birth.

Unlike other natural disasters, such as earthquake and hurricane, wildfire can be caused by nature or human beings, and can spread easily through forests. Therefore, wildfire can become a threat to any forested area in the United States, and using wildfire as the natural experiment will help reduce the analytical problem in our estimation, which might be caused by selection through migration behaviors (it will be difficult for people to avoid wildfires by choosing the residential area because of the general existence of wildlands and the random occurrence of wildfire). As the most densely populated state in the United States, New Jersey continues to grow, creating the land use pressures and leading to more and more people move from urban areas to rural wildland areas.

The increasing number of residents living in the rural wildland areas not only raises the potential population that might be affected by wildfire, but also increases the probability of wildfire occurrences that are caused by human. Moreover, Pinelands and Pine Barrens National Reserves occupy nearly 22% of New Jersey's land area, representing two of the most hazardous wildland fuel types because they burn extremely hot and spread fire rapidly[3]. Our study uses all the major wildfire events (i.e. wildfires burning more than 150 acres or considered significant by the New Jersey Office of Emergency Management) during the years 2003 to 2012 to evaluate the effects of prenatal wildfire exposure on infant birth outcomes. Figure 1-1 shows the distribution of forest and pineland areas in New Jersey, and Figure 1-2 is a historical wildfire occurrence map from 1924 to 2007. As we have observed in Figure 1-2 and summarized from the detailed wildfire records, most of the wildfires in New Jersey are not severe (i.e. burned relatively smaller areas, and lasted for a relatively short period), which makes it worthwhile to generalize our results nationwide, and to apply our analysis to states with more prevalent and significant wildfires in the future.

Wildfire exposure could affect birth outcomes in two ways, by polluting the air and causing maternal stress. In this paper, we control the most impactful pollutants regulated by the Environmental Protection Agency (EPA) in the estimation, which helps us to rule out the possibility that wildfire exposure affecting birth outcomes by releasing pollutants into the air and therefore increasing the probability of being exposed to air pollution *in utero*. By controlling pollution levels in estimation, we are able to isolate the effects of wildfire exposure that work through maternal stress. Therefore, we believe that our estimation at the final stage can capture the effects of maternal stress itself during pregnancy.

[3] State of New Jersey 2014 Hazard Mitigation Plan (2014).

Our results indicate that being exposed to a wildfire event could have a significant adverse effect on birth weight, and thus might increase the risk of having low birth weight. Our results further suggest that such a negative effect is more powerful at the relatively earlier stages of pregnancy. The estimated effect of ever being exposed to any wildfires *in utero* decreases average birth weight by 20 grams. For all births with more than 26 weeks gestational age, the birth weight decrease caused by maternal stress is approximately 39 grams after controlling for the air pollution, which is also a consequence of wildfire outbreaks. By estimating the effect of wildfire exposure in each month of pregnancy based on the full-term birth sample, we find that the early stages (especially the first trimester) of pregnancy are much more fragile to maternal stress, compared to the later stages.

The rest of the paper is organized as follows. Section 2 provides background information on the relationship between wildfire, air pollution, and stress, and reviews the existing literature. Section 3 introduces our data sources. In section 4, we discuss our empirical strategy in detail. Section 5 presents our main results, and section 6 presents the results of robustness checks. Section 7 concludes.

1.2. Background and Literature

Two of the most noticeable adverse effects of wildfire are its emission of smoke and its threat to people's safety and property. Wildfire smoke primarily contains carbon dioxide, particulate matter, water vapor, nitrogen oxides and other compounds, some of which can travel a very long distance and thus affect a broad area and potentially a large population. In addition to its effects on air quality, wildfire if not quickly controlled or occurring in wildlands adjacent to residential areas can threaten property or the lives of those who live in or near wildlands. While significant literature has studied the effect of air pollution, and studies have focused on the effect of stress caused by events similar to wildfire, research using events that involve both air pollution and stress is unique. This paper uses wildfire as a special type of naturally occurring stressful event and aims to disentangle the effect of maternal stress triggered by wildfire from the impact of air pollution caused by wildfire smoke.

In general, two strands of the literature are related to our study: one focuses on the effect of air pollution on adults' and children's health; the other investigates the impact of maternal stress on birth outcomes. Our study is at the conjunction of these two strands, because life/property-threatening wildfire also might release pollutants into the air.

In epidemiology, convincing evidence across different studies related to the adverse effect of air pollution (especially particulate matters and carbon monoxide) suggests that it might increase the morbidity of respiratory symptoms for adults and children (Bowman and Johnston, 2005), and may have an adverse effect on infant birthweight (Breton et al., 2011; Glinianaia, 2004). In health economics, many studies investigate the effect of air pollution, particularly on infant birth outcomes, and find there are adverse effects on infant birth weight and mortality (Currie and Neidell, 2005; Chay and Greenstone, 2003). This converges with the evidence from epidemiology. Wildfire smoke may lead to serious air pollution and can dramatically increase the level of particulate matters and carbon monoxide, possibly resulting in impaired fetal growth due to hypoxia and or oxidative stress (Siddiqui et al., 2008). The literature on the impact of wildfire smoke already has found a negative effect on birth weight (Holstius et al., 2012; Siddiqui et al., 2008). These studies together suggest that ruling out the potential effects of air pollution is crucial if we want to disentangle the impact of maternal stress triggered by the wildfire event from the effect of air pollution caused by wildfire smoke.

We know that maternal signals of stress during pregnancy may have a programming influence on the developing fetus, causing potential adverse consequences as gestation advances (Davis and Sandman, 2010). As the end product of the body's major stress responsive system, cortisol has been observed in higher levels among people who are exposed to greater stress (Wust et al., 2000; Aizer et al., 2009). Using a unique dataset with measures of cortisol levels during pregnancy in maternal fixed-effect models, Aizer, Stroud, and Buka (2009) strengthen the evidence that maternal stress will increase excessive cortisol levels, possibly with adverse effects on child's cognition, health, and educational attainment in later stages of life. Although the effect is not large, they also observe slightly worse birth outcomes linked to excessive cortisol levels. A similar study conducted by Davis and Sandman (2010) uses the data on maternal cortisol measures together with data on the maternal psychological state: they find that elevated maternal cortisol levels caused by pregnancy-specific anxiety have programming influences on the developing fetus, and are associated with a slower rate of development over the first postnatal year and lower scores on the mental development index of the Bayley Scales of Infant Development (BSID) at 12 months.

However, such unique datasets with cortisol measures during pregnancy are not widely accessible, and typically only involve a small group of people. To evaluate the impact of maternal stress with large scale and easily accessible data, some research examines the impact of maternal stress triggered by negative exogenous shocks (i.e. exogenous stressful events), such as life events and natural disasters, which serve as natural experiments. For example, many researchers have looked at the impact of maternal stress triggered by terrorist attacks (such as the 9/11 attack). Most of these studies compare baby's birth outcomes between treated and control groups identified by their mothers' residence area, assuming that mothers who live far from the attacked area would experience less maternal stress than mothers who live in or near the attacked area. In addition, there are also studies compare birth outcomes of infants born to mothers who live in the same area but were pregnant before and after terrorist attacks to investigate the impact of maternal stress. Some of these studies are summarized in Table 1-1.

Another group of studies examines the adverse effect of catastrophic natural disasters (such as earthquakes) on infant birth outcomes. Like the literature on the effect of terrorist attacks, this work compares the birth outcomes of mothers living in affected versus unaffected regions, or compares birth outcomes of infants born to mothers who were pregnant before and after natural disasters. Currie (2013) uses a large-scale individual-level dataset from Texas an follows the same mothers over time to investigate the impact of storms and hurricanes on infant birth outcome, including not only birth weight but also abnormal conditions and complications at delivery and labor. She finds that exposure to a hurricane during pregnancy increases the probability of abnormal conditions for the newborn and meconium aspiration syndrome (MAS). Table 1-1 also summarizes some of the literature on the effect of maternal stress caused by natural disasters.

In summary, many existing literature investigates the causal effect of maternal stress on a variety of birth outcomes, and find different adverse effects on birth outcomes as pregnancy advances. Some of the literature focuses on birth weight, and others focus on birth outcomes such as abnormal conditions of newborns, complications at delivery and Apgar score at birth. However, the accuracy of these estimations might be affected by potential problems such as small sample size, precisely identifying exposure to stressful event and migration behavior in response to the stressful events. Our study focused on estimating the effect of maternal stress caused by wildfire events, by using a large sample from New Jersey birth records and by ruling out the other possible channel through which wildfire outbreaks could affect birth outcomes. By using the birth records data, we are able to estimate the effects of maternal stress triggered by wildfire outbreaks based on a large-scale data set with sufficient variation at individual level. As we can observe from Table 1-1, there are various studies using different types of exogenous shock as the nature experiment to investigate the effects of maternal stress, however, almost none of them has the detailed location information of multiple nature events to precisely identify the exposure to stressful events. With the uniqueness of this confidential data set and information of wildfires, we are able to use detailed residential address to precisely identify the exposure to wildfire by matching the location of each wildfire with mother's residential address information (i.e. mother's detailed address information, as well as the latitude and longitude of the address).

Although wildfire is much more prevalent in the Southwest and Mideast, it affects all forest areas across all cities, counties and states in the United States. For most of the wildfire, they are not as catastrophic as other natural disasters, such as earthquake and hurricane, which provide an unexpected and unpredictable source of mild or moderate stress, allowing us to investigate the effects of maternal stress caused by a mild or moderate stressful natural event. As we mentioned in the previous section, clinical data of cortisol level is very difficult to collect, therefore comprehensive evidence collected by different studies using various types of natural events to estimate the impact of maternal stress is needed to support the clinical study. Our study attempts to contribute to the existing literature by adding the evidence of estimated effects of maternal stress that could be caused by a type of mild or moderate stressful natural events.

1.3. Data

1.3.1. Birth Records

One of the most important datasets that we use in this study is the restricted version of birth records from the New Jersey State Department of Health. It contains all the information (except for very confidential information such as mother's name) collected from birth certificates during the period January 2004 through December 2012. This dataset has all the information we need to identify wildfire exposure (such as date of birth, mother's last normal menses date, and gestational age), as well as a variety of birth outcomes (such as birth weight, abnormal conditions and complications at labor/delivery) and other important variables that we must control (such as mother's demographics, parents' education, medical risk for this pregnancy, maternal behaviors, and prenatal care).

We use the information on newborns' date of birth, gestational age, and mother's last menses date (LMP), together with information on dates of wildfire occurrences, to determine whether the infant has been exposed to any wildfire event *in utero*. We measure exposure to any wildfire in our main estimation as follows: First we create a date variable expressed in year and the month of the year for each month[4] of every pregnancy by counting backward from the date of birth according to the length of gestational age; then we compare the date of each pregnancy month with the date of each wildfire to check if this baby was *in utero* when there is any wildfire happened during that month. As the third step, we compare mother's residential zip code versus the zip code of every wildfire outbreak to confirm whether this baby's mother lived in the affected area during any month of her pregnancy. If the baby was *in utero* when any of the wildfire happened and his/her mother also lived in the affected area when this wildfire occurred, then we define this baby

[4] For example, if the gestation length of one pregnancy is 9 months, each month of this pregnancy refers to the 1st to 9th month during this pregnancy.

as exposed to that particular wildfire during that month of pregnancy. Our indicator variable measures exposure to any wildfire for each month during the pregnancy. It equals 1 if the baby was exposed to any wildfire *in utero* in that month during pregnancy. For example, for an infant with a 9-month gestational length, if there was a wildfire outbreak in his/her mother's residential region during the mother's third month of pregnancy, then the indicator generated for the third month of pregnancy equals 1. To check the accuracy of our approach used to identify wildfire exposure *in utero* and verify the robustness of our estimation, we have also used the approximate date of conception, counting forward to calculate the year-month date of each month during the pregnancy, as an alternative method to define the exposure to wildfire *in utero*.

1.3.2. Wildfire Events

As an essential part of our estimation, we need information about every significant wildfire event[5] from 2003-2012 to identify whether an infant born in 2004-2012 has ever been exposed to any significant wildfire event. All of the pollution data are one year earlier than birth records data, which allows us to capture prenatal exposure to wildfire outbreaks for those who were born in 2004 but conceived in 2003. The historical wildfire events we use to identify the affected group in our paper come from two main sources: historical New Jersey State Hazard Mitigation Plan and Federal Wildland Fire Occurrence Data.

The New Jersey State Hazard Mitigation Plan provides information on significant wildfires according to the New Jersey Forest Fire Service (NJFFS). This includes county of occurrence, a detailed description of each wildfire and the total acres burned. We can also identify start date,

[5] Significant wildfire event is defined as wildfire burning a total of greater than 100 acres or considered significant wildfires by New Jersey State Hazard Mitigation Plan.

control date, and location (such as street, area, township or city) of each wildfire from the description for each wildfire.

The Federal Wildland Fire Occurrence Data provides wildfire records collected by federal land management agencies in the United States, including information on start date, control date, location (i.e. latitude, longitude, and state), fire type, cause, size class, and so on. However, this dataset only collects information of wildfires that happened in the federal owned wildlands. Therefore, we use this dataset as the supplemental resource to help us check the location information of the wildfires that are in both datasets.

We collect the location information of each wildfire from four other supplemental sources: 1) Google map; 2) USPS ZIP Code Lookup website; 3) ZIP code boundary map; and 4) Google news. Since the New Jersey State Hazard Mitigation Plan provides only limited information about the location of each wildfire occurrence (such as started street, district, township or city), we need extra information including a ZIP code list for every township or city, a ZIP code boundary map showing the boundary of each ZIP code, and published news related to each wildfire outbreak. To accurately pin down the ZIP code of the affected area, we use the description from the New Jersey State Hazard Mitigation Plan as the primary source of information and Google news as the secondary source for information about the affected areas. Using the wildfire name, occurrence time and location information from the New Jersey State Hazard Mitigation Plan, we were able to use Google searching news related to each wildfire to collect information of affected area as detailed as possible, in terms of affected or closed street, burned area and evacuated community. With all the information we have collected, we further use Google Map together with the USPS ZIP Code Lookup website and the ZIP code boundary map, to pin down the zip codes for all the affected regions.

Table 1-2 summarizes the number of wildfires by county from 2003 to 2012. It shows that most wildfires happened in counties with large areas of forest. Ocean county and Burlington county occupy most of New Jersey's pine lands and pine barrens, which also have the most frequent wildfire occurrences. In our analysis, we only use the information on significant wildfires, which burned at least 150 acres and are evaluated as dangerous or have caused significant damages by the department of New Jersey Forest Fire Service.

1.3.3. Pollution Data

Our data on pollution comes from U.S. Environmental Protection Agency[6] (EPA), which provides publicly available data on the regulated pollutants that are considered harmful to public health and the environment, at hourly and daily levels. According to National Ambient Air Quality Standards[7] (NAAQS), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and ozone must be monitored at the one-hour level[8], and particulate matter at the 24-hour level. Among these five pollutants, CO and ozone are measured in parts per million (ppm), SO_2 and NO_2 are measured in parts per billion (ppb), and PM2.5 is measured in micrograms per cubic meter based on the local condition[9] (LC) using the Federal Reference Methods[10]. In order to rule out the pathway through which wildfire exposure affects birth outcomes by polluting the air, we control these five pollutants at average monthly concentration levels. The monthly average concentrations of these five pollutants (CO, SO_2, NO_2, ozone and PM2.5) at ZIP code level are the arithmetic means of the weighted monthly averages at monitor-level within each ZIP code.

[6] Data source: http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily

[7] Please check http://www3.epa.gov/ttn/naaqs/criteria.html for details.

[8] Primary standards provide public health protection, including protecting the health of "sensitive" populations such as asthmatics, children, and the elderly.

[9] The concentration was reported based on local temperature and pressure.

[10] For PM2.5 at local conditions, only those data validated from Federal Reference Methods, Federal Equivalent Methods, or other methods that are to be used in making NAAQS decisions are reported.

We construct the variable measuring pollution for each pollutant at ZIP code level for each month in every year from 2003 to 2012 as follows: First, the pollutant data contains hourly readings (i.e. CO, SO_2, NO_2, and ozone), we compute the arithmetic mean of those readings within a day to get the daily average for every pollutant at monitor level. Second, we compute the arithmetic mean of all the daily measures within a month to get the monthly average for every pollutant (i.e. CO, SO_2, NO_2, ozone and PM2.5) at monitor level. Third, we pair each ZIP code with all the monitors in this ZIP code and surrounding states, and then calculate the geodetic distance from ZIP code centroid[11] to the location of each paired monitor[12]. Fourth, we keep only the monitors within 20 miles from ZIP code centroid for each ZIP code, based on the distances we have calculated in step 3. Finally, we calculate the weighted monthly average of concentrations for each pollutant at ZIP code level by weighting monthly average readings from the selected monitors using the inverse distance between ZIP code centroid and monitor as the weights.

Based on the average monthly concentration levels for different pollutants we have calculated above, we then construct average concentration level of different pollutants for every pregnancy month (i.e. 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th and 9th month during pregnancy) for each individual in our sample. Mothers last normal menses date and baby's date of birth are used to assign and compute the concentration level of pollutants in different months of pregnancy for each individual. For every single individual in the sample, we treat the month of birthdate as the last month of pregnancy, and count backward towards mother's last menses date to pin down the month of the year (i.e. January, February and so on) for each pregnancy month. In order to correctly assign measures of pollution, we then match concentration levels of pollutants to each month of pregnancy according to in which month of the year that each pregnancy month is. For alternative

[11] Zip code centroid data is purchased from http://www.zip-codes.com/.
[12] The location of each monitor is identified by its coordinates.

model specifications, we further use the average measures of pollution in each month of pregnancy to construct the average concentration level of pollutants for each trimester of pregnancy. As a robustness check, we have also constructed the measures of pollution by using an alternative method, which uses the estimated conception date counting forward towards baby's birthdate and calculate the concentration level for each pollutant by using the estimated start date and end date of each pregnancy month. The details of the method will be further discussed in the later section.

1.3.4. Sample Construction

We use two main samples for our estimation. The first contains all the observations with gestational length greater than 26 weeks; the second includes infants who are full-term (i.e. with at least 37-week gestational length). Based on these two samples, we further restrict the observations according to these three conditions: First, we retain singleton live births with non-missing information on variables controlled in the regression; second, we drop the observations with abnormally low or high birth weight, so the remaining observations have birth weight ranging from 500g to 6500g; third, we only keep the births of mothers aged 20 to 45; in addition, we exclude the birth records of mothers who do not reside in New Jersey. We initially have 148,167 observations in total, and we have 87,862 birth records left after the restrictions.

1.4. Empirical Method

1.4.1. Identifying Exposure to Wildfire

To be as accurate as possible when identifying prenatal exposure to any wildfire event, we need mother's residential information during pregnancy (including residence address, ZIP code and coordinates) and mother's last menstrual period (LMP) and baby's date of birth[13]. We also need each wildfire event's starting date, controlled date, starting location, and affected areas identified by ZIP code.

Based on that information, we use the following sample criteria: 1) for the affected ZIP code areas, we retain all birth records where mother's residential information matches with any of the affected ZIP codes; 2) by using the description of each wildfire and Google Maps, we narrow the comparison ZIP code areas to only the adjacent ZIP code areas, close to the affected ZIP code areas but never affected by any wildfire[14]; 3) using the coordinates information of mother's residential address, we can calculate the distance from mother's residence to the centroids of all the affected ZIP codes; 4) for the birth records of babies not exposed to any wildfire during mother's pregnancy, we only retain those records with mother's residential address within the adjacent ZIP code area as defined in step 2, and those living less than ten miles[15] away from the centroid of affected ZIP code areas (as calculated in step 3).

Based on these samples, we define the variable of exposure to any wildfire outbreak measured in each month of pregnancy (i.e. the first through last month of pregnancy) as follows: We create a dummy variable, "place", which equals 1 if any of these wildfires happened in the area where mother resides and 0 otherwise. Then, we create another dummy variable, "time_i", for each month of mother's pregnancy: It equals 1 if any of these wildfires happened during the i-th month of mother's pregnancy and 0 otherwise. Next, we use the variable "month_i", which equals "place" multiplied by "time_i", to identify exposure to any wildfire outbreak *in utero* for

[13] Our main results rely on the conception date counted forwards from mother's last menstrual period (LMP). We also use baby's birth date and clinical estimated gestation counting backwards to calculate the approximate conception date in the robustness check part.
[14] The criteria used to identify the comparing group will be altered to check the robustness of our estimation in the later section.
[15] The distance used to construct the comparing group will be adjusted to verify the robustness of the results in the later section.

each month of pregnancy. Therefore, a newborn is defined as "exposed to any wildfire event *in utero* during the i-th moth of mother's pregnancy" if the mother was i-month pregnant and lived in the affected areas (identified by ZIP code) when any of the wildfires happened. In addition, by aggregating the values of monthly indicators calculated above we have created a variable, "exposure", measuring "ever exposed to any wildfire event *in utero*". Both of these methods of measuring wildfire exposure are used in our estimation.

By following the steps above, we are able to construct two samples: the first sample contains 72,737 birth records with gestational length longer than 26 weeks, and the second sample includes 68,375 birth records with at least 37-week gestation. Tables 1-3, 1-4, 1-5 and 6 report the summary statistics for all the variables used in our analysis. We summarize the basic statistics by subgroups (defined by exposure) for both samples. Tables 1-3 and 1-5 report statistics based on the sample with gestational age greater than 26 weeks (46,280 births), while table 1-3 is for the sample without pollution data added in, and table 5 is for the sample with pollution variables. The basic statistics in tables 4 and 6 are summarized based on the sample with gestational age greater than or equal to 37 weeks (i.e. with at least 9 months of pregnancy, 68,627 births in total), without and with the pollution variables respectively. Column 1 in all the tables (i.e. table 1-3 to table 1-5) shows means for the corresponding sample with gestational age greater than 26 weeks or at least 37 weeks, which include all the births born during 2004 to 2012 matching our sample criteria listed in the previous section. We further divide the whole sample into two subgroups according to whether they have been exposed to any wildfire or not. Column 2 in the tables list the mean values of different variables for the newborns who have never experienced any significant wildfire *in utero,* and column 3 summarizes variables for those newborns who have experienced at least one wildfire event *in utero.*

Comparisons of columns 2 and 3 demonstrate that there is no significant difference in the prevalence of adverse birth outcomes (i.e. low birth weight, high morbidity of abnormal conditions and high morbidity of complications at birth/delivery) between the two groups (i.e. infants never exposed to any wildfire and infant exposed to at least one wildfire). For both samples, the proportions of births ever exposed to any wildfire are very similar (around 6 percent), providing the evidence that exposure to wildfire events is determined exogenously due to the unexpected nature of wildfire occurrences. Comparing mothers ever exposed to any wildfire during her pregnancy to those who were not, we find difference is that mothers exposed to wildfires are more likely to be married, less likely to be black or Hispanic, less likely to be high school dropouts, and more likely to have a college degree than the mothers in the other group. There is literature[16] suggest that differences in social economic status can be the potential reason for the differences in health status and physical and mental stress. The differences in mother's characteristics in our sample indicate that mothers exposed to wildfires might have more advantages than mothers who were not exposed because of their social economic status, which implies that the estimated effect of maternal stress triggered by wildfire events in our model might be underestimated for the treatment group, because there is literature suggesting that mothers with worse social economic status (which is the control group in our model) may suffer higher maternal stress and have worse health condition during pregnancy.

[16] Williams et al. (1997); Lazzarino et al. (2014).

1.4.2. Model Specification

Our main estimations are based on two regression models specified as follows, which use different measures of wildfire exposure discussed in the above section.

$$y_{izt} = \beta_0 + \beta_1 * exposure_{izt} + \pi'\boldsymbol{X}_{izt} + \gamma'\boldsymbol{M}_{izt} + \theta'\boldsymbol{C}_{izt} + \mu'\boldsymbol{P}_{izt} + \alpha_c + \delta_t + \varepsilon_{izt} \qquad (1)$$

$$y_{izt} = \beta_0 + \sum_{k=1}^{9} \beta_k * month\_k_{izt} + \pi'\boldsymbol{X}_{izt} + \gamma'\boldsymbol{M}_{izt} + \theta'\boldsymbol{C}_{izt} + \mu'\boldsymbol{P}_{izt} + \alpha_c + \delta_t + \varepsilon_{izt} \qquad (2)$$

In equation (1), $exposure_{izt}$ is an indicator equals to 1 if this newborn i, who was born in year t and his/her mother lived in zip code area z, has ever been exposed to any wildfire *in utero*, and 0 otherwise, no matter how many wildfire events or how many days of exposure he/she experienced *in utero*. We use equation (1) as the regression model for both samples constructed in section 4.1 (i.e. samples with >26 weeks and >=37 weeks gestational length). Regressions based on equation (1) estimate the overall effects of ever exposed to any significant wildfire outbreak without differentiating the timing effect of exposure.

In equation (2), we break down the aggregate indicator “exposure” into 9 indicators of wildfire exposure measured in each month of pregnancy, to investigate which stage would be the most important or fragile stage for the fetus, if his/her mother is exposed to any wildfire during her pregnancy. Indicators month_1 to month_9 represent whether the newborn has ever been exposed to any wildfire through the first month to the last month of the pregnancy. Note that we can only use equation (2) on the full-term sample (i.e. births with at least 37 weeks gestational age), because only mothers who were pregnant more than 36 weeks experienced nine months of pregnancy. Estimation based on equation (2) allows us to investigate the effects of exposure to wildfire outbreaks by taking the time of exposure into account.

In both equations (i.e. equation (1) and (2)), y_{izt} denotes birth outcomes for child i born in year t, whose mother lived in zip code area z, and we focus on the effect of prenatal exposure to wildfire on birth outcomes, including birth weight (measured in grams), the occurrences of low birth weight (LBW), the occurrences of any abnormal conditions at birth[17], Apgar scores, and the occurrences of preterm birth. $\boldsymbol{X}_{izt}$ is a vector of variables representing all the available information on baby's and mother's characteristics, which include indicators of baby's sex, baby's born season, baby's birth order, whether this baby is born by C-Section, mother's age categories[18], mother's race, mother's ethnicity, mother's education categories[19], and whether mother was married during this pregnancy. $\boldsymbol{M}_{izt}$ is a vector of variables measuring maternal behavior during pregnancy, which includes indicators of whether mother used any tobacco, alcohol or drug during her pregnancy, whether mother had adequate prenatal care[20], and how much weight mother gained during her pregnancy[21]. $\boldsymbol{C}_{izt}$ controls for the pre-existing risks of this pregnancy, including indicators of having any medical risk and/or having any congenital anomalies. $\boldsymbol{P}_{izt}$ represents measurements of pollution that every baby was exposed to during his/her mother's pregnancy, including particulate matter (PM2.5), carbon monoxide (CO), sulfur dioxide ($\mathrm{SO_2}$), nitrogen dioxide ($\mathrm{NO_2}$) and ozone[22]. To capture any impact of differences in geographic regions and birth year, county fixed effect α_c and year fixed effect δ_t are controlled in the regression models. In addition, the interaction of county fixed effect and year fixed effect, as well as the time trend

17 The reported abnormal conditions of newborn include: Anemia (Hct.<39/Hgn.<13), Birth Injury, Fetal Alcohol Syndrome (FAS), Hyaline Membrane Disease/RDS, Meconium Aspiration Syndrome, Assisted Ventilation (< 30 Min.), Assisted Ventilation (>=30 Min.), Seizures and other abnormal conditions.

18 Mother's age has been categorized into 3 groups: 20-24, 25-34, and 35+.

19 Mother's education level has been categorized into 5 groups: less than high school, high school diploma, some college, college degree, and graduate education.

20 Adequate prenatal care is measured according to Kessner Index.

21 Weight gained during pregnancy has been categorized into 4 groups: 0/missing, <16 lbs., 16-60 lbs., and 60+ lbs.

22 Our main results are based on equation (1) and (2), using pollution level measured as trimester average.

variable are also controlled in alternative specifications as robustness check. At the end of each regression model, ε_{izt} is added as an error term for each individual.

1.5. Results

1.5.1. The Effects of Exposure to Wildfires on Infant Birth Outcomes

Table 1-7 reports the estimated effects (at the aggregate level) of prenatal exposure to wildfires on birth outcomes (i.e. birth weight, abnormal conditions, the occurrence of cesarean section, pre-term birth, and Apgar score). These results are based on OLS regressions on equation (1), using birth records with at least 26-week gestational length and without controlling average monthly pollution. For all of the regressions in the following tables, individual-level variables[23], pre-existing risks[24], and maternal behaviors[25] during pregnancy are controlled. The summary statistics in tables 1-3, 1-4, 1-5 and 1-6 suggest that if mothers' characteristics and residential location are not controlled in the estimation, we might get misleading results. For instance, mothers who live near the wildland areas are typically more likely to be exposed to wildfires, however, they are also more educated and may have better access to health care. Without controlling these characteristics, the estimated effects of exposure could be biased down or even reversed. Meanwhile, mothers who are exposed to wildfires also tend to have slightly higher rate of risky behaviors, which might worse their babies' birth outcomes. And failure to control for these factors could yield overestimated effects.

[23] Individual level variables control for baby's sex, first child or not, seasonality, mother's age, mother's race, mother's ethnicity, mother's marital status, mother's education, and baby born by C-Section or not.
[24] Pre-existing risks include medical risk for this pregnancy and congenital anomalies.
[25] Maternal behaviors control tobacco use, alcohol use, drug use, mother's weight gain during this pregnancy, and adequate prenatal care or not.

The results in Table 1-7 suggest that wildfire exposure during pregnancy has a statistically significant adverse effect on birth weight for infants who were ever exposed to any wildfire in utero, which decreases birth weight by approximately 20 grams. Because the estimation in Table 7 focuses on birth records with gestational length greater than 26 weeks, and pre- term (i.e. with gestational length less than 37 weeks) is one of the important birth outcomes that has been emphasized by a significant literature and which could also affect birth weight directly, we checked the effect of prenatal exposure to wildfire on the probability of pre-term birth (<37 weeks), and controlled the effect of being premature in the estimation for the other birth outcomes. Column (6) in Table 1-7 shows the estimated effect of wildfire exposure on the occurrence of pre-term birth; it indicates that prenatal exposure to wildfire does not have a significant effect on the occurrence of pre-term birth. There is literature suggests that prenatal exposure to wildfire might cause pre-term birth by shortening gestational length, or full-term birth with relatively slower fetal growth in utero. The results in Table 1-7 find that there is no significant effect on the length of gestation but there is effect on birth weight, implying that prenatal exposure to wildfire might affect infant health at birth by causing intrauterine growth restriction that slows fetal growth *in utero*.

Based on the sample with at least 26-week gestational length, Table 1-8 reports the estimates with control in average monthly air pollution levels. After adding measurements of pollution into our regression, we observe more significant negative effects on birth weight, and with slightly higher magnitudes. One possible explanation for this result is that air pollution could be affected by the wind direction, which might blow pollutants to the areas that are at the downwind of the wildfire location since these pollutants can travel a very long distance. For mothers who live in the downwind regions and not in any of the affected regions of wildfires, they are defined as not exposed to any wildfire. However, wildfires might still affect them by polluting the air in their living areas with the help of wind, causing worse birth outcomes for their babies. If measures of pollutants are not controlled, the effects caused by maternal stress will mix with the potential effects of air pollution. With controlling for the effect of air pollution generated by wildfire smoke, we are able to separate the effects caused by maternal stress from the effects of air pollution and estimate the effect of maternal stress triggered by wildfire exposure itself.

Our results based on Table 1-7 and 1-8 suggest that being exposed to any significant wildfire *in utero* decreases birth weight of a newborn by approximately 39 grams on average. And that effect might work by slowing down fetus' growth in utero. However, effects of exposure to wildfires on birth outcomes other than birth weight are found not statistically significant. Our estimation in the following section will focus on birth weight as a summary measure of a newborn's health for two reasons: first, although birth weight is not a perfect measure of newborn's overall health status, it is a widely accepted and used measure since it has been proved to have critical effects on baby's further development; second, birth outcomes such as Apgar score and abnormal condition at birth might be affected by other contemporary shocks other than maternal stress, which will cause the confounding effects towards these outcomes. Focusing on birth weight, Tables 9 and 10 present the estimated effects of wildfire exposure on infant birth weight using different specifications that control for different sets of explanatory variables, with and without controlling for measures of pollution respectively. The last column in both tables show the results with full control of all the explanatory variables we have in our data set, which is the final estimated effects of being exposed to wildfire outbreaks based on the sample with at least 26 weeks gestational length.

1.5.2. The Timing Effects of Exposure to Wildfires

The results in Tables 1-7 and 1-8 provide evidence of the negative effect of wildfire exposure on newborn's birth weight, which also suggest that exposure to wildfire outbreaks is unlikely to significantly increase pre-term births. In this section, we focus on the full-term births and further test the effects of being exposed to wildfires on birth outcomes at different stages (i.e. first month through the last month of pregnancy) of the pregnancy in order to investigate when prenatal exposure might cause the most significant or dangerous impact on infant birth outcomes. Estimating the effects of exposure at finer time scales will help us identify the critical windows of prenatal exposure to wildfire outbreaks, during which exposure can cause severer adverse effects to infant health, compared to the other stages of pregnancy.

Table 1-11 reports the results based on equation (1) based on all live singleton births with at least 37-week gestational length. Columns (1) and (2) show the estimates with and without controlling for pollution level respectively; they are consistent with our results in Table 1-7 and Table 1-8. Table 1-12 presents the estimates based on equation (2), which looks at the effect of exposure in different months of pregnancy on birth weight, with and without controlling for pollution level respectively. The results in Table 1-12 suggest that being exposed to wildfire during the early stage of pregnancy, the first trimester (i.e. the 1st month of pregnancy to the 3rd month of pregnancy), has the most significant effect on birth weight. After controlling for air pollution, the negative effect of wildfire exposure still exists, at an even higher level, and the impact on the earlier stages becomes more significant. These findings suggest that fetus could be most vulnerable during the earlier stage of pregnancy due to the fact that baby's most critical development happends during the first trimester of pregnancy therefore there is the possibility that fetus at this stage is more sensitive and fragile to the negative outside shock.

Overall, the estimates based on equation (1) and (2) using two different samples (gestation of at least 26 weeks and at least 37 weeks) are consistent, and the effect of wildfire exposure on birth weight persists after controlling for air pollution. Prenatal exposure to any wildfire could significantly decrease infant birth weight, and this effect is more significant if that exposure happened during the earlier stage of pregnancy, especially the first trimester. We observe no significant effect on pre-term births, and the negative effect on birth weight exists among different samples at comparable magnitudes, suggesting that the mechanism through which the adverse effect decreases birth weight is by imposing intrauterine growth restriction.

1.6. Robustness Check

1.6.1. Testing the Sample Selection

There is the possibility that a certain group of mothers are more likely to live in the neighborhoods that have more wildfire outbreaks, therefore lead to sample selection problems. For example, mothers earning higher income or with higher preferences for suburban or rural environment might be more likely to live in the areas that have more wildfire outbreaks. However, on the other side, these mothers might also have better access to better prenatal care, which could mitigate the negative effects caused by wildfire exposure. To make sure that the possible bias caused by selection into sample is not a potential problem for our estimation, we examine whether there is a significant difference in mother's characteristics -including mother's race, ethnicity, age, education, whether mother is married, and mother's risky behaviors during pregnancy- by comparing these variables across treated and control groups.

Tables 1-13 and 1-14 report the results of these tests based on birth records with at least 26-week gestational age. Panel A shows the results based on the sample without restriction of non-missing pollution data; Panel B presents results based on the sample with this restriction. Consistent with what we have observed from the summary statistics in Tables 1-3 to 1-6, we find that mothers who were exposed to any wildfire outbreaks are indeed more likely to be married, less likely to be black, more likely to have some college education, and more likely to give birth at age 25 to 34. Although we do not have variable on household or family income in our data, we could assume mothers in the exposure group are more likely to have higher income and better prenatal care, compared to the other group of mothers, given the facts about mothers' characteristics we have found. These results further imply that the estimates based on these samples are very likely to underestimate the true effects of maternal stress triggered by wildfire outbreaks because the unobserved higher income and better prenatal care can mitigate the possible negative effects caused by wildfire exposure for mothers in the exposure group. Furthermore, there is no significant difference observed for mother's risky behaviors during pregnancy across two groups, indicating that mothers' risky behaviors during pregnancy will not cause potential bias to our estimation. Tables 1-15 and 1-16 are also testing the difference of mothers' characteristics across difference groups of mothers, based on the sample with only full-term births. The results reported in Table 1-15 and 1-16 are consistent with the results reported in Tables 1-13 and 1-14, indicating that the differences we have observed persist across different samples and there is a high probability that the true impact of wildfire exposure has been underestimated.

1.6.2. Restricting the Size of Wildfire

In our main estimation, we have used all the significant wildfires[26], which burned more than 150 acres, to identify who was exposed to significant wildfire outbreaks. In order to check

[26] The significant wildfire is defined as having caused significant damage by the State of New Jersey, in terms of acres burned, houses destroyed, people injury and so on.

the robustness of our results, we tested how the estimated results vary given different restrictions on the severity of wildfires, and the information for wildfires that burned more than 250 acres and 350 acres is used to define the alternative comparison groups.

Tables 1-17 and 1-18 report the results of regressions using equation (1) based on the samples constrained by the alternative restrictions on the size of wildfire. As these tables indicate, the negative effect of wildfire exposure persists across different samples, and our results are robust across samples and are consistent to our main results. In addition, we have observed that the magnitudes of our estimates are increasing as we use more severe wildfires to define the exposed group, which proves our hypothesis that wildfire could cause maternal stress and more severe wildfires will cause greater maternal stress, which would further increase the adverse effect on birth weight.

1.6.3. Adjusting for the Distance

As we mentioned in the previous section, we use distance from mother's residential address to the centroid of affected area to identify the unaffected births in the comparison group. For the unaffected group in our main estimation, we keep only birth records with mothers who reside within ten miles from the zip code centroid of affected areas. Furthermore, these mothers' residential areas should never be affected by any of the wildfires during years 2003 to 2012. These restrictions help to control for the unobserved geographic differences between affected and unaffected groups, and avoid any possibility of having exposed mothers in the control group All these restrictions have kept our control group relatively pure. In order to test the validity of these restrictions, we use different distances (i.e. 5 miles, 15 miles, and 20 miles) to redefine the comparison group and to test the robustness of our main results.

Tables 1-19 and 1-20 present the results of regressions using equation (1) based on alternative control groups defined by different distance restrictions. The results demonstrate that the negative effect caused by wildfire exposure persists across different samples defined by different distance restrictions. The significance and the magnitude of the estimated effect dose not vary a lot across samples, which indicate that our estimates are consistent with our main results, which are robust to the distance restrictions.

1.6.4. Alternative Measurement of Prenatal Exposure to Wildfires

In order to define prenatal exposure to wildfire, we started with baby's birth date and count backward to determine the variable indicating the year and month for each month of pregnancy (i.e. the first month through the last month of pregnancy). We then use this variable to further define whether this baby was exposed to any wildfire *in utero* during each month of pregnancy by comparing the year-month variable with the time of each wildfire occurrence. Given the information we have, there is an alternative method we could use to define the variable exposure, which is to use mother's last menses date (LMP) counting forward to determine the approximate starting date of the pregnancy, and then compare the approximate starting date and exact ending date[27] of pregnancy with the starting and ending dates of each wildfire to determine whether the wildfire happened during the time range of mother's pregnancy. If mother lived in the affected area when the wildfire occurred, and that wildfire occurred within the time range of her pregnancy, then she is defined as exposed to this particular wildfire. The estimates based on this alternative method are presented in Table 1-21. As we observe there, our main estimates are not affected by this alternative method and are robust across samples defined by different methods.

[27] The ending date of each pregnancy should be the date when the baby was born.

1.7. Conclusion

As an emotional and psychological condition, maternal stress is very difficult to quantify. Although cortisol level can be used to measure maternal stress, studies based on such data is hard to be generalized to a large population. Wildfires in New Jersey provide us natural experiments to estimate the impact of exposure to wildfire outbreaks during pregnancy on infant birth outcomes. Focusing on the channel of maternal stress, we find that wildfire exposure could negatively affect birth outcomes by increasing maternal stress. The occurrence of wildfires is an unexpected and unpredictable natural source of stressful exogenous shocks. These shocks could trigger higher maternal stress if experienced during pregnancy.

Although there is a significant literature suggesting that stressful event might cause adverse effects on birth outcomes, the results vary across studies and are limited by potential problems such as small sample size, lack of information needed to precisely identify and measure exposure to stressful events, and migration behavior in response to the stressful events. Wildfire, as a common natural event, could happen anywhere near or in the wildland, and its damage could be minor, moderate, or serious depending on the local conditions (e.g. dry weather and the composition of forest). Therefore, it is difficult for people to avoid wildfires completely by selecting certain areas to live, which helps to mitigate problems caused by migration in our analysis. By using the detailed information on each significant wildfire that occurred between 2003 and 2012 in New Jersey, the news reports related to each wildfire published online and mother's address information, we are able to precisely pin down the affected areas of each significant wildfire at the zip code level and to accurately identify whether a fetus was exposed to any wildfire *in utero*. And this provides us a relatively large sample with sufficient individual level variations.

Our main results suggest that being exposed to wildfire outbreaks *in utero* could significantly decrease infant birth weight. Further, that adverse effect on birth weight is more powerful at the earlier stage (especially the first trimester) of pregnancy. After ruling out other possible channels through which wildfire outbreaks could affect birth outcomes, we find that the adverse effect of wildfire exposure still exists. This implies that such effect could be explained by the mechanism of increasing maternal stress that is triggered by wildfire outbreaks. Because we don't find that prenatal exposure to any wildfire significantly increases the occurrences of having a preterm baby, we can conclude that maternal stress might adversely affect birth weight by imposing intrauterine growth restriction on the fetus who were exposed to any wildfire event *in utero*. Although the estimated effect on birth weight is at a relatively low magnitude, it might be crucial for the newborns whose birth weight falls at the edge of low birth weight since this effect could drive them to the low birth weight tier[28].

The climate and geological condition determine that wildfire occurrences in New Jersey are typically not severe and can be controlled in a relatively short time, implying that stress caused by these wildfires could be moderate or mild. Moreover, for most of the communities near the wildland area, there are programs designed to periodically alert residents about the potential risk of facing wildfire and prepare them for possible wildfire outbreaks, which might help to reduce the stress mothers experienced during wildfire outbreaks. Given these facts, the negative effects caused by stress triggered through wildfire exposure could be at lower magnitudes, compared with other catastrophic natural events. Based on these facts, the results of our study could approximate a lower bound on the effects of maternal stress and it would be worthwhile to extend our study to

[28] We have run the quantile regressions on our sample, and the results suggested that effects of wildfire exposure are most significant for the first 5% quantile, with the birth weight around 2580 grams.

estimate the effects of maternal stress triggered by wildfire outbreaks in states with more frequent and severe wildfires, such as California.

The importance of mother's physical health condition and her nutrition intake during pregnancy has been emphasized by numerous literature in epidemiology and economics, and these conditions are typically checked by doctors periodically during mother's pregnancy. However, it seems that lots of mothers have not realized what bad emotional and psychological condition can do to their babies, and therefore ignored the importance of healthy emotional and psychological condition during pregnancy. Using a large-scale individual-level data set with address information and detailed information of wildfire events, we have found and proved the significant impact of mothers' emotional and psychological condition on infant health at birth, suggesting that mothers should pay enough attention to their emotional and psychological health during her pregnancy, especially during the first trimester. In addition, our study might also provide some meaningful suggestions on the debate over whether it is worthwhile to manage potential wildfire risk by providing the evidence of the possible cost imposed on infant health. Since wildfires might be more likely to occur as more droughts accompany global warming, our study might also contribute to ongoing studies of the possible impact of global warming from a different angle.

1.8. References

Aizer, A., L. Stroud and S. Buka, "Maternal Stress and Child Outcomes Evidence from Siblings," Journal of Human Resources, 51 (2016), 523-555.

Almond, D. and J. Currie, "Human Capital Development before Age Five" Handbook of Labor Economics, 4B (2011), 1315-1486.

Black, S. E., P. J. Devereux and Kjell G. S., "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," The Quarterly Journal of Economics, 122 (2007), 409-439.

Breton, C., C. Park and J. Wu, "Effect of Prenatal Exposure to Wildfire-generated PM2.5 on Birth Weight," Epidemiology, 22 (2011), S66-S66.

Camacho, A., "Stress and Birth Weight: Evidence from Terrorist Attacks," American Economic Review, 98 (2008), 511-515.

Chay, K. Y. and M. Greenstone, "The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession," Quarterly Journal of Economics, 118 (2003), 1121-1167.

Currie, J. and R. Hyson, "Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight," The American Economic Review, 89 (1999), 245-250.

Currie, J., M. Neidell and J. F. Schmieder, "Air pollution and infant health: Lessons from New Jersey," J Health Econ, 28 (2009), 688-703.

Currie, J. and M. Rossin-Slater, "Weathering the storm: hurricanes and birth outcomes," J Health Econ, 32 (2013), 487-503.

Currie, J. and M. Neidell, "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?," The Quarterly Journal of Economics, 120 (2005), 1003-1030.

Davis, E. P. and C. A. Sandman, "The timing of prenatal exposure to maternal cortisol and psychosocial stress is associated with human infant cognitive development," Child Dev, 81 (2010), 131-148.

de Quervain, D. J., B. Roozendaal and J. L. McGaugh, " Stress and glucocorticoids impair retrieval of long-term spatial memory," Nature, 394 (1998), 787-790.

Dickerson, S. S. and M. E. Kemeny, " Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research," Psychological Bulletin, 130(2004), 355-391.

Schetter, C. D., "Psychological science on pregnancy: stress processes, biopsychosocial models, and emerging research issues," Annu Rev Psychol, 62 (2011), 531-558.

Eskenazi, B., A. R. Marks, R. Catalano, T. Bruckner and P. G. Toniolo, "Low birthweight in New York City and upstate New York following the events of September 11th," Hum Reprod, 22 (2007), 3013-3020.

Glinianaia, S. V., J. Rankin, R. Bell, T. Pless-Mulloli and D. Howel, "Particulate air pollution and fetal health: a systematic review of the epidemiologic evidence," Epidemiology, 15 (2004), 36-45.

Holstius, D. M., C. E. Reid, B. M. Jesdale and R. Morello-Frosch, "Birth Weight following Pregnancy during the 2003 Southern California Wildfires," Environmental Health Perspectives, 120 (2012), 1340-1345.

Johnston, F. H., A. M. Kavanagh, D. M. Bowman and R. K. Scott, "Exposure to bushfire smoke and asthma: an ecological study," Med J Aust, 176 (2002), 535-538.

Kirschbaum. C., J. C. Prussner, A. A. Stone, I. Federenko, J. Gaab, D. Lintz, N. Schommer and D. H. Hellhammer, " Persistent High Cortisol Responses to Repeated Psychological Stress in a Subpopulation of Healthy Men," Psychosomatic Medicine, 57 (1995), 468-474.

Lazzarino. A. I., V. Yiengprugsawan, S. Seubsman, A. Steptoe and A. C. Sleigh, "The Associations between Unhealthy Behaviours, Mental stress, and Low Socio-economic Status in an International Comparison of Representative Samples from Thailand and England," Global Halth, 10 (2014).

Lederman, S. A., V. Rauh, L. Weiss, J. L. Stein, L. A. Hoepner, M. Becker and F. P. Perera, "The effects of the World Trade Center event on birth outcomes among term deliveries at three lower Manhattan hospitals," Environmental Health Perspectives, 112 (2004), 1772-1778.

Lipkind, H. S., A. E. Curry, M. Huynh, L. E. Thorpe and T. Matte, "Birth Outcomes Among Offspring of Women Exposed to the September 11, 2001, Terrorist Attacks," Obstetrics and Gynecology, 116 (2010), 917-925.

Eccleston, M., "In Utero Exposure to Maternal Stress: Effect of 9/11 on Bith and Early Schooling Outcomes in New York City," working paper, (2011).

Schlenker, W. and W. R. Walker, "Airports, Air Pollution, and Contemporaneous Health," Review of Economic Studies, 83 (2016), 768-809.

Siddiqui, A. R., E. B. Gold, X. Yang, K. Lee, K. H. Brown and Z. A. Bhutta, "Prenatal exposure to wood fuel smoke and low birth weight," Environ Health Perspect, 116 (2008), 543-549.

Simeonova, E., "Out of Sight, Out of Mind? Natural Disasters and Pregnancy Outcomes in the USA," CESifo Economic Studies, 57 (2011), 403-431.

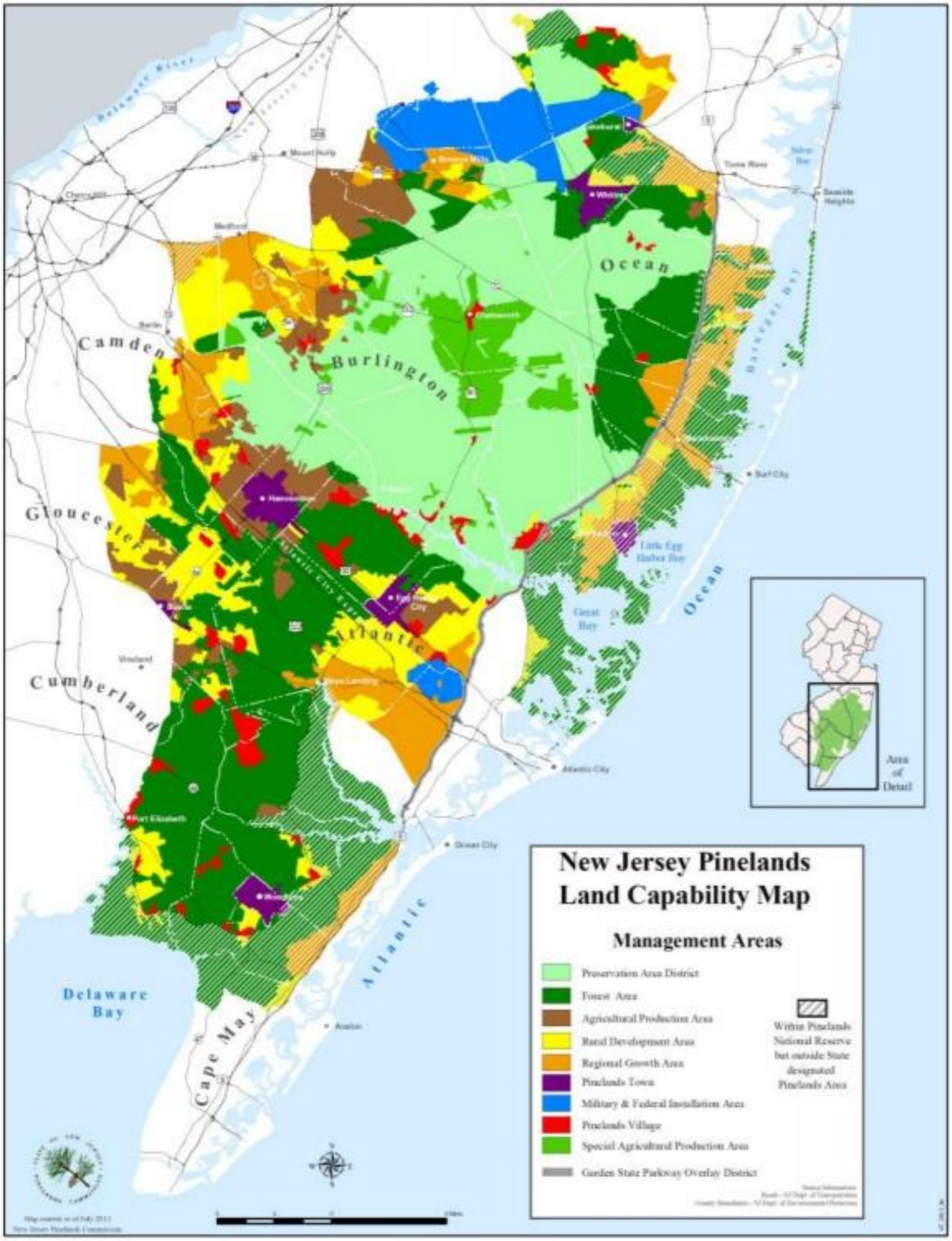
Tan, C. E., H. J. Li, X. G. Zhang, H. Zhang, P. Y. Han, Q. An, W. J. Ding and M. Q. Wang, "The Impact of the Wenchuan Earthquake on Birth Outcomes," PLoS One, 4 (2009).

Williams. D. R., Y. Yu, J. S. Jackson and N. B. Anderson, "Racial Differences in Physical and Mental Health: Socio-economic Status, Stress and Discrimination, " Journal of Health Psychology, 23(1997), 275-282.

Wust, S., I. Federenko, D. H. Hellhammer and C. Kirschbaum, "Genetic factors, perceived chronic stress, and the free cortisol response to awakening," Psychoneuroendocrinology, 25 (2000), 707-720.

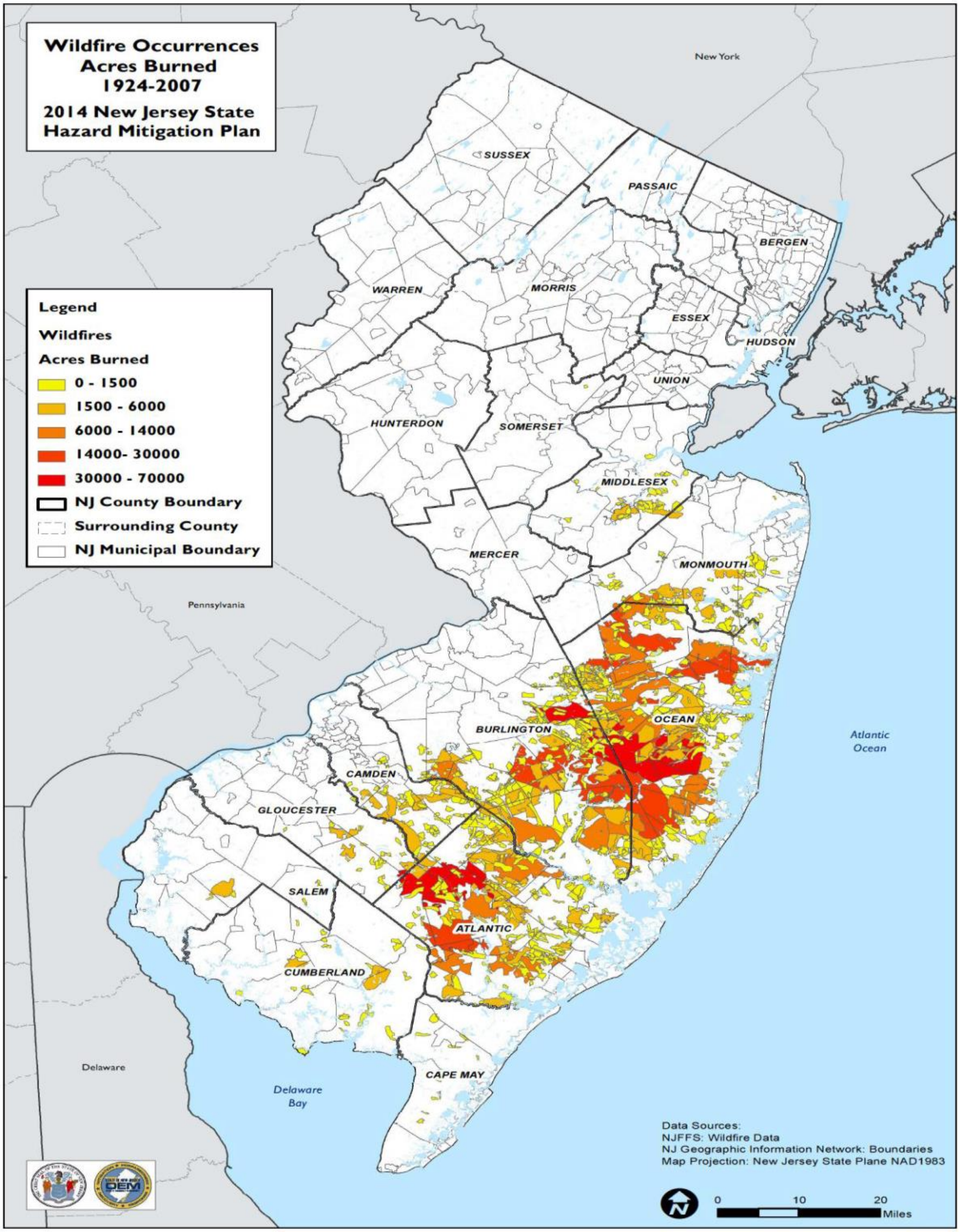
Xiong, X., E. W. Harville, D. R. Mattison, K. Elkind-Hirsch, G. Pridjian and P. Buekens, "Exposure to Hurricane Katrina, post-traumatic stress disorder and birth outcomes," American Journal of the Medical Sciences, 336 (2008), 111-115.

Figure: 1-1: New Jersey Pineland Management Areas



Source: New Jersey Pinelands Commission 2012

Figure 1-2: Past Occurrences of Wildfires in New Jersey



Source: NJFFS 2013

Table 1-1: Literature on Effect of Maternal Stress Caused by Natural Disaster

Study	Event	Data	Measure of Exposure	Outcomes	Findings
Lederman et al. (2004)	911 Terrorist Attack	A sample of 300 women with information collected from medical records and interviews.	Women who lived or worked within 2-mile radius at some time in the 4 weeks after 911 were classified into exposure group. Anyone not meet the condition was in the reference group.	birth weight, birth length, head circumference, ponderal index, SGA births	Women in the first trimester of pregnancy at the time of the attack delivered infants with significantly shorter gestation and smaller head circumference.
Eskenazi et al. (2007)	911 Terrorist Attack	Birth certificate data for NY residents	Birth outcomes are compared across births in the week after 911 and births in the weeks before 911.	low birth weight, preterm birth	911 events are associated with slightly delayed decreased preterm delivery, and delayed increases in LBW among infants exposed in the first two trimesters.
Lipkind et al. (2010)	911 Terrorist Attack	446 women enrolled in the World Trade Center Health Registry.	Women within the Registry with a gestation that overlapped with the period September 11, 2001 through December 1, 2001 are defined as exposed.	birth weight, low birth weight, gestation, preterm birth	Women who lived, worked or were near the World Center on or soon after 911 has pregnancy outcomes similar to women residing more than 5 miles away.
Camacho (2008)	Landmine Explosions	Vital Statistics Records collected in Colombia from 1998-2003	Exposure is identified by matching landmine explosions with mothers' residential area.	birth weight	Intensity of random landmine explosions during a woman's first trimester of pregnancy has significant negative impact on child birth weight.

Table 1-1: Literature on Effect of Maternal Stress Caused by Natural Disaster

Study	Event	Data	Measure of Exposure	Outcomes	Findings
Xiong et al. (2008)	Hurricane Katrina	A sample of 301 women from 2 prenatal care clinics	Women were interviewed about their hurricane experience and their psychological response. The exposures of this study are classified as high hurricane experience, PTSD, and depression.	low birth weight, preterm birth	Women who had high hurricane exposure were at an increased risk of having low birth weight infants. Exposure to specific sever disaster events may lead to even worse outcomes.
Simeonova (2009)	Extreme Weather Events	The national vital statistics from 1968-1988	Exposure is identified by matching extreme weather events with the county of residence for the mother.	prematurity, gestation, low birth weight	Exposure to extreme events increases the risk of preterm births.
Tan et al. (2009)	Wenchuan Earthquake	A sample of 13,003 newborns.	The sample is divided into pre-earthquake group and post-earthquake group according to birth date.	Birth weight, preterm birth, birth defect, Apgar score	Low birth weight, high ratio of low birth weight and low Apgar scores of post-earthquake group were observed.

Table 1-2: Number of Significant Wildfires from 2003 to 2012 by County

County	Number of Significant Wildfires
Atlantic	2
Burlington	5
Camden	2
Cape May	2
Cumberland	0
Essex	0
Gloucester	2
Hudson	0
Hunterdon	1
Mercer	0
Middlesex	5
Monmouth	0
Morris	2
Ocean	7
Passaic	0
Salem	1
Somerset	0
Sussex	0
Union	0
Warren	0

Table 1-3: Variable Means of Births with Gestation >26 Weeks (no pollution)

	Gestation: >26 weeks		
	All NJ Births	Births Never Exposed to Wildfire	Birth Ever Exposed to Wildfire
	(N=72,737)	(N=68,375)	(N=4,362)
Birth weight (measured in grams)	3,372.083	3,372.094	3,371.915
Low birth weight (<2500g)	0.034	0.034	0.034
Any abnormal conditions of newborn (1/0)	0.010	0.010	0.012
Any complications labor/delivery (1/0)	0.534	0.535	0.515
Gestation (measured in weeks)	39.554	39.553	39.569
Preterm (1/0)	0.057	0.057	0.054
Child is male (1/0)	0.509	0.509	0.514
Child born in spring (1/0)	0.250	0.252	0.217
Child born in summer (1/0)	0.263	0.260	0.311
Child born in fall (1/0)	0.252	0.250	0.282
Child born in winter (1/0)	0.236	0.239	0.190
First child (1/0)	0.399	0.399	0.403
C-Section delivery (1/0)	0.355	0.353	0.385
Mother is married (1/0)	0.738	0.735	0.794
Mother is black (1/0)	0.102	0.104	0.066
Mother is Hispanic (1/0)	0.224	0.232	0.095
Mother's age 20-24 (1/0)	0.146	0.145	0.153
Mother's age 25-34 (1/0)	0.605	0.603	0.633
Mother's age 35+	0.230	0.252	0.214
Mother's ed: <HS (1/0)	0.108	0.112	0.051
Mother's ed: HS degree (1/0)	0.231	0.229	0.248
Mother's ed: some college (1/0)	0.194	0.190	0.248
Mother's ed: college (1/0)	0.256	0.256	0.260
Mother's ed: college+ (1/0)	0.211	0.212	0.193
Mother smoked during pregnancy (1/0)	0.057	0.054	0.090
Mother had alcohol during pregnancy (1/0)	0.007	0.007	0.011
Mother had drugs during pregnancy (1/0)	0.008	0.008	0.012
Mother gained 0/missing	0.013	0.013	0.017
Mother gained <16 lbs (1/0)	0.102	0.103	0.097
Mother gained 16-60 lbs (1/0)	0.870	0.870	0.867
Mother gained >60 lbs (1/0)	0.015	0.015	0.019
Adequate prenatal care (Kessner)	0.807	0.808	0.793
Any medical risk for this pregnancy (1/0)	0.415	0.413	0.450
Any congenital anomalies for this pregnancy (1/0)	0.006	0.006	0.007

Table 1-4: Variable Means of Births with Gestation >=37 Weeks (no pollution)

	Gestation: >=37 weeks		
	All NJ Births	**Births Never Exposed to Wildfire**	**Birth Ever Exposed to Wildfire**
	(N=68,627)	**(N=64,502)**	**(N=4,125)**
Birth weight (measured in grams)	3,399.998	3,400.041	3,399.326
Low birth weight (<2500g)	0.023	0.023	0.024
Any abnormal conditions of newborn (1/0)	0.010	0.010	0.012
Any complications labor/delivery (1/0)	0.533	0.534	0.512
Gestation (measured in weeks)	39.759	39.759	39.766
Preterm (1/0)	0.000	0.000	0.000
Child is male (1/0)	0.508	0.507	0.515
Child born in spring (1/0)	0.250	0.252	0.218
Child born in summer (1/0)	0.263	0.260	0.313
Child born in fall (1/0)	0.253	0.251	0.281
Child born in winter (1/0)	0.234	0.237	0.188
First child (1/0)	0.401	0.401	0.401
C-Section delivery (1/0)	0.351	0.349	0.381
Mother is married (1/0)	0.742	0.738	0.795
Mother is black (1/0)	0.100	0.102	0.065
Mother is Hispanic (1/0)	0.221	0.230	0.094
Mother's age 20-24 (1/0)	0.145	0.144	0.154
Mother's age 25-34 (1/0)	0.608	0.606	0.636
Mother's age 35+	0.248	0.250	0.210
Mother's ed: <HS (1/0)	0.106	0.110	0.049
Mother's ed: HS degree (1/0)	0.229	0.227	0.246
Mother's ed: some college (1/0)	0.194	0.190	0.250
Mother's ed: college (1/0)	0.259	0.258	0.262
Mother's ed: college+ (1/0)	0.213	0.214	0.193
Mother smoked during pregnancy (1/0)	0.056	0.054	0.089
Mother had alcohol during pregnancy (1/0)	0.007	0.007	0.010
Mother had drugs during pregnancy (1/0)	0.008	0.007	0.012
Mother gained 0/missing	0.012	0.012	0.017
Mother gained <16 lbs (1/0)	0.100	0.100	0.094
Mother gained 16-60 lbs (1/0)	0.873	0.873	0.870
Mother gained >60 lbs (1/0)	0.015	0.015	0.018
Adequate prenatal care (Kessner)	0.811	0.812	0.796
Any medical risk for this pregnancy (1/0)	0.410	0.407	0.446
Any congenital anomalies for this pregnancy (1/0)	0.006	0.006	0.007

Table 1-5: Variable Means of Births with Gestation >26 Weeks (with pollution)

	Gestation: >26 weeks		
	All NJ Births	**Births Never Exposed to Wildfire**	**Birth Ever Exposed to Wildfire**
	(N=59,867)	**(N=57,880)**	**(N=1,978)**
Birth weight (measured in grams)	3,365.258	3,365.975	3,344.383
Low birth weight (<2500g)	0.035	0.035	0.039
Any abnormal conditions of newborn (1/0)	0.010	0.010	0.014
Any complications labor/delivery (1/0)	0.535	0.534	0.561
Gestation (measured in weeks)	39.545	39.547	39.489
Preterm (1/0)	0.057	0.057	0.055
Child is male (1/0)	0.509	0.509	0.513
Child born in spring (1/0)	0.254	0.253	0.281
Child born in summer (1/0)	0.260	0.258	0.315
Child born in fall (1/0)	0.249	0.249	0.237
Child born in winter (1/0)	0.237	0.240	0.168
First child (1/0)	0.398	0.397	0.418
C-Section delivery (1/0)	0.350	0.349	0.361
Mother is married (1/0)	0.745	0.739	0.898
Mother is black (1/0)	0.105	0.106	0.059
Mother is Hispanic (1/0)	0.247	0.252	0.081
Mother's age 20-24 (1/0)	0.141	0.141	0.122
Mother's age 25-34 (1/0)	0.603	0.601	0.655
Mother's age 35+	0.256	0.257	0.222
Mother's ed: <HS (1/0)	0.117	0.120	0.028
Mother's ed: HS degree (1/0)	0.217	0.218	0.184
Mother's ed: some college (1/0)	0.180	0.178	0.232
Mother's ed: college (1/0)	0.261	0.259	0.298
Mother's ed: college+ (1/0)	0.226	0.224	0.258
Mother smoked during pregnancy (1/0)	0.046	0.046	0.046
Mother had alcohol during pregnancy (1/0)	0.006	0.006	0.006
Mother had drugs during pregnancy (1/0)	0.006	0.006	0.005
Mother gained 0/missing	0.013	0.013	0.014
Mother gained <16 lbs (1/0)	0.100	0.101	0.084
Mother gained 16-60 lbs (1/0)	0.873	0.873	0.888
Mother gained >60 lbs (1/0)	0.014	0.014	0.014
Adequate prenatal care (Kessner)	0.811	0.811	0.800
Any medical risk for this pregnancy (1/0)	0.410	0.411	0.401
Any congenital anomalies for this pregnancy (1/0)	0.006	0.006	0.012

Table 1-6: Variable Means of Births with Gestation >=37 Weeks (with pollution)

	Gestation: >=37 weeks		
	All NJ Births	**Births Never Exposed to Wildfire**	**Birth Ever Exposed to Wildfire**
	(N=56,472)	**(N=54,595)**	**(N=1,877)**
Birth weight (measured in grams)	3,392.900	3,393.609	3,372.256
Low birth weight (<2500g)	0.024	0.024	0.028
Any abnormal conditions of newborn (1/0)	0.010	0.010	0.013
Any complications labor/delivery (1/0)	0.534	0.533	0.559
Gestation (measured in weeks)	39.751	39.753	39.686
Preterm (1/0)	0.000	0.000	0.000
Child is male (1/0)	0.507	0.507	0.511
Child born in spring (1/0)	0.254	0.253	0.283
Child born in summer (1/0)	0.260	0.258	0.320
Child born in fall (1/0)	0.250	0.251	0.234
Child born in winter (1/0)	0.236	0.238	0.164
First child (1/0)	0.399	0.399	0.413
C-Section delivery (1/0)	0.346	0.346	0.355
Mother is married (1/0)	0.748	0.743	0.901
Mother is black (1/0)	0.103	0.104	0.057
Mother is Hispanic (1/0)	0.243	0.249	0.081
Mother's age 20-24 (1/0)	0.14	0.140	0.124
Mother's age 25-34 (1/0)	0.606	0.604	0.658
Mother's age 35+	0.255	0.256	0.218
Mother's ed: <HS (1/0)	0.114	0.117	0.028
Mother's ed: HS degree (1/0)	0.215	0.216	0.181
Mother's ed: some college (1/0)	0.180	0.178	0.232
Mother's ed: college (1/0)	0.263	0.262	0.302
Mother's ed: college+ (1/0)	0.228	0.227	0.258
Mother smoked during pregnancy (1/0)	0.045	0.045	0.046
Mother had alcohol during pregnancy (1/0)	0.006	0.006	0.005
Mother had drugs during pregnancy (1/0)	0.006	0.006	0.004
Mother gained 0/missing	0.012	0.012	0.015
Mother gained <16 lbs (1/0)	0.098	0.098	0.083
Mother gained 16-60 lbs (1/0)	0.876	0.875	0.889
Mother gained >60 lbs (1/0)	0.014	0.014	0.014
Adequate prenatal care (Kessner)	0.814	0.815	0.800
Any medical risk for this pregnancy (1/0)	0.405	0.405	0.392
Any congenital anomalies for this pregnancy (1/0)	0.006	0.006	0.013

Table 1-7: The Effect of Exposure to Wildfire (Gestation >26 weeks, not control for pollution)

	Birth Weight (in grams)	Low birth weight (<2500g)	Any abnormal conditions of newborn (1/0)	Caesarean section (0/1)	Apgar Score at 5 minute < 7 (0/1)	Premature (0/1)
Ever exposed to any wildfire:						
Exposure or not (0/1)	-20.2513*	0.0012	-0.0027	0.0106	-0.0008	0.0010
	(10.7244)	(0.0031)	(0.0031)	(0.0087)	(0.0008)	(0.0027)
Individual level control variables	Yes	Yes	Yes	Yes	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes	Yes	Yes	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes	Yes	Yes	Yes	Yes
Pollution	No	No	No	No	No	No
Preterm	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	72,737	72,737	72,737	72,737	72,737	72,737

Notes: Each column in each panel is a separate OLS regression. The sample covers 2004-2010 cohorts of births. All singleton births with mother aged 20-45 years old are retained.

Individual level variables controlled for baby's sex, first child or not, seasonality, mother's age, mother's race, mother's ethnicity, mother's marital status, mother's education, and baby born by C-Section or not. Pre-existing risks for this pregnancy controlled for medical risk for this pregnancy and congenital anomalies. Maternal behaviors controlled for tobacco use, alcohol use, drug use, mother's weight gain during this pregnancy, and adequate prenatal care or not.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-8: The Effect of Exposure to Wildfire (Gestation >26 weeks, control for pollution)

	Birth Weight (in grams)	Low birth weight (<2500g)	Any abnormal conditions of newborn (1/0)	Caesarean section (0/1)	Apgar Score at 5 minute < 7 (0/1)	Premature (0/1)
Ever exposed to any wildfire:						
Exposure or not (0/1)	-39.0505**	0.0061	0.0004	0.0037	0.0007	0.0018
	(17.1114)	(0.0050)	(0.0018)	(0.0142)	(0.0012)	(0.0048)
Individual level control variables	Yes	Yes	Yes	Yes	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes	Yes	Yes	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes	Yes	Yes	Yes	Yes
Pollution	Yes	Yes	Yes	Yes	Yes	Yes
Preterm	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	59,867	59,867	59,867	59,867	59,867	59,867

Notes: Each column in each panel is a separate OLS regression. The sample covers 2004-2010 cohorts of births. All singleton births with mother aged 20-45 years old are retained.

Individual level variables controlled for baby's sex, first child or not, seasonality, mother's age, mother's race, mother's ethnicity, mother's marital status, mother's education, and baby born by C-Section or not. Pre-existing risks for this pregnancy controlled for medical risk for this pregnancy and congenital anomalies. Maternal behaviors controlled for tobacco use, alcohol use, drug use, mother's weight gain during this pregnancy, and adequate prenatal care or not.

Pollution controls for particulate matter (pm2.5), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and ozone.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-9: The Effect of Exposure to Wildfire on Birth Weight (Gestation >26 weeks, not control for pollution)

	Birth Weight (in grams)			
Ever exposed to any wildfire:				
Exposure or not (0/1)	-22.4251**	-21.7879*	-20.7242*	-20.2513*
	(11.0117)	(11.2203)	(10.8793)	(10.7244)
Individual level control variables	Yes	Yes	Yes	Yes
Pre-existing risks for this pregnancy	No	Yes	Yes	Yes
Maternal behaviors during pregnancy	No	No	Yes	Yes
Preterm	No	No	No	Yes
Pollution	No	No	No	No
County fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Obs. No.	72,737	72,737	72,737	72,737

Notes: Each column in each panel is a separate OLS regression. The sample covers 2004-2012 cohorts of births. All singleton births with mother aged 20-45 years old are retained.

Individual level variables controlled for baby's sex, first child or not, seasonality, mother's age, mother's race, mother's ethnicity, mother's marital status, mother's education, and baby born by C-Section or not. Pre-existing risks for this pregnancy controlled for medical risk for this pregnancy and congenital anomalies. Maternal behaviors controlled for tobacco use, alcohol use, drug use, mother's weight gain during this pregnancy, and adequate prenatal care or not.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-10: The Effect of Exposure to Wildfire on Birth Weight (Gestation >26 weeks, control for pollution)

	Birth Weight (in grams)				
Ever exposed to any wildfire:					
Exposure or not (0/1)	-42.3578**	-43.0793**	-41.9063**	-40.7939**	-39.0505**
	(19.1033)	(19.2166)	(18.0690)	(17.6816)	(17.1114)
Individual level control variables	Yes	Yes	Yes	Yes	Yes
Pre-existing risks for this pregnancy	No	Yes	Yes	Yes	Yes
Maternal behaviors during pregnancy	No	No	Yes	Yes	Yes
Preterm	No	No	No	Yes	Yes
Pollution	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Obs. No.	59,867	59,867	59,867	59,867	59,867

Notes: Each column in each panel is a separate OLS regression. The sample covers 2004-2012 cohorts of births. All singleton births with mother aged 20-45 years old are retained.

Individual level variables controlled for baby's sex, first child or not, seasonality, mother's age, mother's race, mother's ethnicity, mother's marital status, mother's education, and baby born by C-Section or not. Pre-existing risks for this pregnancy controlled for medical risk for this pregnancy and congenital anomalies. Maternal behaviors controlled for tobacco use, alcohol use, drug use, mother's weight gain during this pregnancy, and adequate prenatal care or not.

Pollution controls for particulate matter (pm2.5), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and ozone.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-11: The Effect of Exposure to Wildfire on Birth Weight (Gestation >=37 weeks)

	Birth Weight (in grams)	
	No Control for Pollution	**Control for Pollution**
Ever exposed to any wildfire:		
Exposure or not (0/1)	-21.2018**	-38.2859**
	(10.6301)	(16.6817)
Individual level control variables	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes
Pollution	No	Yes
Preterm	Yes	Yes
County fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Obs. No.	68,627	56,472

Notes: Each column in each panel is a separate OLS regression. The sample covers 2004-2012 cohorts of births. All singleton births with mother aged 20-45 years old are retained.

Individual level variables controlled for baby's sex, first child or not, seasonality, mother's age, mother's race, mother's ethnicity, mother's marital status, mother's education, and baby born by C-Section or not. Pre-existing risks for this pregnancy controlled for medical risk for this pregnancy and congenital anomalies. Maternal behaviors controlled for tobacco use, alcohol use, drug use, mother's weight gain during this pregnancy, and adequate prenatal care or not.

Pollution controls for particulate matter (pm2.5), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and ozone.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-12: The Effect of Exposure to Wildfire on Birth Weight (Gestation >=37 weeks)

	Birth Weight (in grams)	
	No Control for Pollution	**Control for Pollution**
Ever exposed to any wildfire:		
1st month of pregnancy	-34.0909	-52.9452*
	(25.3145)	(27.4874)
2nd month of pregnancy	-38.7679**	-51.5618*
	(18.0022)	(26.4616)
3rd month of pregnancy	-22.3848	-46.1511**
	(15.4467)	(18.5119)
4th month of pregnancy	-13.0495	-28.5007
	(24.9407)	(38.7902)
5th month of pregnancy	-2.2714	-10.3561
	(14.6780)	(22.6005)
6th month of pregnancy	1.8566	1.3985
	(19.2578)	(27.4808)
7th month of pregnancy	-76.5628***	-77.1787*
	(23.7342)	(40.3904)
8th month of pregnancy	-10.2184	-37.7121
	(21.1000)	(25.9482)
9th month of pregnancy	15.6133	-48.1261
	(26.5749)	(42.2638)
Individual level control variables	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes
Preterm	Yes	Yes
Pollution	No	Yes
County fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Obs. No.	68,627	56,472

Notes: Each column in each panel is a separate OLS regression. The sample covers 2004-2012 cohorts of births. All singleton births with mother aged 20-45 years old are retained.

Pollution controls for particulate matter (pm2.5), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and ozone.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-13: Maternal Characteristics and Exposure to Wildfires in New Jersey (Gestation >26 weeks)

	Mother's Age [20, 24]	Mother's Age [25, 34]	Mother's Age [35, 45]	Mother's Ed: <HS Degree	Mother's Ed: HS degree	Mother's Ed: Some College
Panel A: No Pollution						
	0.0004	0.0232	-0.0236	-0.0301	-0.0066	0.0242*
	(0.0233)	(0.0169)	(0.0190)	(0.0287)	(0.0281)	(0.0144)
Panel B: With Pollution						
	-0.0283	0.0530**	-0.0246	-0.0631	-0.0440	0.0280
	(0.0364)	(0.0253)	(0.0321)	(0.0476)	(0.0355)	(0.0198)

	Mother's Ed: College	Mother's Ed: College+	Mother is Married	Mother is Black	Mother is Hispanic
Panel A: No Pollution					
	0.0075	0.0050	0.0442	-0.0267*	-0.0628
	(0.0235)	(0.0293)	(0.0408)	(0.0150)	(0.0495)
Panel B: With Pollution					
	0.0369	0.0423	0.1179**	-0.0385**	-0.1128
	(0.0337)	(0.0409)	(0.0582)	(0.0175)	(0.0829)

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-14: Maternal Behaviors and Exposure to Wildfires in New Jersey (Gestation >26 weeks)

	Tobacco Use	Alcohol Use	Drug Use	Weight Gain: <= 15bls	Weight Gain: <= 60bls	Weight Gain: > 60bls	Weight Gain: 0 or Missing	Adequate Prenatal Care
Panel A: No Pollution								
	0.0118	0.0021	0.0024	0.0034	-0.0106	0.0062	0.0010	0.0062
	(0.0101)	(0.0022)	(0.0018)	(0.0026)	(0.0087)	(0.0099)	(0.0022)	(0.0115)
Panel B: With Pollution								
	-0.0099	-0.0014	-0.0020	-0.0007	-0.0110	0.0137	-0.0021	0.0030
	(0.0069)	(0.0017)	(0.0014)	(0.0032)	(0.0119)	(0.0128)	(0.0025)	(0.0201)

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-15: Maternal Characteristics and Exposure to Wildfires in New Jersey (Gestation >= 37 weeks)

	Mother's Age [20, 24]	Mother's Age [25, 34]	Mother's Age [35, 45]	Mother's Ed: <HS Degree	Mother's Ed: HS degree	Mother's Ed: Some College
Panel A: No Pollution						
	0.0025	0.0236	-0.0261	-0.0297	-0.0070	0.0267*
	(0.0232)	(0.0168)	(0.0192)	(0.0281)	(0.0281)	(0.0144)
Panel B: With Pollution						
	-0.0263	0.0527**	-0.0264	-0.0605	-0.0445	0.0279
	(0.0363)	(0.0255)	(0.0321)	(0.0465)	(0.0356)	(0.0194)

	Mother's Ed: College	Mother's Ed: College+	Mother is Married	Mother is Black	Mother is Hispanic
Panel A: No Pollution					
	0.0068	0.0032	0.0424	-0.0256*	-0.0619
	(0.0236)	(0.0290)	(0.0408)	(0.0149)	(0.0489)
Panel B: With Pollution					
	0.0376	0.0394	0.1164**	-0.0386**	-0.1101
	(0.0336)	(0.0411)	(0.0578)	(0.0172)	(0.0818)

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-16: Maternal Behaviors and Exposure to Wildfires in New Jersey (Gestation >= 37 weeks)

	Tobacco Use	Alcohol Use	Drug Use	Weight Gain: <= 15bls	Weight Gain: <= 60bls	Weight Gain: > 60bls	Weight Gain: 0 or Missing	Adequate Prenatal Care
Panel A: No Pollution								
	0.0117	0.0012	0.0021	0.0043	-0.0113	0.0068	0.0002	0.0073
	(0.0095)	(0.0023)	(0.0020)	(0.0028)	(0.0085)	(0.0096)	(0.0024)	(0.0117)
Panel B: With Pollution								
	-0.0085	-0.0022	-0.0031*	0.0009	-0.0107	0.0128	-0.0029	0.0002
	(0.0065)	(0.0015)	(0.0015)	(0.0033)	(0.0122)	(0.0131)	(0.0027)	(0.0200)

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-17: The Effect of Exposure to Wildfire on Birth Weight (different restrictions on the severity of wildfires)

	Birth Weight	
	>= 250acres	**>= 350acres**
Ever exposed to any wildfire:		
Exposure or not (0/1)	-22.7898**	-24.5993**
	(10.4158)	(11.8444)
Individual level control variables	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes
Pollution	No	No
Preterm	Yes	Yes
County fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Obs. No.	56,560	52,635

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-18: The Effect of Exposure to Wildfire on Birth Weight (different restrictions on the severity of wildfires)

	Birth Weight	
	>= 250acres	**>= 350acres**
Ever exposed to any wildfire:		
Exposure or not (0/1)	-50.2416**	-59.5918**
	(18.6565)	(22.8325)
Individual level control variables	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes
Pollution	Yes	Yes
Preterm	Yes	Yes
County fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Obs. No.	44,754	41,497

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-19: The Effect of Exposure to Wildfire on Birth Weight (different restrictions on distance, no pollution)

	Birth Weight		
	Distance <= 5miles	**Distance <= 15miles**	**Distance <= 20miles**
Ever exposed to any wildfire:			
Exposure or not (0/1)	-22.1434*	-20.9952*	-22.6482**
	(12.5790)	(11.0050)	(11.3290)
Individual level control variables	Yes	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes	Yes
Pollution	No	No	No
Preterm	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Obs. No.	46,248	83,953	90,359

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-20: The Effect of Exposure to Wildfire on Birth Weight (different restrictions on distance, with pollution)

	Birth Weight		
	Distance <= 5miles	**Distance <= 15miles**	**Distance <= 20miles**
Ever exposed to any wildfire:			
Exposure or not (0/1)	-39.7045**	-40.8144**	-41.2506**
	(16.0242)	(17.2923)	(17.2742)
Individual level control variables	Yes	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes	Yes
Pollution	Yes	Yes	Yes
Preterm	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Obs. No.	40,632	63,069	63,429

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1-21: The Effect of Exposure to Wildfire on Birth Weight (defining exposure using an alternative method)

	Birth Weight (in grams)	
	No Pollution	**With Pollution**
Ever exposed to any wildfire:		
Exposure or not (0/1)	-19.4586*	-34.3843**
	(10.0015)	(16.7128)
Individual level control variables	Yes	Yes
Pre-existing risks for this pregnancy	Yes	Yes
Maternal behaviors during pregnancy	Yes	Yes
Pollution	Yes	Yes
Preterm	No	Yes
County fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Obs. No.	72,769	59,913

Notes: Each column in each panel is a separate OLS regression. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

CHAPTER 2:

The Impact of Access to Restaurants on Maternal Weight Gain and Infant Birth Outcomes

2.1. Introduction

The importance of infant health at birth has been widely studied and proved to have significant long-term effect on health outcomes, educational attainment, adult earnings and so forth. Purposed by Barker in 1986, fetal origins hypothesis states that how well the fetus develops in *utero* might have significant effect on their developmental health conditions and their wellbeing at the later stage of life. And such effect could impact not only infancy health in the short-run, but also the long-run effects in their childhood and adulthood. In epidemiology, a widely accepted explanation for such effect is the fetal programing process, which could be altered by a stimulus or insult during fetus vulnerable developmental period and cause a long-lasting or permanent effect on fetal development. And such effect will further affect infants' birth outcomes. Various factors have been studied and some of them have been proved as having intensive effect on birth outcomes, among which, nutrition received by fetus *in utero* during mothers' pregnancy has been emphasized as one of the most important factors in both economic and epidemiological literature.

Convincing evidences (Barker 1992, 1993) exist and suggest that perinatal undernutrition during pregnancy might cause adverse effect on infant health at birth, including low birth weight, complications at birth, diabetes and cardiovascular disease in adulthood. As obesity becomes a prevalent problem that threatens both children's and adults' health, a variant of the original fetal origins hypothesis has been proposed, which suggests that not only under-nutrition, over-nutrition during pregnancy can also cause adverse effects on infant health at birth[29]. According to this hypothesis, excessive maternal bodyweight (i.e. mother gain overweight during pregnancy) might

[29] Barker (2007); Erikson et al. (2001); Oken and Gillman (2003); Pettitt and Jovanic (2001); Whitaker and Dietz (1998).

change the intrauterine environment during fetal development process, leading to permanent changes in the hypothalamus, pancreatic islet cells, adipose tissue, or other biological systems that can directly or indirectly alter the development of fetus, and thus affect birth outcomes (Currie and Ludwig (2010)). The majority of literature studied the effects of undernutrition, but not enough studies investigated the potential adverse effects of over-nutrition. However, as the development of our society and the improvement of the quality of our life, having “super baby” because of over-nutrition and excessive maternal weight gain during pregnancy becomes mothers’ new concern gradually. In order to investigate the potential adverse effects of over-nutrition and excessive maternal weight gain, and let people be aware of the possible adverse effects, our study focuses on estimating the impact of access to different types of restaurants on excessive maternal weight gain during pregnancy and infant health at birth, including birth weight, Apgar score, abnormal conditions of newborn, and complications at birth.

In the debate over the causes of obesity, availability of fast-food restaurants is the one that often gets blamed as an important determinant of increasing obesity rates. Existing studies find that increasing number of restaurants in the neighborhood, especially fast-food restaurants is likely to be accompanied by higher proportion of overweight and obesity population in that region (Chou et al. 2004). Based on these studies, there is another group of researchers investigated the impact of restaurants on maternal health during pregnancy, and they have found out that a nearby fast-food restaurant in the neighborhood might increase the probability for mother to gain excessive maternal weight during pregnancy (Currie et al. (2010)). Based on the fact that lots of the previous researches suggest the strong link between the availability of restaurants, especially fast-food restaurants, and excessive weight gain, our study investigates the impact of excessive maternal weight gain on a various set of birth outcomes, and we also exam the assumption that access to restaurants is one of the main exogenous determinants that could cause excessive maternal weight gain and further lead to different adverse effects on infant birth outcomes.

Our main estimation focuses on the effect of access to restaurants on maternal weight gain and infant health at birth. We first check whether easy access to restaurants (we look at fast-food restaurants and full-service restaurants separately) could cause adverse effects on infant birth outcomes. Based on the effects we have confirmed at the first step, we further check if the availability to restaurants is one of the important determinants of excessive maternal weight gain, which further could further lead to the adverse impact on infant birth outcomes. Our results suggest that increasing number of restaurants could cause the increased probability of gaining excessive maternal weight gain during pregnancy, and further cause having super-sized baby, complications at delivery, and decrease in Apgar score. In contrast, increase in grocery stores and fresh food markets might improve newborns' health condition at birth.

The rest of the paper is organized as follows. Section 2 provides background information on the effect of excessive maternal weight gain and overweight infant, and a review of the existing literature. Section 3 introduces the main data sources we are using. In section 4, we discuss the empirical strategy in details. Section 5 presents our main results, following by section 6 presenting the results of robustness checks. In the last section, we conclude based on our estimation results.

2.2. Background and Literature

Fetal origins hypothesis proposed that the in-utero environmental shocks could cause persisting effects on the developmental health conditions by altering the fetal programming process. Quite a lot attention has been addressed to the impact of under-nutrition, suggesting that perinatal under-nutrition might increase the probability of low birth weight, which could further increase the risk of diabetes and cardiovascular disease in adulthood (Barker (1993, 2005)). In the recent decade, obesity becomes one of the prevalent health concerns, which draws more attention on the causes of health issues related to obesity. A mass of studies addressed on the adverse effects of over-nutrition emerge. Both studies in epidemiology and economics have suggested that over-nutrition *in utero*, which in most of the cases presents in the form of very high birth weight, can also cause adverse effects on infant health and affect their health conditions in adulthood.

As one of the most important birth outcomes, both low birth weight and high birth weight play a crucial role in determining infant health at birth. In our study, main attention is addressed on the effect of high birth weight that is beyond the normal level. Macrosomia is a typical term, which is used to describe the condition that a newborn has an excessive birth weight. There are a few different methods used to define Fetal Macrosomia, and the two most common methods used to define macrosomia are using the threshold of birth weight being at least 4000g (8lb 13oz) or the birth weight level that is greater than 90th percentile for gestational age after correcting for neonatal sex and ethnicity (American College of Obstetricians and Gynecologists (ACOG)). Figure 2-1 shows a histogram of the birth weight distribution based on our sample. The observations with birth weights higher than 4000 grams concentrate on the right tail of the bell-shaped curve. Macrosomia, defined by the above methods in the literature, is associated with a lot potential complications for both infants and mothers. Due to the presence of macrosomia, the risks of shoulder dystocia, brachial plexus injury, skeletal injuries, meconium aspiration, prenatal asphyxia, hypoglycemia, and fetal death increase significantly (Mohammadbeigi et al. (2013)). Moreover, for the infants born with macrosomia, they have been found at a higher risk of developing diabetes mellitus, hypertension, and obesity in adulthood. Moreover, macrosomia does not only cause risk for infants, but also threatens mothers' health. Macrosomia is reported to have adverse effects on maternal health, including increasing the occurrence of oxytocin, cesarean delivery, postpartum hemorrhage, infection, 3rd- and 4th-degree perineal tears, thromboembolic events, and anesthetic accidents during pregnancy or delivery (Hermann et al. (2010)).

Two strands of literature are related to our topic. The first group of researches focuses on the effect of maternal weight gain on birth outcomes, especially birth weight. And the other group of studies investigate the effect of access to restaurants, especially fast-food restaurants, on maternal weight gain during pregnancy. Our estimation is closely related to both of these two strands and focuses on investigating the effect of maternal weight gain on a set of various birth outcomes. In addition, we want to exam if the access to restaurants, especially fast-food restaurants, is one of the channels that cause these adverse effects by increasing the probability for mother to gain excessive weight during her pregnancy.

A vast literature has studied the effect of mother's gestational weight gain on infant's birth weight. Ludwig and Currie (2010) investigate the association between pregnancy weight gain and birth weight by restricting their sample to only the multiple births that could be identified by the same mothers. And their estimation suggests that there is a consistent association between pregnancy weight gain and infant birth weight. They found that for those mothers who have gained more than 24 kilograms during pregnancy, the average birth weight of their babies is 148.9 grams heavier than the babies whose mothers have gained 8-10 kilograms during pregnancy. The associations between maternal weight gain and other birth outcomes, such as birth complications and birth defects, are also studied by other researches, which suggest that mothers had very high weight gain during pregnancy might experience higher risk of birth defects and more complications during pregnancy or at delivery (Watkins et al. (2003)).

Another group of studies have estimated the impact of availability of restaurants on weight gain in general, and children or maternal weight gain in particular. Significant literature exists supporting the hypothesis that increased availability to restaurants, especially fast-food restaurants, could increase the risk of obesity or overweight[30]. A study conducted by Chou, Grossman and Rashad (2006) employs the First, Second, and Third National Health and Nutrition Examination Surveys (NHANES I, II, and III) and investigates the determinants that could affect body mass index or obesity. The results from their study proved that access to restaurants is one of the most important reasons that could increase obesity. Another study by Currie, Vigna, Moretti and Pathania (2010) investigates the effect of fast-food restaurants on maternal weight gain and obesity in particular. Their estimation indicates that the number of fast-food restaurants within 0.5 miles of residential area has a significant effect on mothers gaining excessive weight during pregnancy.

These existing literatures provide strong support to the strategy that we will use for the estimation in this paper, since access to restaurants, especially fast-food restaurants, could significantly affect maternal weight gain during pregnancy, and therefore further adversely affect infant health and maternal health. However, we find few literature that studies how excessive maternal weight gain affect infant birth outcomes, or evaluate mother's BMI and estimate how the access to restaurants affect the probability for mother to gain excessive weight during her pregnancy. Since reasonable weight gain during pregnancy is essential to the healthy development for infants, we need to apply certain credential to define excessive maternal weight gain in order to investigate the adverse effect of unnecessary or even redundant weight gain on infant health or maternal health. The uniqueness of our data allows us to evaluate whether the maternal weight gain is excessive or not by following the suggestions of ACOG. And we have calculated an

[30] Anderson and Matsa (2011); Lhila (2011); and Dunn, Sharkey and Horel (2011).

indicator of excessive maternal weight gain by using mother's weight and height information before and after her pregnancy. In our analysis, we first investigate the effect of access to restaurants on infant birth outcomes, we then further exam how the access to restaurants affect mother's weight gain during her pregnancy based on the existence of the possible adverse effects. If there is a strong link between the ease of access to restaurants and excessive maternal weight gain, then excessive maternal weight gain might be the main channel, through which access to restaurants cause adverse effects on infant birth outcomes.

2.3. Data

2.3.1. Data on Birth Records

The most important data set that provides our estimation with a tremendous amount of restricted birth records and a large variety of variables is the Linked Patient Discharge Data and Birth Cohort File of California. We obtained this dataset from the Office of Statewide Health Planning and Development (OSHPD) of California. This data set contains birth records and all infant readmission records occurring within the first year after the infants were born. In addition, the maternal antepartum and postpartum hospital records for the nine months prior to delivery and one year post-delivery are also included in this data set. All the information about birth and hospital records are collected and paired from the following resources[31]: 1) California Patient Discharge Data; 2) Vital Statistics Birth Certificate Data; 3) Vital Statistics Death Certificate Data; 4) Vital Statistics Fetal Death File; 5) Vital Statistics Birth Cohort File; 6) Emergency Department Data; and 7) Ambulatory Surgery Center Data. We have employed all the birth records from January 2007 to December 2010, which has information of mother's pre-pregnancy weight, post-pregnancy

[31] Please check the official website for the details of data sources: https://www.oshpd.ca.gov/HID/Data_Request_Center/Types_of_Data.html.

weight, and height. This uniqueness of our data allows us to identify whether mother has gained excessive weight during her pregnancy by using her BMI before and after her pregnancy.

In order to estimate the possible effects of excessive maternal weight gain, a variable that measures whether mother gained excessive maternal weight during her pregnancy is necessary, which is determined by mother's height and mother's weight before and after her pregnancy. Our birth records data provides all the required information mentioned above for the identification of excessive maternal weight gain during pregnancy. In addition, this data set also provides various birth outcomes (such as birth weight, abnormal conditions, complications at labor/delivery, Apgar scores and so forth), hospital records (such as diagnosis codes and procedure codes) and other important information about parents and infants (such as mother's demographics, parents' education, risk of this pregnancy, maternal behaviors, and prenatal care). Our data source will allow us to exam the effects of excessive maternal weight gain on a large variety of infant birth outcomes.

The indicator of whether mother gained excessive weight during pregnancy is calculated based on the guidelines from American College of Obstetricians and Gynecologists (ACOG) and The Institute of Medicine[32]. This maternal weight gain indicator is derived according to mother's actual weight gain during pregnancy and mother's pre-pregnancy BMI[33]. Table 2-1 presents the criteria for calculating the indicator of excessive maternal weight gain, which is adopted from ACOG and The Institute of Medicine's guideline. For women who are underweight before pregnancy, the suggested weight gain range is 28-40 pounds; for women who are in the normal weight range before pregnancy, the suggested weight gain is 25-35 pounds; and for women who

[32] We have compared the guidelines for the medically suggested maternal weight gain during pregnancy provided by these two institutes. Both of them provide the consistent guidelines.
[33] Mother's pre-pregnancy BMI is calculated by the formula: $BMI{=}weight(in\ pounds){*}703/height(in\ inches)^2$.

are overweight/obese before pregnancy, the suggested weight gain is 15-25 pounds. The uniqueness of our data source provides mother's height and pre-pregnancy weight to calculate mothers' pre-pregnancy BMI, which allows us to assign a relatively accurate value to maternal weight gain indicator by strictly following the guidelines.

Our sample based on Linked Patient Discharge Data and Birth Cohort File is constructed based on the following restrictions: First, we only retain all the singleton births in our sample; second, we exclude all the records related to births from mothers who had pre-pregnancy diabetes or gestational diabetes (i.e. diabetes during this pregnancy); third, we retain all the singleton live births with gestational age greater than or equal to 37 weeks but less than 41 weeks, with non-missing information on any of the variables we will use in our models; fourth, we only keep the birth records with birth weight bounded between 500 grams and 7000 grams. Moreover, we exclude the birth records with mother's age below 20 or above 45 years old. With these restrictions imposed on the sample, we have 1,233,095 observations left in total. And Table 2-2 presents the summary statistics for all the variables used in our analysis.

2.3.2. Data on Restaurants

The number of different types of restaurants and stores in each ZIP code region are used as measurements of how ease mothers can access to restaurants (both fast-food and full-service restaurants, which usually serve high-calorie food) or stores (which usually provides more options towards healthy food, especially compared to fast-food) that can provide more relatively healthier options. And the type of restaurant is determined according to the definitions provided by the United Census Bureau. All the information about the number of restaurants and stores at ZIP code level is collected from County Business Patterns Data: Complete ZIP code Industry Detail File[34],

[34] Please find more information on: http://www.census.gov/programs-surveys/cbp/data/datasets.html.

from 2007 to 2010. Different types of restaurants and stores, including limited-service restaurant, full-service restaurant, supermarket and grocery store, meat market, fish and seafood market, and fruit and vegetable market, are classified by the North American Industry Classification System (NAICS) [35].

According to the classification of North American Industry Classification System (NAICS), there are two main types of restaurants: Limited-service restaurants and full-service restaurants. Limited-service restaurant is also call fast-food restaurants, defined as "establishments primarily engaged in providing food services (except snack and nonalcoholic beverage bars) where patrons generally order or select items and pay before eating" (NAICS 2007-2010). While, full-service restaurants are "the restaurants primarily engaged in providing food services to patrons who order food and are served while seated (i.e., waiter/waitress service) and pay after eating" (NAICS 2007-2010). Since individual's eating habit also depends on the availability to other types of stores and markets (which plays the role as the substitutes of unhealthy fast-food), we have also taken the possible effects of these stores and markets into consideration, which include: 1) super market and grocery stores, defined as stores that primarily engaged in retailing a general line of food, such as canned and frozen foods; fresh fruits and vegetables; and fresh and prepared meats, fish, and poultry; 2) meat markets, defined as the stores primarily engaged in retailing fresh, frozen, or cured meats and poultry; 3) fish market, defined as the markets engaged in retailing fresh, frozen, or cured fish and seafood products; and 4) fruit and vegetable markets, defined as the stores engaged in retailing fresh fruits and vegetables.

By using 5-digit ZIP code of mother's residential address, we are able to merge the variable about the number of restaurants and stores into the sample we have constructed above.

[35] For more information about coding of restaurants and store, please visit: http://www2.census.gov/programs-surveys/cbp/technical-documentation/reference/naics-descriptions/naics2002.txt

2.4. Empirical Specification

2.4.1. The Availability of Restaurant, Maternal Weight Gain and Infant Health

A persuasive explanation about the positive relationship between the availability of restaurants and maternal weight gain argues that: cooking is time consuming; however, getting food in a restaurant, especially taking out fast-food in a restaurant is time saving. Given the financial budget constraint families are facing, dining in a full-service restaurant might be costly so that people might do it occasionally. While taking out food from a fast-food restaurant not only save time but also relatively cheaper, compared to most of the full-service restaurants. Therefore, when the supply of fast-food restaurants increases in the residential area, the availability to the relatively cheaper and time saving food increases, which might raise the consumption of fast-food for families who do not have enough time to prepare food and also having a relatively tight budget. The more fast-food restaurants in the region, the more likely that people will visit the restaurant in the case we described above. For women who are pregnant, this positive relationship might be even stronger, since it is likely that they tend to care about their weight gain less due to their pregnancy.

However, the relation between availability of restaurant and maternal weight gain could also be negative or zero. For some of the pregnant mothers, food from restaurants is just a substitute of the unhealthy home-cooked food. Since every meal provided by the restaurants comes with certain amount or size, mothers who are taking unhealthy restaurants food as substitutes of home-cooked unhealthy food might even lose some weight if they are taking in less amount of food when they order outside.

Given these possibilities, the impact of availability of restaurant on maternal weight gain could be uncertain, and one of this paper’s goals is to investigate whether restaurant availability (especially fast-food restaurants) is responsible for the increase in the possibility of excessive maternal weight gain, which further leads to the adverse effects on infant health. In our paper, the ease of accessing to a certain type of restaurants is measured by the number of restaurants in the ZIP code region, based on mother's residential information. The higher the number is, the easier access mothers have. Both the number of limited-service and full-service restaurants are taken into consideration. In addition, we have also taken the access to different types of market and stores into consideration, because food from these stores is substitutes to the food served in restaurants. Increasing availability of these markets and stores might reduce people's inconvenience and time spent on getting materials to prepare healthy food, therefore reduces the incentive to dinning in restaurants. And this change in mother's dinning behavior will further affect her nutrition intake and therefore cause the change in infant birth outcomes.

2.4.2. Model Specification

The first step of our estimation is to investigate if easy access to restaurants could cause any adverse effect on infant health at birth, including birth weight, APGAR score, any abnormal conditions, and any complications at birth. Upon the existence of the potential adverse effects, we further investigate whether the availability of restaurants, especially fast-food restaurants, cause these adverse birth outcomes by increasing the likelihood for mothers who have easy access to restaurants to gain excessive weight during pregnancy. In other words, we will check if maternal weight gain is the channel, through which access to restaurants adversely affect infant health at birth. We will look at the potential effects of different types of restaurants and stores on infant birth outcomes and on the probability of gaining extra weight during pregnancy.

Our main results are estimated by two regression models specified as follows, which exam the effects of access to different types of restaurants and stores on various birth outcomes (i.e.

equation (1)), and the relation between access to restaurants and maternal weight gain (i.e. equation (2)). The regression models are as following:

$$outcomes_{izt} = \beta_0 + \beta_1 * N_{izt}^{fast} + \beta_2 * N_{izt}^{full} + \beta_3 * N_{izt}^{stores} + \pi'\boldsymbol{X}_i + \gamma'\boldsymbol{M}_i + \theta'\boldsymbol{P}_i + \alpha_c + \delta_t + \varepsilon_{izt} \quad (1)$$

$$wgtover_{izt} = \beta_0 + \beta_1 * N_{izt}^{fast} + \beta_2 * N_{izt}^{full} + \beta_3 * N_{izt}^{stores} + \pi'\boldsymbol{X}_i + \gamma'\boldsymbol{M}_i + \theta'\boldsymbol{P}_i + \alpha_c + \delta_t + \varepsilon_{izt} \quad (2)$$

Among all these three models (i.e. equation (1)-(2)), $wgtover_i$ is an indicator equals to 1 if mother gained excessive weight during pregnancy, which is measured by using pre-pregnancy Body Mass Index (BMI) and weight gain during pregnancy according to the guideline from ACOG and The Institute of Medicine. $outcomes_{izt}$ is a vector of different birth outcomes, including macrosomia (equals 1 if birth weight greater than 4000g), Apgar score at 1 minute, Apgar score at 5 minutes, occurrence of cesarean delivery, presence of any complications at delivery[36] and so forth.

Both the number of limited-service restaurants (fast-food restaurants) and the number of full-service restaurants are considered in the regression. Since fast-food restaurants tend to provide unhealthy food with higher calorie, however, full-service restaurants tend to provide relatively healthier food, the existence of one type restaurants might affect the effects of the other type. N_{izt}^{fast} represents total number of fast-food restaurants in each ZIP code region, and N_{izt}^{full} indicates total number of full-service restaurants in each ZIP code region. While, N_{izt}^{stores} counts the number of market and stores at ZIP code level, which include super markets and grocery store, meat stores, fish and sea food market, and fruit and vegetable markets. We anticipate the existence of these markets might have counter effects toward the effects of fast-food restaurants.

[36] There are 31 types of complications at delivery defined by OSHPD, which will be provided in the appendix.

Among the explanatory variables, $\boldsymbol{X}_i$ is a vector of variables that includes indicators of baby's sex, baby's born season, and baby's birth order. $\boldsymbol{M}_i$ is a vector of variables that contains mother's age categories, mother's race, mother's ethnicity, mother's education categories, and whether mother ever took the WIC food during her pregnancy. $\boldsymbol{P}_i$ is a vector of three variables indicating the type of payment sources for mother's prenatal care. In addition, α_c controls for county fixed effect, δ_t controls for year fixed effect, and ε_{izt} is the error term.

2.5. Results

2.5.1. Summary Statistics

Table 2-1 summarizes the sample mean and standard deviation for all the variables used in our estimation. Overall, our sample has 1,233,095 observations that have no missing values on any of the variables used in the analysis. The availability of restaurants is measured by the number of limited-service restaurants and full-service restaurants. Since the ease of access to stores that provide health food and cook materials might have confounding effect, we have controlled the access to grocery stores, meat stores, fish and seafood markets, and fruit and vegetable markets in each ZIP code region. On average, there are approximately 7 limited service restaurants and 6.5 full-service restaurants in each ZIP code region, however, the average number of other types of stores and markets stores within each ZIP code region is only 2.

The average weight of the whole sample is around 3402 grams, among which, 9% babies were born as overweight with birth weight higher than 4000 grams, and 1.2% were born as fetal Macrosomia (i.e. birth weight is greater than 4500 grams). Among all the birth records, 5% of the babies' mothers are black, and 22.1% of them are high school drop-offs. 43.7 % of mothers are taking WIC food, and 43.7% of them are using Medicaid as the source of payment for their doctor/hospital visits. According to the statistics based on the whole sample, in a typically ZIP code region, the average number of grocery stores and markets is relatively less than the number of fast-food or full-service restaurants.

In addition to the summary statistics of the whole sample, we have also checked the basic statistics for different subsamples based on race, ethnicity and education level. Tables 2-2 to 2-4 report the summary statistics for these subsamples correspondingly. For babies born by mothers who are black, their average birth weight is lower than the overall average, and there is approximately one fast-food and one full-service restaurant less in their living regions. For the other subsamples, the statistics indicate that they all have similar averages as the overall sample. We will further check whether the effects of the availability to restaurants vary on different groups of mothers as part of the results.

2.5.2. Regression Results and Robustness Check

Table 2-5 reports the estimated results based on equation (1), which measures the effects of restaurants availability on infant birth outcomes. Our results suggest that the number of restaurants and other types of stores dose have statistically significant effects on the occurrence of having super-sized babies (i.e. birthweight greater than 4000 grams), increasing the occurrence of have complications at delivery and the C-section rates. The results indicate that one more fast-food restaurant established in mother's residential neighborhood would increase the probability of having over-sized baby by 0.01 percentage point, increase the occurrence of Cesarean delivery by 0.04 percentage points, and increase the occurrence of complications at delivery by 0.11 percentage points. Comparing to fast-food restaurants, full-service restaurant might help to mitigate the risk of having adverse effects cause by fast-food restaurants. Moreover, increasing one more store (i.e. grocery stores, meat stores, fish and seafood markets, and fruit and vegetable markets), in contrast, can reduce the probability of having adverse birth outcomes.

As part of the estimation, we also conduct the same analysis on different subsamples based on race, ethnicity and mother's education level, and the results are shown in the Tables 2-6 to 2-9. Basically, the effects of restaurants availability on infant birth outcomes are consistent across different subsamples with reasonable variation as we expect. The effects of having more fast-food restaurants can worsen infant's birth outcomes, however, the magnitude of the effects varies across different subsamples. The same patterns are observed for the impact of full-service restaurants and stores that provide more healthy choices.

Since we have observed the adverse effects of fast-food restaurants, we want to further investigate whether excessive maternal weight gain is the channel that leads to these adverse effects. Therefore, as the second step of our analysis, we have tested the link between availability of restaurants and the probability of gaining excessive maternal weight gain during pregnancy. The results are reported in Table 2-10 for the whole sample and Table 2-11 to 2-14 for the subsamples. The estimation results in the second step demonstrate that more fast-food restaurants could increase the risk for mothers to gain excessive weight gain during her pregnancy, and this excessive weight gain is not recommended for both mother and baby's health. This result confirms our assumption and suggest that there is strong link between the access to restaurants and the probability of gaining excessive maternal weight gain, which further implies that increasing the probability of excessive maternal weight gain is one of the channels through which the impact of restaurants affects infant birth outcomes.

To check the robustness of our results, we have tried different specifications and estimation methods, the results are robust. We have also changed the measurements of access to restaurants (we use the number of restaurants at county level instead of Zip code level), and our results are still robust to the different measurements of restaurants at different geographic levels.

2.6. Conclusion

This study investigates the effect of access to restaurants on infant birth outcomes, using the number of restaurants in mother's residential region (identified by ZIP code) as the measure of availability of restaurants. The estimated results suggest that increasing availability of fast-food restaurants might significantly increase the risk of having worse health outcomes at birth, including occurrence of Cesarean section delivery, complications at delivery and Apgar score.

A new fast-food restaurant opened in mother's residential area might increase the probability of mother gaining overweight during pregnancy and further cause adverse effects on infant birth outcomes, while one more full-service restaurant might offset this adverse effect caused by fast-food restaurants. On the other side, a new store (i.e. super markets, grocery store, meat stores, fish and sea food market, and fruit and vegetable markets) would decrease the probability of mother gaining overweight during pregnancy and reverse the negative effects caused by fast-food restaurants by providing more healthier choices to the local families.

We have applied the same analysis method to different subgroups based on race, ethnicity and mother's education level. And the main results are consistent across different subgroups, although the magnitude and significant level might vary slightly across these subgroups.

As obesity becomes one of the health issues faced by a large population in our society, we often hear the voice that advocates to reduce the number of fast-food restaurants or to increase the tax charged on fast-food. The results in our study provide support for this voice, at the same time our results might also provide an alternative method to reduce obesity, which suggest that instead of regulating the number of fast-food restaurants in the region, increasing the number of other types of stores (such as grocery stores and fresh food market) and promote the easy access to healthier food might be another solution that could also be efficient.

2.7. References

Alviola, P. A., R. M. Nayga, M. R. Thomsen, D. Danforth, and J. Smartt, "The effect of fast-food restaurants on childhood obesity: A school level analysis," Economics & Human Biology, 12 (2014), 110-119.

Anderson, M. L., and D. A. Matsa, "Are Restaurants Really Supersizing America?," American Economic Journal-Applied Economics, 3 (2011), 152-188.

Barker, D. J. P., "Obesity and early life," Obesity Reviews, 8 (2007), 45-49.

Barker, D. J. P., P. D. Gluckman, K. M. Godfrey, J. E. Harding, J. A. Owens, and J. S. Robinson, "Fetal Nutrition and Cardiovascular-Disease in Adult Life," Lancet, 341 (1993), 938-941.

Chou, SY, JT. Liu, and C. J. Huang, "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System," Journal of Health Economics, 23(2004), 565-587.

Davis, B., and C. Carpenter, "Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity," American Journal of Public Health, 99 (2009), 505-510.

Dunn, R. A., J. R. Sharkey, and S. Horel, "The effect of fast-food availability on fast-food consumption and obesity among rural residents: An analysis by race/ethnicity," Economics & Human Biology, 10 (2012), 1-13.

Eriksson, J., T. Forsen, J. Tuomilehto, C. Osmond, and D. Barker, "Size at birth, childhood growth and obesity in adult life," International Journal of Obesity, 25 (2001), 735-740.

Hockaday, T. D. R., and C. S. Yajnik, "Free Fatty acids (FFA) and endothelial dysfunction; role of increased oxidative stress and inflammation - to: Hales CN, Barker DJP (1992) Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. Diabetologia 35 : 595-601 - Comment," Diabetologia, 46 (2003), 303-304.

Lhila, A., "Does access to fast food lead to super-sized pregnant women and whopper babies?," Economics & Human Biology, 9 (2011), 364-380.

Ludwig, D. S., and J. Currie, "The association between pregnancy weight gain and birthweight: a within-family comparison," Lancet, 376 (2010), 984-990.

Oken, E., and M. W. Gillman, "Fetal origins of obesity," Obesity Research, 11 (2003), 496-506.

Rosenheck, R., "Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk," Obesity Reviews, 9 (2008), 535-547.

Rashad, I., M. Grossman, and SY. Chou, "The Super Size of America: An Economic Estimation of Body Mass Index and Obesity in Adults," Eastern Economic Journal, 32(2006), 133-148.

Watkins, M. L., S. A. Rasmussen, M. A. Honein, L. D. Botto, and C. A. Moore, "Maternal obesity and risk for birth defects," Pediatrics, 111 (2003), 1152-1158.

Whitaker, R. C., and W. H. Dietz, "Role of the prenatal environment in the development of obesity," Journal of Pediatrics, 132 (1998), 768-776.

Figure 2-1: Birth Weight Distribution

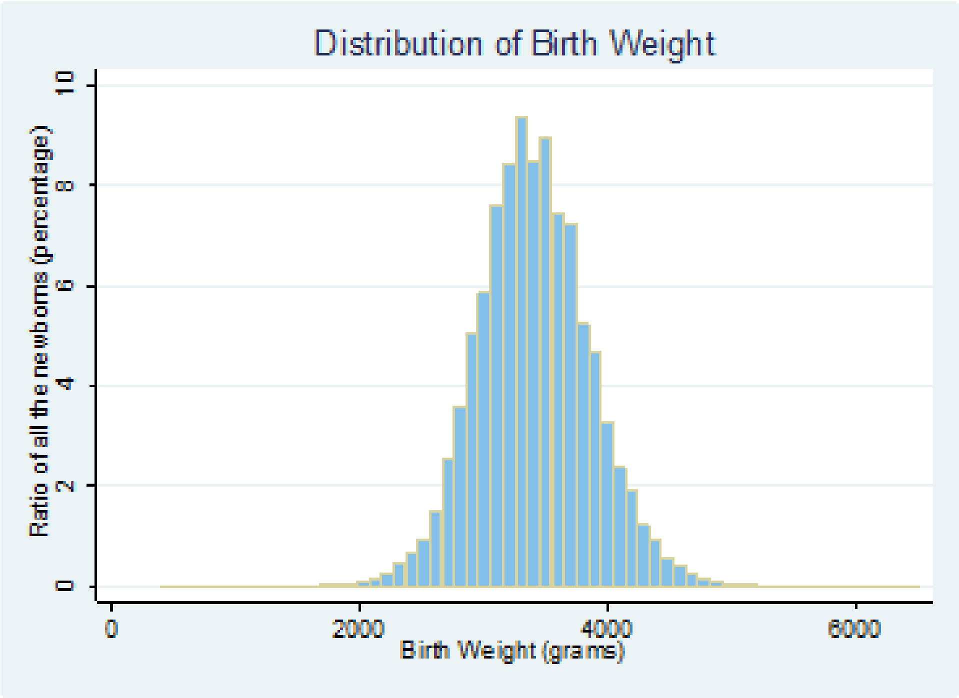


Table 2-1: Summary Statistics (whole sample)

Variable Name	Mean	N
Birth weight (measured in grams)	3,402.380	1,233,095
Fetal Macrosomia (>4000g)	0.091	1,233,095
Fetal Macrosomia (alternative) (>4500g)	0.012	1,233,095
Gestation (measured in weeks)	39.062	1,233,095
Any complications during pregnancy (1/0)	0.607	1,233,095
Any complications of newborn (1/0)	0.062	1,233,095
Any complications labor/delivery (1/0)	0.653	1,233,095
C-Section delivery (1/0)	0.314	1,233,095
Maternal weight gain (measured in pounds)	30.290	1,233,095
Excessive maternal weight gain (1/0)	0.452	1,233,095
Apgar score at 1 minute	8.257	1,233,095
Apgar score at 5 minutes	8.946	1,233,095
Child is male (1/0)	0.510	1,233,095
Child born in spring (1/0)	0.243	1,233,095
Child born in summer (1/0)	0.258	1,233,095
Child born in fall (1/0)	0.260	1,233,095
Child born in winter (1/0)	0.239	1,233,095
First child (1/0)	0.357	1,233,095
Mother is black (1/0)	0.051	1,233,095
Mother is Hispanic (1/0)	0.466	1,233,095
Mother's age 20-24 (1/0)	0.239	1,233,095
Mother's age 25-34 (1/0)	0.571	1,233,095
Mother's age 35+	0.190	1,233,095
Mother's ed: <HS (1/0)	0.221	1,233,095
Mother's ed: HS degree (1/0)	0.254	1,233,095
Mother's ed: some college (1/0)	0.245	1,233,095
Mother's ed: college (1/0)	0.183	1,233,095
Mother's ed: college+ (1/0)	0.096	1,233,095
WIC food	0.497	1,233,095
Medicaid	0.437	1,233,095
Private insurance	0.513	1,233,095
Selfpay	0.014	1,233,095
Other payment source	0.033	1,233,095
Fast food restaurants	7.246	1,233,095
Full-service restaurants	6.528	1,233,095
Grocery stores and markets	2.218	1,233,095

Table 2-2: Summary Statistics (subsample: African American)

Variable Name	Black	non-Black
	(N=62,816)	**(N=1,170,279)**
Birth weight (measured in grams)	3,296.913	3,408.041
Fetal Macrosomia (>4000g)	0.066	0.092
Fetal Macrosomia (alternative) (>4500g)	0.009	0.012
Gestation (measured in weeks)	39.012	39.064
Any complications during pregnancy (1/0)	0.615	0.606
Any complications of newborn (1/0)	0.076	0.061
Any complications labor/delivery (1/0)	0.670	0.652
C-Section delivery (1/0)	0.352	0.312
Maternal weight gain (measured in pounds)	31.927	30.202
Excessive maternal weight gain (1/0)	0.537	0.448
Apgar score at 1 minute	8.166	8.262
Apgar score at 5 minutes	8.918	8.948
Child is male (1/0)	0.510	0.510
Child born in spring (1/0)	0.236	0.243
Child born in summer (1/0)	0.255	0.258
Child born in fall (1/0)	0.262	0.260
Child born in winter (1/0)	0.246	0.239
First child (1/0)	0.352	0.357
Mother is black (1/0)	1	0
Mother is hispanic (1/0)	0.029	0.489
Mother's age 20-24 (1/0)	0.347	0.233
Mother's age 25-34 (1/0)	0.512	0.575
Mother's age 35+	0.141	0.192
Mother's ed: <HS (1/0)	0.111	0.227
Mother's ed: HS degree (1/0)	0.339	0.250
Mother's ed: some college (1/0)	0.386	0.238
Mother's ed: college (1/0)	0.112	0.187
Mother's ed: college+ (1/0)	0.052	0.099
WIC food	0.639	0.489
Medicaid	0.512	0.433
Private insurance	0.405	0.519
Selfpay	0.011	0.014
Other payment source	0.067	0.031
Fast food restaurants	6.472	7.287
Full-service restaurants	5.038	6.608
Grocery stores and markets	2.098	2.224

Table 2-3: Summary Statistics (Hispanic)

Variable Name	Hispanic (N=574,316)	non-Hispanic (N=658,779)
Birth weight (measured in grams)	3,406.683	3,398.628
Fetal Macrosomia (>4000g)	0.090	0.091
Fetal Macrosomia (alternative) (>4500g)	0.012	0.012
Gestation (measured in weeks)	39.028	39.091
Any complications during pregnancy (1/0)	0.520	0.682
Any complications of newborn (1/0)	0.052	0.070
Any complications labor/delivery (1/0)	0.573	0.722
C-Section delivery (1/0)	0.318	0.312
Maternal weight gain (measured in pounds)	28.053	32.239
Excessive maternal weight gain (1/0)	0.430	0.471
Apgar score at 1 minute	8.338	8.186
Apgar score at 5 minutes	8.959	8.935
Child is male (1/0)	0.507	0.512
Child born in spring (1/0)	0.236	0.249
Child born in summer (1/0)	0.259	0.257
Child born in fall (1/0)	0.264	0.256
Child born in winter (1/0)	0.241	0.238
First child (1/0)	0.280	0.424
Mother is black (1/0)	0.003	0.093
Mother is Hispanic (1/0)	1	0
Mother's age 20-24 (1/0)	0.311	0.176
Mother's age 25-34 (1/0)	0.546	0.593
Mother's age 35+	0.142	0.231
Mother's ed: <HS (1/0)	0.406	0.060
Mother's ed: HS degree (1/0)	0.306	0.209
Mother's ed: some college (1/0)	0.204	0.281
Mother's ed: college (1/0)	0.061	0.289
Mother's ed: college+ (1/0)	0.023	0.160
WIC food	0.745	0.280
Medicaid	0.653	0.250
Private insurance	0.303	0.697
Selfpay	0.016	0.013
Other payment source	0.026	0.039
Fast food restaurants	7.556	6.976
Full-service restaurants	5.950	7.031
Grocery stores and markets	2.542	1.935

Table 2-4: Summary Statistics (Subsample: education level)

Variable Name	Mother edu >= high school (N=960,269)	Mother edu< high school (N=272,826)
Birth weight (measured in grams)	3,403.835	3,397.257
Fetal Macrosomia (>4000g)	0.091	0.090
Fetal Macrosomia (alternative) (>4500g)	0.012	0.012
Gestation (measured in weeks)	39.079	38.999
Any complications during pregnancy (1/0)	0.635	0.505
Any complications of newborn (1/0)	0.063	0.059
Any complications labor/delivery (1/0)	0.688	0.529
C-Section delivery (1/0)	0.314	0.318
Maternal weight gain (measured in pounds)	31.249	26.912
Excessive maternal weight gain (1/0)	0.468	0.396
Apgar score at 1 minute	8.234	8.337
Apgar score at 5 minutes	8.944	8.954
Child is male (1/0)	0.512	0.504
Child born in spring (1/0)	0.245	0.236
Child born in summer (1/0)	0.258	0.259
Child born in fall (1/0)	0.260	0.261
Child born in winter (1/0)	0.238	0.244
First child (1/0)	0.407	0.183
Mother is black (1/0)	0.058	0.026
Mother is Hispanic (1/0)	0.355	0.854
Mother's age 20-24 (1/0)	0.218	0.311
Mother's age 25-34 (1/0)	0.583	0.531
Mother's age 35+	0.199	0.158
WIC food	0.386	0.885
Medicaid	0.321	0.847
Private insurance	0.627	0.115
Selfpay	0.014	0.015
Other payment source	0.037	0.018
Fast food restaurants	7.058	7.906
Full-service restaurants	6.599	6.276
Grocery stores and markets	2.040	2.842

Table 2-5: The Effects of Access to Restaurants on Infant Birth Outcomes (whole sample)

	Fetal Macrosomia	Cesarean Delivery	Complications at Delivery	Apgar Score (at 1 minute)	Apgar Score (at 5 minutes)
Fast food restaurants	0.0001**	0.0004***	0.0011***	-0.0005	-0.0002
	(0.0001)	(0.0001)	(0.0004)	(0.0005)	(0.0002)
Full-service restaurants	-0.0001***	-0.0003***	-0.0002	-0.0003	0.0002***
	(0.0000)	(0.0001)	(0.0003)	(0.0003)	(0.0001)
Stores and markets	0.0001	-0.0003	-0.0052***	0.0036**	-0.0004
	(0.0001)	(0.0003)	(0.0010)	(0.0014)	(0.0008)
Control for baby's characteristics	Yes	Yes	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes	Yes	Yes
WIC food	Yes	Yes	Yes	Yes	Yes
Payment Type	Yes	Yes	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses. The estimates are based on OLS regression.

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

Table 2-6: The Effects of Access to Restaurants on Infant Birth Outcomes (subsample: African American)

	Fetal Macrosomia	Cesarean Delivery	Complications at Delivery	Apgar Score (at 1 minute)	Apgar Score (at 5 minutes)
Fast food restaurants	-0.0002	0.0003	0.0022***	-0.0019*	-0.0002
	(0.0002)	(0.0004)	(0.0008)	(0.0011)	(0.0004)
Full-service restaurants	0.0001	-0.0000	0.0006	-0.0002	0.0003
	(0.0002)	(0.0003)	(0.0005)	(0.0009)	(0.0003)
Stores and markets	-0.0004	-0.0005	-0.0062***	0.0056**	0.0004
	(0.0004)	(0.0010)	(0.0016)	(0.0024)	(0.0008)
Control for baby's characteristics	Yes	Yes	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes	Yes	Yes
WIC food	Yes	Yes	Yes	Yes	Yes
Payment Type	Yes	Yes	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses

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

Table 2-7: The Effects of Access to Restaurants on Infant Birth Outcomes (subsample: Hispanic)

	Fetal Macrosomia	C-section Delivery	Complications at Delivery	Apgar Score (at 1 minute)	Apgar Score (at 5 minutes)
Fast food restaurants	0.0001	0.0005***	0.0005	0.0002	0.0000
	(0.0001)	(0.0002)	(0.0006)	(0.0007)	(0.0003)
Full-service restaurants	-0.0001**	-0.0003**	0.0004	-0.0013***	0.0001
	(0.0001)	(0.0001)	(0.0004)	(0.0005)	(0.0001)
Stores and markets	0.0002	-0.0003	-0.0055***	0.0050***	-0.0005
	(0.0002)	(0.0003)	(0.0011)	(0.0019)	(0.0012)
Control for baby's characteristics	Yes	Yes	Yes	Yes	Yes
Control for mother's characteristic	Yes	Yes	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes	Yes	Yes
WIC food	Yes	Yes	Yes	Yes	Yes
Payment Type	Yes	Yes	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes	Yes	Yes

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

Table 2-8: The Effects of Access to Restaurants on Infant Birth Outcomes (subsample: education level >= high school)

	Fetal Macrosomia	C-section Delivery	Complications at Delivery	Apgar Score (at 1 minute)	Apgar Score (at 5 minutes)
Fast food restaurants	0.0001**	0.0002*	0.0014***	-0.0007*	-0.0003***
	(0.0001)	(0.0001)	(0.0004)	(0.0004)	(0.0001)
Full-service restaurants	-0.0001**	-0.0003***	-0.0005*	-0.0001	0.0002***
	(0.0000)	(0.0001)	(0.0003)	(0.0003)	(0.0001)
Stores and markets	-0.0000	-0.0001	-0.0054***	0.0034***	0.0002
	(0.0001)	(0.0003)	(0.0010)	(0.0011)	(0.0003)
Control for baby's characteristics	Yes	Yes	Yes	Yes	Yes
Control for mother's characterist	Yes	Yes	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes	Yes	Yes
WIC food	Yes	Yes	Yes	Yes	Yes
Payment Type	Yes	Yes	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses

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

Table 2-9: The Effects of Access to Restaurants on Infant Birth Outcomes (subsample: education level < high school)

	Fetal Macrosomia	C-section Delivery	Complications at Delivery	Apgar Score (at 1 minute)	Apgar Score (at 5 minutes)
Fast food restaurants	0.0001	0.0008***	0.0003	0.0003	0.0003
	(0.0001)	(0.0002)	(0.0006)	(0.0008)	(0.0005)
Full-service restaurants	-0.0002***	-0.0003**	0.0002	-0.0010*	0.0002
	(0.0001)	(0.0001)	(0.0004)	(0.0005)	(0.0002)
Stores and markets	0.0002	-0.0008	-0.0039***	0.0043**	-0.0009
	(0.0002)	(0.0005)	(0.0012)	(0.0022)	(0.0016)
Control for baby's characteristics	Yes	Yes	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes	Yes	Yes
WIC food	Yes	Yes	Yes	Yes	Yes
Payment Type	Yes	Yes	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parentheses

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

Table 2-10: The effect of fast food and full-service on maternal weight gain (whole sample)

	Excessive Maternal Weight Gain (1/0)		
	(1)	(2)	(3)
Fast food restaurants	0.0005***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)
Full-service restaurants	-0.0005***	-0.0005***	-0.0005***
	(0.0001)	(0.0001)	(0.0001)
Stores and markets	-0.0005**	-0.0005*	-0.0005*
	(0.0003)	(0.0003)	(0.0003)
Control for baby's characteristics	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes
WIC food	No	Yes	Yes
Payment Type	No	No	Yes

Notes: Robust standard errors in parentheses

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

Table 2-11: The effect of fast food and full-service on maternal weight gain (subsample: African American)

	Excessive Maternal Weight Gain (1/0)		
	(1)	(2)	(3)
Fast food restaurants	-0.0004	-0.0004	-0.0004
	(0.0004)	(0.0004)	(0.0004)
Full-service restaurants	-0.0001	-0.0001	-0.0001
	(0.0003)	(0.0003)	(0.0003)
Stores and markets	0.0005	0.0004	0.0004
	(0.0007)	(0.0007)	(0.0007)
Control for baby's characteristics	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes
WIC food	No	Yes	Yes
Payment Type	No	No	Yes

Notes: Robust standard errors in parentheses

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

Table 2-12: The effect of fast food and full-service on maternal weight gain (subsample: Hispanic)

	Excessive Maternal Weight Gain (1/0)		
	(1)	(2)	(3)
Fast food restaurants	0.0006***	0.0006***	0.0006***
	(0.0001)	(0.0001)	(0.0001)
Full-service restaurants	-0.0003***	-0.0003***	-0.0003***
	(0.0001)	(0.0001)	(0.0001)
Stores and markets	-0.0010***	-0.0009***	-0.0009**
	(0.0003)	(0.0003)	(0.0003)
Control for baby's characteristics	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes
WIC food	No	Yes	Yes
Payment Type	No	No	Yes

Notes: Robust standard errors in parentheses

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

Table 2-13: The effect of fast food and full-service on maternal weight gain (subsample: education level >= high school)

	Excessive Maternal Weight Gain (1/0)		
	(1)	(2)	(3)
Fast food restaurants	0.0004***	0.0004***	0.0004***
	(0.0001)	(0.0001)	(0.0001)
Full-service restaurants	-0.0004***	-0.0004***	-0.0004***
	(0.0001)	(0.0001)	(0.0001)
Stores and markets	-0.0003	-0.0002	-0.0002
	(0.0003)	(0.0003)	(0.0003)
Control for baby's characteristics	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes
WIC food	No	Yes	Yes
Payment Type	No	No	Yes

Notes: Robust standard errors in parentheses

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

Table 2-14: The effect of fast food and full-service on maternal weight gain (subsample: education level < high school)

	Excessive Maternal Weight Gain (1/0)		
	(1)	(2)	(3)
Fast food restaurants	0.0008***	0.0008***	0.0008***
	(0.0002)	(0.0002)	(0.0002)
Full-service restaurants	-0.0006***	-0.0006***	-0.0006***
	(0.0002)	(0.0002)	(0.0002)
Stores and markets	-0.0011**	-0.0011**	-0.0010**
	(0.0004)	(0.0004)	(0.0004)
Control for baby's characteristics	Yes	Yes	Yes
Control for mother's characteristics	Yes	Yes	Yes
Control for seasonality	Yes	Yes	Yes
Control for mother's age	Yes	Yes	Yes
Control for mother's education	Yes	Yes	Yes
Control for county fixed effect	Yes	Yes	Yes
Control for birth year fixed effect	Yes	Yes	Yes
WIC food	No	Yes	Yes
Payment Type	No	No	Yes

Notes: Robust standard errors in parentheses

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

CHAPTER 3:

Environment and Emotion: An Evaluation of Air Pollution Effects based on Social Media Data

3.1. Introduction

The importance of a healthy environment to human being has been emphasized and advocated for decades. Most of the developed countries have experienced severe air pollution in the past and lots of developing countries are experiencing air pollution problem in current era. There are also ongoing debates on whether countries should trade their clean air for development or not, and if yes by how much we can afford to trade for. To answer these questions, it is necessary for us to first comprehensively understand the consequences of air pollution.

As the fundamental topic in the study of environment and air pollution, the impact of air pollution on human being has been studied by lots of researchers in various directions. It is widely accepted that air pollution can cause negative impact on physical health. Studies in epidemiology have proved that air pollution can increase the morbidity of respiratory symptoms for both adults and children (Bowman and Johnston, 2005; Moretti and Neidell, 2011), increase the risk of cardiovascular diseases, and even increase the mortality (Chen et al., 2013). There are sufficient researches in economics using different methods and approaches (e.g. air pollution caused by emissions from coal fired power plants and wildfires) to study the effects of air pollution on physical health, and they have found the consistent results as the studies conducted in epidemiology.

However, substantial adverse effects on physical health are not the only impact caused by air pollution. There is also epidemiological evidence showing that air pollution is strongly associated with people’s mental health outcomes, including but not limited to depression, anxiety and suicide thoughts. But such effects are much less studied in economics for the possible causation. One of the main reasons for the lack of studies on the mental health effects of air pollution is the difficulty of getting sufficient data that can relatively accurately measure the mental health. Even though there might be survey data available that can allow us measure individual's mental health, studies using such data might still encounter problems such as small sample size or difficulty to link the individual records with air pollution data. As mental health is crucial to individual and social well-being, correctly evaluate the cost and benefits of reducing pollution by taking the potential effects imposed on affected group's mental health is important to the long-run development of the society. Therefore, studies focusing on the causal effects of air pollution on mental health are important to fill the gap in the literature. And our paper attempts to contribute the existing literature by partially filling this gap.

Given the uniqueness of our data collected from social media (i.e. postings on Tweeter), we are able to employ a relatively large and representative sample in our analysis. By searching every single world in each posting posted by each individual in our sample, we can construct an emotion score by applying the Linguistic Inquiry and Word Count (LIWC) method. And the emotion score is a number that measures the happiness and sadness by taking the relevant happy and sad words into calculation, which help to measure emotion in a quantitative way. By using the emotion score as the dependent variable and the measurements of air pollution as one category of the explanatory variables in the analysis, we are able to quantify the effects of air pollution on people's emotional health.

Our results suggest that increase in the concentration level of sulfur dioxide or particulate matters decreases both the positive emotion scores and the polarity values (i.e.

the spread between positive and negative emotion scores) significantly, but we do not see the significant effects on negative emotion score. The results suggest that air pollution could adversely affect people's positive emotion. Our estimation provides the evidence of air pollution on people's mental health, adding the piece that help to evaluate the cost of air pollution more comprehensively.

The rest of the paper is organized as follows. Section 2 provides background information on the relationship between wildfire, air pollution and stress, and reviews the existing literature. Section 3 introduces the data sources we are using. In section 4, we discuss the empirical strategy in details. Section 5 presents our main results, while section 6 shows the results of robustness checks. The last section concludes.

3.2. Background and Literature

As we mentioned in the previous part, the effects of air pollution on physical health is already well documented, and the path through which the pollutants cause such effects is relatively clear. However, the mechanism about how air pollution impact mental health still remains unclear. Several hypotheses have been purposed to explain the effects of air pollution imposed on mental health, and the nervous system pathology is one of the most acceptable explanation. There are animal studies showing that the nervous system responds to air pollution exposure with neuro inflammatory responses, which might cause damage to the neurovascular unit, producing autoantibodies against neural and tight-junction proteins (Calderon-Garciduenas and et al., 2016; Brockmeyer et al., 2016; Block et al. 2009), and this response could further causes psychiatric symptoms such as depression and anxiety.

Using a survey data collected from 537 participants among the elderly population, Lim and et al. (2012) have found that increases in PM2.5, NO_2, and ozone may increase the depressive symptoms among the elderly. Another study conducted by Vert and et al. (2017) investigates the impact of NO_2 on the occurrence of depression symptom, and suggests that for each $10\mu g/m^3$ increase in the concentration of NO_2, the odds of depression will be doubled. In addition, there are researches estimate the impact of air quality by using the emergency department visit data, and they find that the emergency department attendance for depressive episodes were significantly higher for the particular combinations of air pollutants at certain times (e.g. NO_2 in summer). However, there are also studies that did not find any association between air pollution and depression (Wand and et al., 2014; Zijlema et al., 2016).

According to the existing literature in epidemiology, depressive symptom is not the only emotional effect air pollution is associated to. Anxiety disorder, psychosis and suicide attempts caused by the negative emotion are also closely related to air pollution. Power and et al. (2015) find that people who are exposed to higher concentration levels of particular matters are more likely to have the symptoms of anxiety. There are also other studies using various data and methods investigating the relation of severity of air pollution and occurrence of suicide attempts, most of which suggest the positive correlation between suicide attempts and air pollution (Yackerson and et. Al., 2014; Szyszkowicz and et. al. 2010; Lin and et. al., 2016).

Fast development of the economy brings tremendous benefits to us. However, it also brings negative side effects at the same time. Economic development creates a more competitive living environment, which increases the stress or other negative emotions faced by individuals in their daily life. Meanwhile, the rapid urbanization reduces the coverage of green land, and increases human caused air pollution, which could further worsen the negative emotion we are facing through the channels we have discussed above. Therefore, study of the casual effects of air pollution on emotions is very important for us to understand the cost of air pollution and to provide better evaluations on all the potential impacts caused by air pollution, and give more accurate suggestions to the policy makers.

As an alternative of the direct measurement of emotion, the transparent text analysis program - Linguistic Inquiry and Word Count (LIWC) - provides an option to indirectly measure individual’s emotion by collecting and analyze the words in the postings on social media platform. By adopting this method on the tweets data we have, we are able to employ a relative large and representative sample to estimate the potential effects of air pollution on emotion, including both positive and negative emotion. Our study contributes the existing literature by filling the gap of lacking causal studies and by providing evidence of the causal effects of air pollution on emotion.

3.3. Data

3.3.1. Data on Tweets Records

The tweets data used in our analysis is obtained from Gnip, Inc., a social medial application programming Interface (API) aggregation company, owned by Twitter. Each tweet is the actual posting online collected from different account owner in the state of Pennsylvania on one of the 63 randomly selected days from 2012 to 2013. To be more representative, each of the 63 days is randomly chosen from the 1st to 10th, 11th to 20th, and

21st to the end of each month during the year. For the details of the selection of these days, please refer to Table 3-7. All the tweets are posted and recognized in English.

The emotion index is constructed by using the transparent text analysis program - Linguistic Inquiry and Word Count (LIWC)[37]. The LIWC system has two central features - the processing component and a set of built-in dictionaries. The processing feature would go through every single word in the content and compare each of the words with the words in the dictionaries in the system, which contain a pool of various words that people could use to describe different categories of emotion. By comparing the actual words (referred as target word) with the words in the dictionary (referred as dictionary word), the processing feature can help us identify and classify the words into different psychologically-relevant categories. The dictionary version we use in the processing step is LIWC2015, which is composed of approximately 6400 words, word stems and selected emotions, and the LIWC2015 emotion categories are designed hierarchically. The processing module first reads and counts for all words in a given text, and then calculates and reports the percentage of total words that match each of the dictionary categories.

For instance, a Tweets post contains 100 words, we use LIWC to analyze this post and compare each word in the posting to the LIWC2015 dictionary and find out that there are 10 pronouns and 12 positive emotion words used in this post. LIWC then further converts the number of words into percentages, that means there are 10% of pronounces and 12% positive emotion words in this post. For every word in the dictionary, it might be classified into one or more emotion categories, corresponded to different scale scores. For example, the word cry is part of five word categories: Sadness, Negative Emotion, Overall

[37] A relatively comprehensive review of this method is summarized by Tausczik and Pennebaker (2010).

Affect, Verb, and Past Focus. Therefore, if there is word cry in the target post, then it will be classified into these 5 emotion categories and the scale scores of all these 5 categories will be counted towards the emotion score of this post[38].

The emotion score of each tweet i is calculated as E_i:

$$E_i = \frac{a_i^E}{a_i}$$

Where E_i is the emotion score, which could be positive or negative emotion. a_i^E Is the amount of words which are categorized as emotion E, positive or negative. And a_i is the total number of words every tweet post contains. For example, a tweet contains 5 positive emotion words among all the 200 words in the posting, which is equivalent to 2.5 percent of all the words. Therefore, the positive emotion score calculated by the LIWC will be 2.5. The higher the positive emotion score is, the more positive emotion words a posting has.

3.3.2. Data on Pollution

Our data on pollution is the public available data from Environmental Protection Agency[39] (EPA) of the United States, which provides publicly available data of the regulated pollutants that are considered harmful to public health and the environment, at hourly level. According to National Ambient Air Quality Standards[40] (NAAQS), primary standards[41] require carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and ozone to be monitored at one-hour level, and particulate matters to be monitored at 24-

[38] For the details of LWIC, please refer to the LWIC manual.
[39] Data source: http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily
[40] Please check http://www3.epa.gov/ttn/naaqs/criteria.html for details.
[41] Primary standards provide public health protection, including protecting the health of "sensitive" populations such as asthmatics, children, and the elderly.

hour level. All the hourly readings of every pollutant from each monitor located at different regions are available for download on EPA's website. We have downloaded the hourly concentrations for CO, SO_2, NO_2, ozone and particulate matters (PM2.5) from EPA's website. Among these five pollutants we are using, CO and ozone are measured in parts per million (ppm), SO_2 and NO_2 are measured in parts per billion (ppb), and PM2.5 is measured in micrograms per cubic meter based on the local condition[42] (LC) using the Federal Reference Methods[43].

We intend to investigate whether pollution can cause negative effects on people's emotion. In our model, we estimate the effects of concentration levels of CO, SO_2, NO_2, ozone and particular matter (these five pollutants are defined as harmful to human health and closely monitored by EPA) on people's emotional scores measured by the method described above. The concentration levels of these five pollutants are the arithmetic means of the weighted hourly averages at monitor-level within each Zip code region. The steps we followed to calculate the weighted hourly average are: First, we compute the arithmetic mean of all the hourly readings within a day to get the daily average for every pollutant at monitor level. Second, we pair each ZIP code with all the monitors in state of Pennsylvania and its surrounding states that have their borders adjacent to Pennsylvania, and calculate the geodetic distance from ZIP code centroid[44] to the location of each paired monitor by using the latitudes and longitudes. Third, we only keep all the monitors within 20 miles from ZIP code centroid for each ZIP code, based on the distances we have calculated. As

[42] The concentration was reported based on local temperature and pressure.
[43] For PM2.5 at local conditions, only those data validated from Federal Reference Methods, Federal Equivalent Methods, or other methods that are to be used in making NAAQS decisions are reported.
[44] Zip Code centroid data is purchased from http://www.zip-codes.com/.

the last step, we calculate the weighted hourly average of concentrations for each pollutant at ZIP code level by weighting hourly average readings from the selected monitors using the inverse distance between ZIP code centroid and monitor as the weights.

The measurement of pollution is assigned to each tweet account by using their Zip code information and the date and time (i.e. year, month, day and the hour of the day) when the tweets are posted.

3.3.3. Data on Holidays and Weekends

Since most of people tend to be happier during holiday season or during weekend, holiday and weekend might significantly affect people's emotion too. Therefore, we have used holiday and weekend indicators in our analysis to control the effects due to holidays and weekends. We define the holiday indicator according to the federal holiday schedule, which is obtained from the U.S. Office of Personal Management (OPM). Among the 63 days randomly selected in our sample, only one date is the federal holiday.

3.4. Empirical Method

In our analysis, each tweet posted online will be evaluated by the LIWC system and a corresponded emotion score will be calculated according to the words showed up in the tweet, and that emotion score can measure both the positive and negative emotion. Based on Berrios and et al. (2015) study, a mass of existing literature on sentiment study use the polarity score as one of the measurements of emotion[45]. Since it is purposed that people process positive and negative emotion in parallel, the difference/spread between positive

[45] Thelwall and et al. (2010, 2011); Stieglitz and Dang-Xuan (2013); Ferrara and Yang (2015).

and negative emotion score might be a good measurement in order to capture the overall sentiment in terms of emotion intensity. The polarity score is used to calculate the spread between positive and negative emotion scores, which is defined as the following:

$$Polarity_i = |Emotion\_Positive_i| - |Emotion\_Negative_i|$$

where i indicates each tweet posted online. According to the words used in each tweet, a positive emotion score and a negative emotion score will be constructed by evaluating each word used in the posting, and the polarity score is calculated for each posting using the emotion scores and the formula above.

The following models are used to estimate the impact of air pollution on the emotion scores and polarity scores calculated based on tweets contents:

$$Emotion_{p_{ijzt}} = \beta_0 + \beta_1' * \boldsymbol{Pollution_{ijzt}} + \beta_2' * \boldsymbol{Account_{ijz}} + \beta_3 * Holiday_t$$
$$+ \beta_4' * Y_t + \beta_5' * M_t + \beta_6' * H_t + \alpha_z + \varepsilon_{ijzt} \quad (1)$$

$$Emotion\_n_{ijzt}$$
$$= \beta_0 + \beta_1' * \boldsymbol{Pollution_{ijzt}} + \beta_2' * \boldsymbol{Account_{ijz}} + \beta_3 * Holiday_t$$
$$+ \beta_4' * Y_t + \beta_5' * M_t + \beta_6' * H_t + \alpha_z + \varepsilon_{ijzt} \quad (2)$$

$$Polarity_{ijzt} = \beta_0 + \beta_1' * \boldsymbol{Pollution_{ijzt}} + \beta_2' * \boldsymbol{Account_{ijz}} + \beta_3 * Holiday_t + \beta_4' * Y_t$$
$$+ \beta_5' * M_t + \beta_6' * H_t + \alpha_z + \varepsilon_{ijzt} \quad (3)$$

Where i indicates each tweet posting, j indicates every user account, z indicates the Zip code region where the account registered for, t indicates the time when the tweet is posted. Therefore, $Emotion\_p_{ijzt}$ represents the positive emotion score of tweet i, which is posted by account j that is in Zip code area z, at time t. $Emotion\_n_{ijzt}$ represents the negative emotion score, and $Polarity_{ijzt}$ represents the spread between positive and negative emotion score for each tweet i. $Holiday_t$ and Day_t controls the effects of federal holiday and the day of the week (i.e. Monday to Sunday). In addition, we have also controlled year fixed effects (Y_t), month fixed effect (M_t), hour fixed effect (H_t) and Zip code fixed effect (α_z) in our analysis. And ε_{ijzt} is the error term.

In order to estimate the effects of air pollution, we have $\boldsymbol{P_{ijzt}}$ as a vector representing measurements of pollution that each account user is exposed to during the time t when he/she posted the tweet, which include particulate matter ($PM2.5$), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and ozone. And $\boldsymbol{Account_{ijz}}$ is a vector of account characteristics, which include the number of accounts that the user is following and the number of followers account j has by time t, and these characteristics can help us indirectly identify the characteristics of the account user. The coefficients of the pollution levels are the estimated effect we focus on. Overall, we have employed three sets of regressions based on the model (1)-model (3) to estimate the effects of air pollution on different types of emotion and the spread between emotions.

3.5. Results

3.5.1. Summary Statistics

Overall, we have 1,773,087 tweets records in the sample, with no missing values for any of the variable that used in the analysis. One thing worth mention is that all these records are from 98,494 unique accounts with unique user ID. We differentiate tweets from different accounts by using the unique user ID. One account can have several postings on the same day or different days. Table 3-1 summarizes the mean of each variable we have used in our analysis.

If we pool all the tweets postings together without distinguishing accounts, the average positive emotion score among all the postings is 5.118 and the average negative emotion score is 3.495. This means that among all the postings, the average percentage of positive emotion words among all the words in all the postings is 5.118 percent, and the negative emotion words weighted 3.495 percent among all the words. The average anger value, anxious value and sad value are 1.820, 0.234, and 0.509 respectively. For each account, it has around 1199 favorites, 703 followers and it is following 562 accounts on average. Moreover, the average concentration level of Carbon Monoxide (CO) is 2.196 parts per 100,000, and it is 2.507 parts per 10,000 for ozone. The average concentration level for Nitrogen Dioxide (NO2) and Sulfur Dioxide (SO2) are 11.025 and 1.371 parts per billion respectively, and the average concentration level for PM2.5 is 9.646 micrograms per cubic meter.

3.5.2. Empirical Results

Table 3-2 reports the main results on positive emotion score. Column (1) to column (7) report the results from models with control in different sets of explanatory variables. All the estimates are based on OLS regressions. Column (7) presents the model with all the available explanatory variables in control. As the table indicates, 1 unit (measured in parts per billion) increase in Sulfur Dioxide will decrease the positive emotion score by 0.0117, which means it will increase the total positive words in a Tweet by 0.0117 percent at the 10% significance level.

Table 3-3 and Table 3-4 present the estimated effects of air pollution on negative emotion score and polarity value respectively, with the same layout of Table 3-2. Column (7) in Table 3-3 shows that none of the pollutants has significant effects on the negative emotion score. However, in Table 3-4, from column (7), we have observed the significant effects of increased concentration levels of Sulfur Dioxide and particulate matters on the polarity value, higher concentration levels of these two pollutants cause lower spread between positive and negative emotion scores. The results indicate that 1 unit increase of Sulfur Dioxide will significantly decrease the polarity value by 0.0173, while 1 unit increase of PM2.5 decreases the polarity value by 0.0049. As we mentioned in the previous part, polarity value is a proxy of overall happiness level evaluated based on each tweet. The decrease in polarity value could be caused by the decrease in positive emotion score, or increase in negative emotion score, or both. No matter which scenario causes the decrease, they all demonstrate that increase in the concentration level of these pollutants could make people feel worse emotionally. Although the magnitude of the effects might be relatively small, it is consistent with the existing literature, which uses the similar index to measure emotion.

3.5.3. Robustness Check

We have conducted several robustness checks as part of our analysis and the results are in Table 3-5 and Table 3-6. Results in Table 3-5 are based on the regression clustered at individual level, and the results are very similar as our main results. Among all the tweets, there are approximate 45% postings that have 0 positive and negative emotion scores. For these tweets with 0 emotion scores, the reason could be either the posting is actually natural that does not have any emotion in it, or the posting does have emotion word in it but the word used to describe the emotion is not in the LIWC dictionary. To exam how these zero scores could affect our estimation, we repeat the same analysis on a subsample that does not have any tweet with both positive and negative emotion score as 0. Table 3-6 reports the results based on the subsample with non-zero polarity value.

3.6. Conclusion

The data collected from Tweeter accounts allows us to construct an emotion score based on the contents of each tweet, by using the LIWC method. For each Tweet, LIWC counts the positive emotion words and negative emotion words out of the total words used in the posting, and then come up with a percentage of positive or negative emotional words as the emotion scores. Based on the positive and negative emotion scores, we can further calculate a polarity value, which is a proxy of the overall happiness level in sentiment. By merging the hourly concentration levels of the five critical pollutants with the Tweets data, we are able to estimate the effects of the change in pollutant concentration on people's emotion expressed in terms of postings through social media. The uniqueness of our data provides us the possibility to closely monitor people's emotion change at hourly level.

We have found that increase in the concentration level of sulfur dioxide or particulate matters could decrease both the positive emotion scores and the polarity values, but they do not have significant effects on negative emotion score. A possible explanation could be that people are more likely to share their happy moment and less likely to express their negative emotion through social media. And polarity value might capture the overall happiness level better. Our finding on the negative effects of these two pollutants is consistent with some of the existing literature. Although the magnitudes of our estimates are relatively small, it is reasonable based on the fact that our data is at hourly level. The emotional change caused by air pollution might not change sharply at hourly base, and that could possibly explain the relatively small magnitude of the estimated effects. However, such small magnitude could aggregate along time line if air pollution lasts for a long time, which might cause even server negative effects on people's emotion.

Our analysis contributes to the existing literature by adding a study on the causal effects of air pollution on emotion, and study is done on the hourly data set. Our estimation provides the evidence of air pollution on people's mental health, adding the piece to evaluate the cost of air pollution more comprehensively. And our results suggest that air pollution not only impact physical health, but also affect people's emotion negatively. Therefore, policy makers might need to take such impact into consideration when they are considering trading clean air for economic development. As the next step to carry this study further, we plan add weather factors (i.e. the amount of rainfall, days of snow, days of sunny and so on) in the model since weather is also purposed to be one of the important factors that could affect emotion.

3.7. References

Calderon-Garciduenas, L., E. Leray, P. Heydarpour, R. Torres-Jardon and J. Reis, “Air pollution, a rising environmental risk factor for cognition, neuroinflammation and neurodegeneration: The clinical impact on children and beyond,” Revue Neurologique, 172(1)(2016), 69-80.

Brockmeyer, S. and A. D'Angiulli, “How air pollution alters brain development: the role of neuroinflammation,” Translational Neuroscience, 7(1)(2016), 24-30.

Block, M.L. and L. Calderon-Garciduenas, “Air pollution: mechanisms of neuroinflammation and CNS disease,” Trends in Neurosciences, 32(9)(2009), 506-516.

Lim, Y.H., H. Kim, J.H. Kim, S. Bae, H.Y. Park and Y.C. Hong, “Air pollution and symptoms of depression in elderly adults,” Environmental Health Perspectives, 120(7)(2012), 1023-1028.

Lin, G-Z, L. Li, Y-F. Song, Y-X. Zhou, S-Q. Shen and C-Q. Ou, “The impact of ambient air pollution on suicide mortality: a case-crossover study in Guangzhou, China,” Environmental Health, 15(1)(2016), 90.

Power, M.C., M.A. Kioumourtzoglou, J.E. Hart, O.I. Okereke, F. Laden and M.G. Weisskopf, “The relation between past exposure to fine particulate air pollution and prevalent anxiety: observational cohort study,” BMJ (Clinical research ed), 350(2015), h1111.

Szyszkowicz, J.B. Willey, E. Grafstein, B.H. Rowe and I. Colman, “Air pollution and emergency department visits for suicide attempts in vancouver, Canada,” Environmental Health Insights, 4(2010), 79-86.

Vert, C., G. Sánchez-Benavides, D. Martínez, X. Gotsens, N. Gramunt, M. Cirach, and et al. “Effect of long-term exposure to air pollution on anxiety and depression in adults: A cross-sectional study,” International Journal of Hygiene and Environmental Health, 220(6)(2017), 1074-1080.

Wang, Y., M.N. Eliot, P. Koutrakis, A. Gryparis, J.D. Schwartz, B.A. Coull, and et al. “Ambient air pollution and depressive symptoms in older adults: results from the MOBILIZE Boston study,” Environmental Health Perspectives, 122(6)(2014), 553-558.

Yackerson, N.S., A. Zilberman, D. Todder and Z. Kaplan, “The influence of air-suspended particulate concentration on the incidence of suicide attempts and exacerbation of schizophrenia,” International Journal of Biometeorology, 58(1)(2014), 61-67.

Zijlema, W.L., K. Wolf, R. Emeny, K.H. Ladwig, A. Peters, H. Kongsgard, and et al. “The association of air pollution and depressed mood in 70,928 individuals from four European cohorts,” International Journal of Hygiene and Environmental Health, 219(2)(2016), 212-219.

Table 3-1: Summary Statistics

Variable Name	Mean	Std. Dev.
Panel A: Statistics at Tweet Level (N=1,773,087)		
Positive Emotion Score	5.118	8.742
Negative Emotion Score	3.495	7.582
Anger Value	1.820	5.585
Anxious Value	0.234	1.882
Sad Value	0.509	2.758
Carbon Monoxide (CO)	2.196	1.492
Nitrogen Dioxide (NO2)	11.205	7.63
Sulfur Dioxide (SO2)	1.371	1.784
Ozone	2.507	1.427
PM2.5	9.646	6.057
Federal Holiday	0.012	0.109
Panel B: Statistics at Account Level (N=98,494)		
Account Favorites Number	1,198.994	1550.187
Account Followers Number	702.744	8124.493
Account Following Number	561.615	1793.751

Notes: there are 98,494 unique accounts.

CO and ozone are measured in parts per million (ppm), SO_2 and NO_2 are measured in parts per billion (ppb), and PM2.5 is measured in micrograms per cubic meter based on the local condition (LC) using the Federal Reference Methods.

Table 3-2: The Effect of Air Pollution on Positive Emotion Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pollutants							
CO	0.0183**	0.0178**	-0.0080	-0.0215**	-0.0231**	-0.0122	-0.0067
	(0.0089)	(0.0090)	(0.0085)	(0.0090)	(0.0095)	(0.0093)	(0.0095)
SO2	0.0044	0.0045	0.0027	-0.0069	-0.0047	-0.0104*	-0.0117*
	(0.0059)	(0.0060)	(0.0059)	(0.0058)	(0.0060)	(0.0060)	(0.0061)
NO2	-0.0049***	-0.0047**	-0.0026	-0.0036*	-0.0036*	0.0014	0.0016
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0018)	(0.0019)	(0.0019)
Ozone	-0.0190**	-0.0183*	-0.0210**	-0.0260***	-0.0252**	-0.0129	-0.0124
	(0.0096)	(0.0096)	(0.0093)	(0.0099)	(0.0103)	(0.0135)	(0.0135)
PM2.5	-0.0017	-0.0016	0.0004	0.0006	0.0004	-0.0024	-0.0033*
	(0.0017)	(0.0017)	(0.0017)	(0.0016)	(0.0018)	(0.0018)	(0.0018)
ZIPcode fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Federal holiday	No	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Month fixed efffects	No	No	No	Yes	Yes	Yes	Yes
Day fixed effects	No	No	No	No	Yes	Yes	Yes
Hour fixed effects	No	No	No	No	No	Yes	Yes
Account characteristics	No	No	No	No	No	No	Yes
Observations	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087

Notes: Each column in each panel is a separate OLS regression. The sample covers unique 98,494 Tweeter accounts.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 3-3: The Effect of Air Pollution on Negative Emotion Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pollutants							
CO	-0.0416***	-0.0408***	-0.0167**	-0.0128*	-0.0109	-0.0012	-0.0004
	(0.0074)	(0.0073)	(0.0075)	(0.0075)	(0.0075)	(0.0073)	(0.0073)
SO2	0.0078*	0.0076*	0.0093**	0.0095**	0.0099**	0.0061	0.0056
	(0.0044)	(0.0044)	(0.0044)	(0.0043)	(0.0045)	(0.0044)	(0.0045)
NO2	-0.0074***	-0.0077***	-0.0097***	-0.0089***	-0.0101***	-0.0025	-0.0023
	(0.0016)	(0.0016)	(0.0017)	(0.0017)	(0.0018)	(0.0018)	(0.0018)
Ozone	-0.1380***	-0.1391***	-0.1367***	-0.1299***	-0.1447***	-0.0151*	-0.0134
	(0.0072)	(0.0072)	(0.0072)	(0.0086)	(0.0091)	(0.0090)	(0.0091)
PM2.5	0.0051***	0.0051***	0.0032**	0.0023*	0.0031**	0.0017	0.0017
	(0.0013)	(0.0013)	(0.0012)	(0.0013)	(0.0015)	(0.0015)	(0.0015)
ZIPcode fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Federal holiday	No	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Month fixed efffects	No	No	No	Yes	Yes	Yes	Yes
Day fixed effects	No	No	No	No	Yes	Yes	Yes
Hour fixed effects	No	No	No	No	No	Yes	Yes
Account characteristics	No	No	No	No	No	No	Yes
Observations	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087

Notes: Each column in each panel is a separate OLS regression. The sample covers unique 98,494 Tweeter accounts.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 3-4: The Effect of Air Pollution on Polarity Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pollutants							
CO	0.0598***	0.0586***	0.0087	-0.0086	-0.0122	-0.0110	-0.0063
	(0.0116)	(0.0117)	(0.0113)	(0.0118)	(0.0122)	(0.0120)	(0.0122)
SO2	-0.0034	-0.0031	-0.0066	-0.0164**	-0.0147*	-0.0165**	-0.0173**
	(0.0074)	(0.0075)	(0.0074)	(0.0072)	(0.0075)	(0.0075)	(0.0077)
NO2	0.0025	0.0030	0.0071**	0.0054*	0.0065**	0.0040	0.0039
	(0.0026)	(0.0027)	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
Ozone	0.1190***	0.1208***	0.1157***	0.1039***	0.1195***	0.0022	0.0010
	(0.0125)	(0.0123)	(0.0120)	(0.0132)	(0.0136)	(0.0165)	(0.0167)
PM2.5	-0.0068***	-0.0067***	-0.0027	-0.0018	-0.0027	-0.0042	-0.0049*
	(0.0023)	(0.0023)	(0.0021)	(0.0022)	(0.0025)	(0.0025)	(0.0026)
ZIPcode fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Federal holiday	No	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Month fixed efffects	No	No	No	Yes	Yes	Yes	Yes
Day fixed effects	No	No	No	No	Yes	Yes	Yes
Hour fixed effects	No	No	No	No	No	Yes	Yes
Account characteristics	No	No	No	No	No	No	Yes
Observations	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087	1,773,087

Notes: Each column in each panel is a separate OLS regression. The sample covers unique 98,494 Tweeter accounts.
*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 3-5: The Effect of Air Pollution on Emotion Scores (alternative estimation)

	Positive Scores	Negative Scores	Polarity
Pollutants			
CO	-0.0067	-0.0004	-0.0063
	(0.0093)	(0.0070)	(0.0124)
SO2	-0.0117**	0.0056	-0.0173**
	(0.0058)	(0.0051)	(0.0083)
NO2	0.0016	-0.0023	0.0039
	(0.0021)	(0.0016)	(0.0028)
Ozone	-0.0124	-0.0134	0.0010
	(0.0124)	(0.0088)	(0.0160)
PM2.5	-0.0033	0.0017	-0.0049*
	(0.0020)	(0.0015)	(0.0027)
ZIP code fixed effects	Yes	Yes	Yes
Federal holiday	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes
Account characteristics	Yes	Yes	Yes
Observations	1,773,087	1,773,087	1,773,087

Notes: Each column in each panel is a separate OLS regression. The sample covers unique 98,494 Tweeter accounts.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 3-6: Subsample with Non-zero Emotion Scores

	Positive Scores	Negative Scores	Polarity
Pollutants			
CO	-0.0050	0.0089	-0.0139
	(0.0134)	(0.0124)	(0.0213)
SO2	-0.0072	0.0195***	-0.0267**
	(0.0087)	(0.0070)	(0.0135)
NO2	-0.0003	-0.0064**	0.0061
	(0.0030)	(0.0030)	(0.0051)
Ozone	-0.0154	-0.0191	0.0036
	(0.0190)	(0.0160)	(0.0286)
PM2.5	-0.0067**	0.0023	-0.0090**
	(0.0028)	(0.0028)	(0.0045)
ZIP code fixed effects	Yes	Yes	Yes
Federal holiday	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes
Hour fixed effects	Yes	Yes	Yes
Account characteristics	Yes	Yes	Yes
Observations	971,431	971,431	971,431

Notes: Each column in each panel is a separate OLS regression. The sample covers unique 98,494 Tweeter accounts.

*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 3-7: Random Selected Days

Year	Month	Day1	Day2	Day3
2012	1	2	17	27
2012	2	7	19	26
2012	3	1	13	28
2012	4	6	17	26
2012	5	3	20	24
2012	6	9	13	23
2012	7	2	13	24
2012	8	4	12	21
2012	9	3	19	26
2012	10	10	19	21
2012	11	5	13	24
2012	12	6	12	29
2013	1	3	13	31
2013	2	9	16	26
2013	3	4	17	24
2013	4	6	17	25
2013	5	3	18	21
2013	6	4	18	26
2013	7	2	16	26
2013	8	8	12	30
2013	9	9	11	24
2013	10	1	13	28
2013	11	4	16	21
2013	12	6	12	30

Li Zeng

Contact Information

Department of Economics, College of Business and Economics, Lehigh University
621 Taylor Street, Bethlehem, PA 18015
Email: `lizeng212@gmail.com`
Website: `www.li-zeng.com`

Research and Teaching Interests

Primary:	Health Economics, Applied Econometrics
Secondary:	Environmental Economics, Industrial Organization, Labor Economics

Dissertation

Title:	The Effects of Environmental Factors on Maternal Health, Infant Health and Mental Health
Committee:	Shin-Yi Chou (Co-Chair), Muzhe Yang (Co-Chair), Seth Richards-Shubik, Cheng Chen

Education

2012–2018	Ph.D., Economics, Lehigh University
2009–2010	M.A., Economics, University of Akron

Research and Teaching Experience

2017–2018	Adjunct Professor Kutztown University, Department of Business Administration
2012–2017	Teaching Assistant Lehigh University, Department of Economics
2015–2016	Research Assistant Lehigh University, Department of Economics
2016 Summer	Instructor Lehigh University, Department of Economics
2015 Summer	Instructor Lehigh University, Department of Economics

2015 Summer	Research Assistant Lehigh University, Department of Economics
2014 Summer	Research Assistant Lehigh University, Department of Economics
2011–2012	Adjunct Professor Kent State University, Department of Mathematical Sciences
2010–2011	Research Assistant University of Akron, Department of Economics
2009–2011	Teaching Assistant University of Akron, Department of Economics

Working Papers

1. Li Zeng, "Wildfires and Infant Health at Birth: Evidence from New Jersey."
2. Li Zeng, "The Impact of Access to Restaurants on Maternal Weight Gain and Infant Birth Outcomes."

Work in Progress

1. "The Effects of Provider Payment Reductions on Infant Health Outcomes: Evidence from California"
2. "The Impact of Maternal Employment on Child Well-Being" with Shin-Yi Chou and Katia Ponomareva
3. "Environment and Emotion: An Evaluation of Air Pollution Effects based on Social Media Data" with Shin-Yi Chou and Xuan Li
4. "The Association Between Length of Study and Adolescent Mental Health" with Xiao Meng and Jie Peng

Teaching Activities

Courses Taught at Kutztown University

Introduction to Economics: Fall 2017

Principle of Macroeconomics: Fall 2017, Spring 2018

Courses Taught at Lehigh University

Principles of Economics: Summer 2016

Money, Banking, and Financial Markets: Summer 2015

Courses Taught at Kent State University

Modeling Algebra: Spring 2012

Algebra for Calculus: Fall 2011

Teaching Assistant (TA)/Graduate Assistant (GA) Activities

Recitations Taught as TA at Lehigh University

Money, Banking, and Financial Markets:

Spring 2014, Fall 2014, Spring 2015, Fall 2016, Spring 2017

Principles of Economics:

Fall 2012, Spring 2013, Fall 2013

GA for Courses

Applied Microeconomic Analysis: Summer 2014, Summer 2017 (Lehigh University)

Statistical Methods: Summer 2013 (Lehigh University)

Microeconomic Theory I (graduate): Spring 2011 (U Akron)

Statistics for Econometrics (graduate): Fall 2010 (U Akron)

Applications of Mathematical Models to Economics (graduate): Fall 2010 (U Akron)

Principles of Economics: Fall 2009, Spring 2010 (U Akron)

Professional Activities

Conference Presentations

1. The 6th Biennial Conference of the American Society of Health Economists: "Wildfires and Infant Health at Birth: Evidence from New Jersey", Philadelphia, PA, June 2016.
2. Eastern Economic Association Annual Meeting: "The Impact of Maternal Weight Gain during Pregnancy on Birth Outcomes", Washington, D.C., February 2016.
3. Eastern Economic Association Annual Meeting: "Wildfires and Infant Health at Birth: Evidence from New Jersey", New York, NY, February 2015.

Invited Presentations

1. Rowan University (Department of Political Science and Economics) "Wildfires and Infant Health at Birth: Evidence from New Jersey", Glassboro, NJ, January 2017.

Other Conference Activities

1. Eastern Economic Association Annual Meeting, Discussant for two papers, Washington, D.C., February 2016.

2. National Bureau of Economic Research Summer Institute: Health Economics Workshop, Attendee (Invited), Cambridge, MA, July 2015.

3. Eastern Economic Association Annual Meeting, Discussant, New York, NY, February 2015.

Other Professional Activities

1. Referee for *Journal of Labor Research*

2. Eastern Economic Association Annual Meeting (2016): helped Professor Shin-Yi Chou with the organization of Health, Education and Welfare section

Professional Society

1. American Society of Health Economists
2. American Economic Association
3. Eastern Economic Association

Certificates

Teacher Development Program: Level I (Lehigh University)

Teacher Development Program: Level II (Lehigh University)

Skills

Statistics and Data: Stata, SAS, EVIEWS, SQL

Computer Software: Maple, Tableau, LaTeX, Microsoft Office

Languages: English, Chinese

References

Shin-Yi Chou — Professor and Chair of the Department of Economics, Lehigh University
Research Associate, National Bureau of Economics Research
621 Taylor Street, Bethlehem, PA 18015
syc2@lehigh.edu, (610) 758-3444

Muzhe Yang — Associate Professor, Department of Economics, Lehigh University
621 Taylor Street, Bethlehem, PA 18015
muzheyang@lehigh.edu, (610) 758-4962

Irina Panovska Assistant Professor, Department of Economics, Lehigh University
621 Taylor Street, Bethlehem, PA 18015
irp213@lehigh.edu, (610) 758-1068