Three Essays on Environmental Economics: Effects of Air Pollution on Health and Human Capital

by

Sajad Vahedi

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2017 by the
Graduate Supervisory Committee:

Daniel Silverman, Chair
Matthew Wiswall
Nicolai Kuminoff

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

This dissertation consists of three chapters. Chapter one examines whether spending different amount of time outdoors on weekends and weekdays change the estimates of the impact of ground level ozone on the incidents of respiratory disease and asthma in California. This chapter contributes to the literature that focuses on the short term effect of air pollution on public health. Using the American Time Use Survey data, I find that on average people spend 50 min outdoors on weekends more than weekdays. Incorporating this difference in estimating the health impact of ozone changes the results significantly, especially for adults 20-64. The specification also allows me to find a precise estimate for each day of the week.

In chapter two I estimate the effect of exposure to ozone on skills of children aged 3 to 15 years. I use the Letter-Word (LW) test scores from the Panel Study of Income Dynamics (PSID) as a measure of children's skills. Due to omitted variable bias, OLS estimate of ozone effect on children's skill is positive and imprecisely estimated. To mitigate the omitted variable bias I use the instrumental variables approach. This method accounts for endogeneity of pollution. The effect of ozone on children's skills becomes negative but only marginally significant.

In chapter three, I estimate a production function of skill formation for children 3 to 15 years old and simultaneously account for their childhood exposure to ozone. I find that a one standard deviation increase in ozone leads to a 0.07 standard deviation reduction in the LW test scores on average. The LW test score of 3 year olds drops by 0.10 standard deviation in response to one standard deviation increase in pollution levels, while for the 14 year olds this effect is only half as much, 0.04 standard deviation. I also find that households exhibit compensatory behavior and mitigate the negative effect of pollution by investing more on their children. I quantitatively demonstrate that certain policies, such as a reduction in pollution levels or income

transfers to families, can remediate the negative impact of childhood exposure to pollution on adult outcomes.

To my family for their unconditional love and support.

To those who showed me how to be a better person.

ACKNOWLEDGMENTS

While I am listed as the sole author of this dissertation, this work would not be possible without the guidance and advice of wonderful people I have a privilege to call my mentors. The kindness and continuing support through all the years that Kerry Smith provided to me will always be an example of a true scholar who generously shared his experience and wisdom. Daniel Silverman's questions always highlighted the fundamental directions that my work should take. Matthew Wiswall was paramount in making the process of writing the dissertation exciting by constantly stimulating new ideas and discussing them with me. Nicolai Kuminoff was keen at pointing out all the details that made my work rigorous and thorough, and that I would have easily overlooked otherwise.

I am also grateful for the support and advice of many other ASU faculty members including Kelly Bishop, Alvin Murphy, Greg Veramendi, Eleanor Dillon, Kevin Reffett, Nancy Roberts, and Manjira Datta.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table | Page

LIST OF FIGURES

Figure Page

Chapter 1

ACCOUNTING FOR EXPOSURE TIME IN ESTIMATING THE EFFECT OF OZONE ON RESPIRATORY DISEASE AND ASTHMA

1.1 Introduction

People spend different amount of time outdoors based on their preferences and time constraints at different periods of time. Amount of time spent outdoors impacts the degree of exposure to pollution and, therefore, leads to a different health outcome. Quasi-experimental studies that examine the health impact of pollution usually do not have access to the information about individual's outdoor time. So controlling for the outdoor time is one of the challenges faced by the researchers who are trying to estimate the health impact of pollution.

This chapter test the hypothesis about a systematic difference between times that people spend outdoors on different days of the week. The intuition behind the hypothesis is that people have different time constraints on weekends as compared to weekdays, so they would spend different amount of time outdoors on these days. This outdoor time difference leads to a different exposure to pollution and, therefore, different health impact of pollution on weekends and weekdays.

In order to examine the effect of this difference on the estimate of the health impact of pollution, I set up the analysis in this chapter in two stages. In the first stage I develop a two constraints model to explain individuals' decision of outdoor leisure time on weekends and weekdays. In order to test the results of this model with observational data, I test the difference between individuals' weekend and weekday outdoor time using the American Time Use Survey (ATUS) data.

In the second stage I use a nonpublic version of daily hospitalization data in California that is accessible from the Office of Statewide Health Planning and Development (OSHPD). For any given zip code, I obtain the daily hospitalization rates from the OSHPD dataset; the pollution and meteorological data comes from the California Air Resource Board and the National Climatic Data Center, respectively. Using the variation of pollution and hospital admissions over time and zip codes, I can estimate the health impact of ozone. I examine the effect of incorporating the outdoor time difference in estimating the health impact of ozone. In order to estimate the health impact of ozone I use a conventional model that has been employed in the literature and add the interaction of the ozone level and the weekend fixed effect. The interaction term allows me to control for different exposure time between weekends and weekdays in order to obtain a better estimate of the biological effect of ozone.

The results of the first stage indicate that people spend significantly more time outdoors on weekends than weekdays. The magnitude of this difference on average is around 50 minutes that is large as compare to the mean, 80 minutes, and median, 45 minutes, outdoor time in the dataset. I call this difference the Weekend Effect throughout the study. [1]

The second set of results confirms that incorporating the weekend effect significantly changes the estimate of the health impact of ozone and the estimation gives a precise estimate for each day of the week. This change is consistent with the intuition and more noticeable for adults ages 20-64.

The weekend effect is interesting in own right. First, it shows how individuals' time constraints and preferences affect their time allocation among different types of

[1] It is different than the weekend ozone effect that illustrate the systematic difference between ozone pollution levels on weekends as compared to weekdays in some region in California.

activities. Second, it introduces a method to partially correct the measurement error regarding pollution measure. This is an important issue in estimating the health impact of any pollution when the exposure to pollution is influenced by time spent outdoors.

1.2 Prior Literature

An ultimate goal of many public regulations regarding emissions and pollution is the enhancement of public health. To evaluate these policies we need to estimate cost and benefits of the policies, but there are some difficulties associated with this estimation. Not only it's hard to map the regulations into improvement in emissions, pollution, and eventually public health, but it is also difficult to translate health improvement into a dollar valuation. For example, it is not easy to estimate how much of the improvement in air quality and public health is because of the Clean Air Act (CAA). Even if we have a fairly credible estimate of this relationship, it is difficult to get a dollar valuation of improvement in health outcome. The health outcomes include birth weight, premature births, asthma hospitalizations, heart attack, change in public utility, saved lives, and so forth.

Our understanding and knowledge of health impacts of air pollution come from health science and epidemiological research. In the health science, toxicology, researchers estimate the impact of a toxin on health outcome of a group of humans or animals in a controlled setting. For instance, a group of people is exposed to a pollutant and their pulmonary function is recorded to estimate the dose-response of human to ozone pollution. There are some shortcomings of using this method in policy making context. Ethical concerns over using human subjects in this type of research is one of the issues. At the same time, if the experiment is conducted on animals, then using the research results for human population is problematic. Even if the subjects are humans, generalizing the results to vulnerable groups such as children and elderly that are not usually in the treatment group is not clear.

Based on studies in toxicological literature, the dose-response function is not linear and there are disproportionately greater responses from higher level of pollution, Lefohn *et al.* (2010). [2] It is also not possible to simulate the real world pollution level and its variation, e.g. it is harmful to expose human subjects to high level of pollution in the lab. Therefore, it is not straightforward to map the results from the lab setting to the real world.

However, epidemiological research explains the relationship between pollution and public health using real world observations. What is missing from this literature is human behavior and responses. The main difference between economic and epidemiological models of environmental health, is that the economic models consider economic agent's behavior and they typically do quasi-experimental analysis. People based on their preferences, information, and constraints decide how much to work, how much to spend on consumption goods, where to live, how to avoid negative externalities such as pollution, and how much to spend on medical care to prevent disease or treat their health problems. In the economic literature, researchers' aim is to account for any of these behaviors that affect the association between pollution and health outcomes in order to estimate the health impact of pollution.

[2]Dose-response can be viewed as a damage function that maps individual's exposure to a particular contamination to a health problem, Zivina and Neidell (2013). To estimate the health impact of pollution it is important to choose an appropriate dose-response function. Pollutants may have a temporary and contemporaneous impact, but also long lasting impacts. The damage function can also differ based on age and health status of a person.

1.2.1 Health Impact of Pollution

In order to study the health impact of a pollutant using quasi-experimental analysis, the ideal data would have: an accurate measurement of health outcome, exact measure of contamination level that an individual has contact with, the duration of an individual's exposure to the contamination, individual's characteristics, actions that an individual has taken to avoid the pollution impacts, and other environmental factors that affect individual's health. [3] Availability of each of this information can affect the estimate of the health impact of pollution or the interpretation of the estimate. There are five sources of bias in the estimate of the health impact of pollution: (i) exposure measure; (ii) avoidance behavior; (iii) assigning pollution to individuals; (iv) residential sorting; and (v) environmental confounding. In order to clarify the effect of each of these sources, assume the following relationship between the health outcomes and pollution:

$$Health_{i,t} = \alpha x_{i,t} + \nu_{i,t}, \tag{1.1}$$

where

$$\nu_{i,t} = A_{i,t} + I_{i,t} + f(W_{i,t}) + g(t) + \epsilon_{i,t}$$

In equation (1.1), i and t represent individual and time. $Health$ is the measure of the health outcomes and x is the pollution level that an individual has been exposed. The error term, ν, includes avoidance behavior in response to pollution in order to avert its negative impacts, A, individuals physical and socioeconomic characteristics, I, environmental conditions such as other pollutants and weather conditions, $f(W)$, seasonal effects, $g(t)$, and random factors, $\epsilon_{i,t}$. Equation (1.1) is a comprehensive

[3] Contamination is the amount of a toxic materials in a particular location and time. These materials can be in the air, water, or soil. E.g. carbon monoxide, ozone, lead, particulate matter, and sulfur monoxide.

model that nests all the models that been used in the literature in order to estimate the health impacts of a pollutant of interest.

We are interested in getting an unbiased estimate of α as the biological impact of pollution on health outcome. If we use an imprecise measure of pollution, or if some of the determinants of ν are correlated with both of the health and pollution and we do not control for them, then our estimate of α will be biased. In the following five subsections I explain how each of these mechanism affect the estimate of the biological health impact of pollution.

Exposure Measure

One of the main limitations of the studies in this literature, is lack of data on actual exposure of individuals to a pollutant, pollution in equation (1.1). Therefore, in order to estimate the health impact of pollution, studies usually use pollution level of individuals' residence as a proxy for actual pollution. For example, assume that pollution level on weekends is high and individuals spend more time outdoors, but both of pollution levels and outdoor time on weekdays is low. Therefore, by using residential pollution level as pollution in equation (1.1) without any correction for weekends and weekdays, we will overestimate the health impact of pollution. In other words, using the residential pollution level for variable pollution in equation (1.1) without a proper correction leads to the measurement error in pollution estimate. Since the specification is a linear model that is widely used in the literature, this measurement error gives an inconsistent estimate of the biological health impact of pollution, Cameron and Trivedi (2005). So, the closer we get to the exposure measure, the better estimate we would get.

In this study, I want to partially correct for the measurement error of pollution in equation (1.1). Using the American Time Use Survey (ATUS) data, we can see a significant difference between the times that people spend outdoors on weekends as compared to weekdays. This outdoor time difference comes from an individual's characteristics and in general his life style, and not from his temporarily decision that is correlated with pollution level as a defensive response to that. [4] This difference leads to a different exposure to pollution on weekends and weekdays. So the identical pollution level on weekends and weekdays can lead to a different health outcome on these days. In order to account for different health impact of pollution on weekends and weekdays, in addition to using residential pollution level as pollution in equation (1.1), I also include the interaction of weekend fixed effect with the pollution level in the estimation equation. This method helps me to get a different values of α for weekends and weekdays.

To the best of my knowledge, this is the first study that uses ATUS to correct for the measurement error of pollution in order to estimate the health impact of ozone.

Assigning Pollution to Individuals

Second source of measurement error of pollution comes from the method of assigning the pollution level to an individual. Usually the pollution level of an individual's residence is considered as pollution that the individual has been exposed to. For instance, Currie *et al.* (2009) uses exact address of an individual to assign more accurate pollution level and some use a broader area such as zip code, Source Receptor Area (SRA) in Southern California, or county, Moretti and Neidell (2009), Neidell and Kinney (2010). There is a trade-off between choosing a small region versus a broader region to assign pollution level to an individual. Exact residence address gives a better estimate of the potential pollution level that an individual could be

[4] I will extend the discussion about the determinants of outdoor times on weekends and weekdays in subsection The Weekend Effect.

exposed, but loses some information if an individual travel far from his residence. However, the general preference in the literature is to use the exact address if this information is available. In this study I use the average pollution level within 20miles radius of an individual's residence zip code centroid. [5] This radius provides a good average measure of potential pollution for both large and small zip codes. Large zip codes are usually rural areas and people do not commute very far; and also small ones are more densely populated areas (e.g. Los Angeles) and people may travel between zip codes to work or for shopping, and this radius encircles multiple zip codes.

Avoidance Behavior

Avoidance behavior is defined as a precautionary action that is taken by individuals to avoid the harmful impact of pollution. The avoidance behavior can be considered as spending less time outdoors, or expenditure on equipment that protects an individual from exposing to pollution (e.g. air filter), or any other action or expenditure that avert the negative impact of pollution. This behavior is represented by variable A in equation (1.1). The idea is that people respond accordingly to variation of pollution level. [6] So variable A in equation (1.1) is correlated with pollution, and it is also correlated with health through averting the negative health impact of pollution. Therefore, not controlling for A leads to a biased estimate of α due to omitted variable.

Using the American Time Use Survey data at national level, Back *et al.* (2013) show that after controlling for climatic and geographical factors, children and old people tend to spend less time outdoors as the air quality reaches to unhealthy level. However, the authors illustrate that this behavior is not mainly driven by EPA's Air

[5]This method has also been employed in some studies that estimate the health impact of such pollutants. See, for example: Currie and Neidell (2005), Neidell (2004), and Knittel *et al.* (2011).

[6]This behavior is more prevalent among families that have children, older individuals, or someone who is vulnerable to pollution. When the pollution level is high, these families try to stay at home or take medicine, or take any necessary action.

Quality Index (AQI). Although this result does not preclude the presence of avoidance behavior of people in response to air pollution level. Neidell (2009) show that people respond to ozone forecast and smog alert that are provided by local authorities. The authors find that people decrease attendance at two major outdoor facilities [7] within the boundaries of the South Coast Air Quality Management District (SCAQMD). [8] In addition to spending less time outdoors in response to pollution level, people may take other precautionary actions to avert the harmful impacts of pollution that has not been observed in these studies due to lack of data on those actions.

Recent studies that control for avoidance behavior in two different ways: Neidell (2009), Neidell and Kinney (2010), and Neidell (2004) including information about pollution that is provided to public in the estimation model, and Moretti and Neidell (2009), Knittel *et al.* (2011), and Schlenker and Walker (2016) using an instrumental variable for the pollution. The first method is based on the fact that people respond to the information about the pollution that is self-observable (smog) or is provided by public agencies (e.g. smog alert, ozone forecast). When people receive this information, they adjust their behavior accordingly to avert harmful impacts of pollution, Zivin and Neidell (2009). Neidell (2004) was one of the first attempts to understand the association between air pollution and childhood asthma. Because of the data limitation he uses aggregate monthly data to estimate the impact of some of the air pollutants on childhood asthma. [9] Since different age groups of kids have potentially various behavioral (accomplished by their family) and biological responses to pollution, he divides the kids younger than 18 into 5

[7] The Los Angeles Zoo and Botanical Gardens, and Griffith Park Observatory.

[8] Another study by Zivina and Neidell (2014) find that people change their time allocation in response to temperature (this study does not examine the effect of pollution). They find large reduction in labor supply in industries that are more exposed to climate as temperature increases beyond 85 degree, and reduction in outdoor leisure activities for unemployed individuals

[9] O_3, CO, NO_2, and PM_{10}.

age-groups. [10] By dividing into the age groups and also including smog alert in his estimation model, he controls for avoidance behavior. [11] Few years later, Neidell (2009) and Neidell and Kinney (2010) use similar method to embed the information on pollution that is provided to public into the estimate of the health impact of pollution. The main differences with the previous study are: interacting the smog alert and ozone forecast with the pollution level, using more detailed daily data with better control for environmental confounding, and focusing on smaller study region (Southern California) that has better coverage of pollution information and is populated area with similar regional geography. [12] Before including the information variables, smog alert and ozone forecast, in the estimation, Neidell (2009) also shows that this information actually decrease attendance at two major outdoor facilities within the boundaries of the SCAQMD. In all these studies controlling for avoidance behavior does matter, and dropping this information underestimates the health impact of pollution. For example, Neidell (2009) and Neidell and Kinney (2010) show that including avoidance behavior in the estimation increases the estimate of the health impact of ozone by a factor of 1.5-2.5, and 20%-130% among different age groups, respectively.

In the second approach, studies use instrumental variable to control for measurement error that comes from avoidance behavior. For example, Moretti and Neidell (2009) use daily variation of boat traffic in two major ports of Los Angeles as an instrument for ozone pollution. The idea is that the daily variation of boat traffic is exogenous and the boat traffic neither is included in ozone forecast nor is

[10] 0-1, 1-3, 3-6, 6-12, and 12-18 years old.

[11] Not interacting the smog alert with the pollution level, but including it as an explanatory variable.

[12] Weather conditions such as maximum/minimum temperature, sun cover, humidity, wind speed, and also carbon monoxide and nitrogen dioxide.

reported by media, and also it is not observable by people to impact their daily behavioral responses. The authors show that boat traffic has major impact on ozone level (relevancy) , and is orthogonal to ozone forecast and some weather conditions (validity). [13] The idea of Schlenker and Walker (2016) and Knittel *et al.* (2011) is the same, only they use daily airport runway congestion and automobile congestion as an instrumental variable, respectively. All these studies use two stage least square (2SLS) method for their estimation. They all show that controlling for avoidance behavior using instrumental variable significantly changes the results. For example, Moretti and Neidell (2009) show that the estimate of the health impact of ozone using instrumental variable is 4 times greater than the standard OLS estimation.

Residential Sorting

The pollution levels are not randomly assigned to different residential locations; rather, people sort among various communities. For example, rich families usually live in the areas with better air quality, and they even spend more money on their health care. The opposite is true for poor people who live in more polluted areas. If we do not control for these differences, our estimate of the health impact of pollution, α, will be biased upward. In other words, there are some characteristics of individuals that are correlated with pollution and also have direct effect on $Health$ (e.g. income level) in equation (1.1). If we do not include them in equation (1.1) as variable I, the estimation will suffer from omitted variable bias. [14]

One way to overcome this issue is to collect detailed information on individual's characteristics in order to control for sorting issue, Currie and Neidell (2005) and Currie *et al.* (2009). These studies estimate the impact of pollution on infants' health.

[13] Although this impact decrease with the distance from the ports.

[14] Residential characteristics can also be seen as an individual's characteristics, while it is the same for everyone who lives in the same residence.

They have access to a rich data that includes information on infant's health status at birth, and information about the mother's characteristics such as race, education, and marital status. Including this information as variable I in equation (1.1) and other appropriate controls such as the mother's and zip code-month fixed effects can control for factors that are directly correlated with child's health and also are correlated with pollution level through the residence decision of the mother. Therefore, including this information solves the residential sorting issue due to omitted variable. The authors do not compare their results with the estimation without controlling for the individuals' characteristics. However, their results suggest a strong negative impact of carbon monoxide on infant health. Currie *et al.* (2009) also show that this negative impact is larger for older and smoking mothers.

If this kind of rich data is not available, another way of overcoming sorting issue is to use a proper instrumental variable for pollution (a conventional way in order to overcome the omitted variable bias). Ransom and Pope (1995) use the pollution levels change due to closing for a year and re-opening of steel mill in Utah Valley, where the steel mill is the major source of pollution. This change in pollution level due to closing and re-opening of steel mill is fairly exogenous, and residential sorting is a household's long run decision. So by using the instrumental variable for the openness of the steel mill, they should not worry about residential sorting issue. They use neighboring Cache Valley community as a control group that is not affected by the steel mill pollution because of the geographical characteristics. Cache valley is very similar to Utah Valley in many aspects such as demographic, weather, and housing characteristics and the only difference between them is their pollution levels. The authors find a positive and statistically significant impact of steel mill operation on respiratory hospital admission. [15]

In this study, using daily data and also zip code fixed effect, I partially control for the residential sorting issue. In the estimation, I use the daily variation in ozone pollution to explain the daily hospital admission rates, and unobserved factors that directly impact an individual's health and do not vary in daily basis, Moretti and Neidell (2009). At the same time, since residents within zip codes are fairly homogeneous in some dimensions, by including zip code fixed effect I can control for the unobserved characteristics of individuals.

Environmental Confounding

The last source of endogeneity is environmental confounding. Usually pollutants have seasonal trend, and also are correlated with some other pollutants and weather conditions. For example, ozone is correlated with carbon monoxide, nitrogen dioxide, sun light, and temperature and usually between late spring and early fall has the highest level of the year. On the other hand, the other pollutants and weather conditions impact environmental quality and therefore the public health, Deschnes and Greenstone (2011), and also health outcomes may have some seasonal pattern, Currie *et al.* (2009). These factors are $f(W)$ and $g(t)$ in equation (1.1). So dropping these variables from the estimation of the health impact of ozone leads to a biased estimate due to omitted variable issue. Fortunately, data on the environmental information is available with very high frequency. So, almost all the

[15] In order to deal with the sorting issue Currie and Walker (2009) use shock in pollution level due to introducing of electronic toll collection (E-ZPass). They use difference-in-differences estimation for the period of before and after the introduction of the policy. They show that the policy significantly improves the infants' health. Since the time period is short, this improvement cannot be explained by residential characteristics changes. However, both Ransom and Pope (1995) and Currie and Walker (2009) evaluate the health impact of the shocks, but do not provide the estimate of the biological health impact of pollution that is the focus of my study.

studies in the literature, as well as this study, control for the environmental conditions that are correlated with the pollutant of interest. To control for seasonal trend studies usually include fixed effects of month, quarter of the year, year, the weeks since the child's birth dummy in infant health studies, spline function of the proper date depending on each case, or (and) their interactions. I include day of the week and month-year fixed effects and I'll explain them more in the estimating strategy section.

In sum, first two sources of biases are due to measurement error of pollution, and the last three ones are because of omitted variables of $A, I, f(W)$, and $g(t)$ in equation (1.1). In this study, I want to partially correct the measurement error by adding the interaction of weekend fixed effect with the pollution level in the estimation equation. I also use a reasonable radius around an individual's residence to assign the pollution level to an individual to mitigate the second source of bias. In the current version of the study I do not control for the avoidance behavior. By using daily data and zip code fixed effect I partially control for the residential sorting issue. Finally, I account for the environmental confounding issue by controlling for some other pollutants, weather conditions that are correlated with ozone, and time fixed effects.

1.2.2 Outdoor Leisure Choice

In order to correct the measurement error of pollution in equation (1.1) as I explained in previous sections, I need to show the difference between outdoor leisure time on weekends and weekdays. This subsection and the next section are the explanation of the literature on recreation choice and my conceptual model of explaining outdoor leisure choice on weekends and weekdays. This will be the theoretical base for the empirical section in order to show the difference between outdoor leisure time on weekends and weekdays.

Recreation is a consumption good requiring both a money price and time. The role of time for consumption as an influence to individual's demand for market goods has been emphasized since Becker's classic paper, Becker (1965).

One way to incorporate a role for time, is to introduce a time constraint in an individual's utility maximization problem. Thus, an individual faces both the budget constraints, in monetary terms and a time budget constraint that he needs to allocate in order to consume market goods. This issue is more important when we consider time intensive goods, such as going online or visiting a national park. Goolsbee and Klenow (2006) utilize this framework in order to estimate the consumer surplus from using internet that is a time intensive good. They develop a model that assumes an individual must allocate both monetary and time resources in deciding a consumption pattern. They obtain a larger value for consumer surplus than one can calculate using only expenditure of individuals on internet service. Zivina and Neidell (2014) use a similar framework to estimate the impact of climate change on individuals' time allocation between labor, outdoor leisure, and indoor leisure. These two studies are different in a sense that they focus on different time intensive goods. But, much importantly their models are similar such that an individual jointly decides on time allocation among labor and the time intensive goods, based on their time cost and their contribution to his utility function.

In real life, most people do not have the opportunity to freely choose their time allocation among labor and leisure. Shaikh and Larson (2003) and Larson and Shaikh (2004) use a two constraints' model (money and time constraints) to estimate the welfare measures for change in recreation cost and environmental quality, and the marginal value of time, respectively. The critical assumption they have is that individuals' labor decision is not jointly chosen with leisure time, rather it is a long-term decision. So when individuals want to allocate their time on recreation, the time budget is given. This constraint helps them to use individuals' decision within given time constraint to estimate the marginal value of time.

Palmquist *et al.* (2010) and also most general model of Phaneuf (2011) use two constraints' model to demonstrate an individual's utility maximization problem, when the labor decision has been already made by an individual. The argument is that recreation activities need discrete blocks of time, and transferring time between blocks of time endowment is costly. Although, individuals' labor time is not very flexible, yet they are flexible in choosing among leisure activities and time saving products (hiring someone to mow lawns). So an individual decides how much to buy the time saving good from the market and how much to produce himself by spending his own time, i.e. exchange the money with his free time that can be allocated to leisure activities. So using revealed and stated preferences about these time saving activities they are able to estimate the shadow value of time.

In the current study, similar to the recent studies I use the two constraints' model, but I focus on an individual's weekly problem and distinguishing between weekends and weekdays due to time flexibility and constraints. An individual has already decided about his weekly labor time and he allocates his free time on different leisure activities. An individual solves his weekly problem such that he faces different time constraints on weekends and weekdays and it is costly to transfer time between these days. I also distinguish between outdoor and indoor leisure. As Palmquist *et al.* (2010) argue, leisure activities need discrete blocks of time and this can vary among activities, so people tend to move those activities that need more time to days with larger available time. Usually individuals' time is more flexible and they have more free time on weekends than on weekdays, and outdoor leisure are usually more time consuming than indoor leisure (playing golf, or basketball versus watching TV series). So we expect to see that people move their time consuming outdoor leisure to weekends as compare to weekdays. The next section is detailed explanation of the conceptual model.

1.3 Conceptual Framework of Outdoor Leisure Choice

In my model of individual behavior, and individual's utility depends on consumption good, X, outdoor leisure, l_o, and indoor leisure l_i. The individual consume these goods in period 1, weekday, and period 2, weekend. The labor hours is predetermined and exogenous for the two periods. There is money price for the consumption good and time price for both the leisure goods. The individual solve the following maximization problem:

$$V(I, p_1, q_1, p_2, q_2, T_1, T_2) = \max_{c, l_{i1}, l_{o1}, l_{i2}, l_{o2}} U(c, l_{i1}, l_{o1}, l_{i2}, l_{o2}) \tag{1.2}$$

subject to

$$c = wL = I$$

$$p_1 l_{i1} + q_1 l_{o1} = T_1$$

$$p_2 l_{i2} + q_2 l_{o2} = T_2$$

where w, L, and I are an individual's wage rate, labor time, and income, respectively. T_t is the available free time on period t. I assume a general case such that the relative time price of outdoor and indoor leisure can be different and this can vary between periods too. This means that an individual in order to watch TV or work out, spends different time between weekends and weekdays. The time price difference between indoor and outdoor leisure comes from the type of the indoor and outdoor activities. I also suppress the consumption good on two periods for notational convenience.

Maximizing the utility function subject to money and time constraints in equation (3.4) gives the solution for an individual's problem and indirect utility function $V(I, p_1, q_1, p_2, q_2, T_1, T_2)$. The interiority of solution and using the envelope theorem leads to:

$$V_{T_t} = \mu_t(I, p_1, q_1, p_2, q_2, T_1, T_2) \tag{1.3}$$

$$V_I = \lambda(I, p_1, q_1, p_2, q_2, T_1, T_2)$$

$$\rho_t = \frac{V_{T_t}}{V_I} = \frac{\mu_t}{\lambda}$$

$$l_{it} = -\frac{V_{p_t}}{V_{T_t}}$$

$$l_o t = -\frac{V_{q_t}}{V_{T_t}} \qquad t = 1, 2$$

Where μ_t, λ, and ρ_t represent Lagrange multiplier of time constraint of period t, money budget constraint, and opportunity cost of time at period t, respectively. The subscript of indirect utility V means the derivative of V with respect to that subscript. The opportunity cost of time is directly related to shadow value of time and it varies between period 1 and 2, but it is not straightforward to see how the shadow value of time is related to the parameters of the problem. This difficulty is also true about the decision variables of leisure time that are given by Roy's identity in fourth and fifth expressions of the equation (1.3). In order to discuss the relationship between parameters and the shadow values and the decision variables I consider the following utility function:

$$U(c, l_{i1}, l_{o1}, l_{i2}, l_{o2}) = f(X) + \theta l_{i1}^{\alpha} l_{i2}^{\beta} + (1 - \theta) l_{o1}^{\alpha} l_{o2}^{\beta} \tag{1.4}$$

The form of utility function means that indoor and outdoor leisure can be substitute, and the parameter of θ define the importance of indoor leisure over outdoor leisure. The outdoor (indoor) leisure on weekends is a necessary good for outdoor (indoor) leisure on weekdays, and both are separable from consumption good. Assuming the existence of interior solution, and using the first order conditions leads to:

$$l_{i1} = T_1/(p_1 + (\frac{p_2}{q_2})^{\frac{-\beta}{\alpha+\beta-1}} p_1^{\frac{\beta-1}{\alpha+\beta-1}} q_1^{\frac{\alpha}{\alpha+\beta-1}} (\frac{1-\theta}{\theta})^{\frac{-1}{\alpha+\beta-1}}) \tag{1.5}$$

$$l_{o1} = T_1/(q_1 + (\frac{p_2}{q_2})^{\frac{\beta}{\alpha+\beta-1}} p_1^{\frac{\alpha}{\alpha+\beta-1}} q_1^{\frac{\beta-1}{\alpha+\beta-1}} (\frac{1-\theta}{\theta})^{\frac{1}{\alpha+\beta-1}})$$

By the symmetry, the similar expression can be derived for the leisure time of period 2. After imposing a few standard assumptions of $0 < \alpha\beta$, $0 < \alpha + \beta < 1$, and $0 < \theta < 1$ we can see from equation (1.5) that the outdoor (indoor) leisure is negatively related to its own time price, i.e. $\frac{\partial l_{i1}}{\partial p_1} < 0$ and $\frac{\partial l_{o1}}{\partial q_1} < 0$ and positively related to the available time endowment, $\frac{\partial l_{i1}}{\partial T_1} > 0$ and $\frac{\partial l_{o1}}{\partial T_1} > 0$, and their importance in the utility function, i.e. $\frac{\partial l_{i1}}{\partial \theta} > 0$ and $\frac{\partial l_{o1}}{\partial (1-\theta)} > 0$. These results are very intuitive and as we expected. Using first order conditions the shadow value of time for period 1 is presented in the following equation (it is similar for the period 2):

$$\mu_1 = \frac{\theta\alpha}{p_1} l_{i1}^{\alpha-1} l_{i2}^{\beta} \tag{1.6}$$

This equation together with the equation (1.5) give the relationship between the shadow value of time and the parameters. The shadow value of time at period 1 decreases as the scarcity of time goes away, i.e. $\frac{\partial \mu_1}{\partial T_1} = (\alpha - 1) constant$. So if an individual has less free time on weekdays than on weekends, then an individual's marginal value of extra time on weekdays will be higher than on weekends.

In the empirical part of this study, I analyze the outdoor leisure time difference between weekends and weekdays. Later I will explain the estimating strategy by detail, but here I want to explain the relationship between the theoretical and the empirical models. From equation (1.5), I can drive the expression for $l_{o2} - l_{o1}$:

$$
\begin{aligned}
l_{o2} - l_{o1} = \; & T_2/(q_2 + (\tfrac{p_1}{q_1})^{\frac{\alpha}{\alpha+\beta-1}} p_2^{\frac{\beta}{\alpha+\beta-1}} q_2^{\frac{\alpha-1}{\alpha+\beta-1}} (\tfrac{1-\theta}{\theta})^{\frac{1}{\alpha+\beta-1}}) \\
& -T_1/(q_1 + (\tfrac{p_2}{q_2})^{\frac{\beta}{\alpha+\beta-1}} p_1^{\frac{\alpha}{\alpha+\beta-1}} q_1^{\frac{\beta-1}{\alpha+\beta-1}} (\tfrac{1-\theta}{\theta})^{\frac{1}{\alpha+\beta-1}}) \\
l_{o2} - l_{o1} = \; & h(\alpha, \beta, \theta, T_1, T_2, p_1, q_1, p_2, q_2)
\end{aligned}
\tag{1.7}
$$

Three parameters of α, β, and θ are taste characteristics of an individual that affect the outdoor leisure choice. In the empirical section I control for these characteristics by age, sex, marital status, residential metropolitan status, family income, and household size. I also control for the labor force status (employed, unemployed) and weekly labor hours that affect time constraint T_1 and T_2. Marital status, residential metropolitan status, and household size can affect the time constraint too. If we accept the arguments of Palmquist *et al.* (2010), then we have a framework that maintains people can trade time for money by outsourcing some time consuming home maintenance tasks. In this context, wealthy people most likely to "buy" time by time saving products. So income level can also affect the time constraint. Finally, based on blocks of time needed to do every outdoor activity, an individual decides to do the activity on weekends or weekdays, and this comes from the time price of activities p_1, q_1, p_2, and q_2.

1.4 Data

1.4.1 American Time Use Survey Data

I obtain the information on individuals' activities from the American Time Use Survey (ATUS) data for the years 2005-2012. The ATUS is an annual survey of how individuals age 15 and over spend their time on various activities such as leisure, working, household activities, childcare, and sport activities. They ask subset of people who are chosen from Current Population Survey (CPS) to report their diary of their activities in one day (one person only for a single day). There is detailed information on what activity individuals did, where they did, when they did, who was with them during each activity. I also merge the ATUS with the CPS data to get more demographic information about individuals that is not available from the ATUS data.

To calculate an individual's outdoor leisure time I choose activities based on the nature of activities and the location that is performed. The list of activities that I choose as outdoor activities are provided in Table 1.1. Some of the activities such as hiking, hunting, or golfing are clearly outdoor activities. If an activity can be performed either indoors or outdoors, such as biking or running, I choose those cases that has been performed outdoors.

Table 1.4 provides summary statistics of the outdoor leisure time and the explanatory variables that are classified based on matching criteria. In Table 1.4 and all the following tables, units of outdoor time is minutes per day. The reason for presenting summary statistics of these variables is that in the estimation part, I estimate the effect of two latter variables on outdoor leisure. Table 1.4 yields some interesting insights. Older respondents spend less time outdoors. The family size of respondents who are younger than 45 years of age is very similar. But family size of people older than 45 are smaller than the other age groups, and this can be because their children do not live with them anymore. We can also see that labor force are mainly people between 26-64 ages who have the highest working time. If I drop individuals between 20-25 from the first age group and 65-70 from the last group, this difference in labor hours would be more precise. Another observation is that males spend more time outdoors and work more in comparison to females. The same is true for married people as compare to single people. [16] We do not see that

[16] In this study I considered all the people that are not currently married as single (they can be divorced, separated, never married, or widowed).

much difference between the rest of the categories, but full/part time students. Full time students spend more time outdoors and work less, and they live in larger families as compare to part time students. These are just observations from the raw data without controlling for other determinants that can explain these differences.

Table 1.6 reports summary statistics for different outdoor activities. I drop those activities from Table 1.1 that has less than 30 observations, and I ordered them in Table 1.6 based on their average time. As we can see, there is large diversity between activities based on their mean and median values, where the mean time varies from 15-245 and median time from 5-197 minutes. Order of activities based on their median and interquartile range in some cases are different than their order based on mean, but in most cases is similar. The columns 7 and 8 provide mean value of the activities on weekdays and weekends, respectively. The last column of Table 1.6 gives the percentage of the number of observation that has been observed on weekends.

The Figure 1.1 draws the relationship between percentages of the observation that has been occurred on weekends versus activities, while the activities are ordered based on their mean values. Figure 1.1 shows that activities with greater mean time tend to be done mostly on weekends. The percentage increases from 40% for less time consuming activities to 80% for more time consuming activities. [17] For example people tend to go hunting or boating on weekends that take few hours, but they can go running for 40 minutes whenever they want. Instead of mean values of activities, I use median and interquartile range and the results are similar. Figures 1.11 and 1.12 present these result.

[17] Total number of observation in the data set is almost evenly distributed between weekends and weekdays.

1.4.2 Health Outcome

The data on health outcomes is a nonpublic version of the hospitalization data that is accessible from the Office of Statewide Health Planning and Development (OSHPD) for the years 2004-2011. The data includes all the records of individuals who are discharged from hospitals in California. In the nonpublic version of the data, there is detailed information on a patient's exact date of admission, residential zip code, age, sex, major diagnostic category, and the chief cause of the admission. I obtain the respiratory diseases and the asthma admissions as a measure of health outcomes from the data which is classified based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). [18] People usually visit the hospital if the impact of pollution is very severe. Therefore, the hospital admission on respiratory disease as a measure of health outcome underestimates the health impact of ozone (even if I include other health problem such as heart attack in the number of hospitalization).

Since the health impact of ozone differs among age groups, to calculate the dependent variable I divide the dataset into three age groups: 5-19, 20-64, and 64+. I drop the children younger than five, because it is difficult to clearly diagnose asthma at this age, Neidell (2009). Table 1.8 gives the summary on the respiratory disease and asthma hospitalization and also on zip code population that are separated by age category. The values in the table are the daily average hospitalization across zip codes for the whole data set, i.e. number of hospitalization per day, per zip code.

[18] The OSHPD health data also includes the Major Diagnostic Categories (MDCs) of: 1. Circulatory system 2. Ear, nose, mouth, throat 3. Nervous system 4. Skin, subcutaneous tissues and breast diseases and disorders 5. Newborn and neonate conditions began in perinatal period 6. Factors on health status and other contacts with health service 7. Human immunodeficiency virus infection. The first category (e.g. heart attack) is also expected to be related to air pollution. At this stage of the study, I only focus on the Respiratory System diseases, and another category can be done in the next steps.

The Figure 1.2 gives the annual total counts of respiratory disease hospitalization over the study period. We can see a slight decline in number of hospitalization over time for patients who are older than 65 years of age. However, this value for two other age categories is quite stable. In California over the same time period we see around 4 million increase in the total population of individuals older than 65, 2 million for individuals of ages 20-64, and almost constant population for children at ages 5-19. Therefore, as a matter of absolute values there has been an improvement in the health of elderly and adults, but not children. [19]

Since the patient's zip code is the highest geographical resolution that is available from the health data, I calculate the dependent variables in the zip code-daily level separately for each age category. For a particular zip code x at day t, I count the number of patients who are admitted to any hospital in California and their residential zip code is x and asthma (respiratory diseases) being the chief cause of admission. [20] I divide this number by the zip code population as the hospitalization rate of the zip code x at day t. Therefore, the dependent variable is the daily number of asthma (respiratory diseases) admissions per zip code divided by the zip code population. I divide the number of admission by the zip code population, in order to eliminate the size effect of population: regardless of pollution level, the larger the zip code population, the higher the hospital admission. I obtain the annual population of zip codes per age group, by linearly interpolating population from the US 2000 and 2010 Census data.

Table 1.9 gives some information on hospitalization rate for different range of zip code population. Because the number of asthma hospitalization always is equal or

[19] I draw the similar graph as Figure 1.2 with only difference that I divide values in Figure 1.2 by the total population of each age category and the results is very similar. Figure 1.13 presents the result.

[20] I use the residential zip code of the patients, and not the hospital zip code.

smaller than respiratory disease, so in Table 1.9 I use the asthma hospitalization in order to see the population effect on the health measure clearly. In Table 1.9 for each given range of population, the second column is the total number of zip codes, third column is the total number of asthma admission from all the zip codes for the years of 2004-2011, fourth column is the average asthma admission per zip code, and fifth column is the daily value of the forth column. As we expected, the total admission per zip is positively correlated with the population of the zip codes, fourth column. This correlation does not change even after calculating the daily value, fifth column. One explanation can be that these zip codes are mostly urban areas with higher pollution. But this hypothesis can be tested more specifically in the estimation part, to see which factors explain this variation.

The Figures 1.3 and 1.4 give a better visual insight about the distribution of hospitalization number and population in CA. For example, the range of 0-2000 in Figure 1.3 means that total number of hospitalization from these zip codes was less than 2000 for the entire period of study. In Figure 1.4 values are the average population for the years 2004-2011. The white colors in both figures are missing values. As we can see, the number of hospitalization and population of zip codes are correlated. I present a similar graph only for Los Angeles County in the Figures 1.14 and 1.15, the result is very similar. In both figures, the concentration is mostly in South coast, San Joaquin valley, Sacramento valley, and San Francisco bay area.

1.4.3 Pollution and Weather

Daily pollution data comes from the California Air Resource Board. The data includes daily ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2) in monitor level. It is necessary to control for CO and NO_2 because of their correlation with ozone level, and also their direct impact on daily environmental quality, and therefore, on public health, Brauer *et al.* (2008), Currie and Neidell (2005). Particulate matter less than 10 μm in diameter (PM_{10}) is also one of the main pollutant that is related to the health problem, Knittel *et al.* (2011), Currie *et al.* (2009). However, PM_{10} is not included in this study because it is not available on a daily basis. In most, but not all the monitors, PM_{10} is recorded once every six days. [21] Also, because of the high correlation between PM_{10}, CO, and NO_2, Currie and Neidell (2005), omitting PM_{10} will not bias the estimates of the health impact of ozone.

Weather conditions are correlated with ozone levels, and also affect time allocation of people between activities, and time spent outside, Zivina and Neidell (2014), so they affect individual's exposure to ozone. Weather conditions are also expected to have direct effect on health. [22] Therefore, almost all the studies in the literature, Neidell (2009) and Neidell and Kinney (2010) and Currie and Neidell (2005), control for the weather conditions, and (Knittel *et al.* (2011)) specifically show the importance of including weather conditions in the estimation of the health impact of pollution. To be consistent with the literature, I include daily maximum and minimum temperatures, average precipitation level, and maximum relative humidity. [23] The weather data comes from the National Climatic Data Center and it is publicly available. For the weather conditions, I obtain the weather data for Arizona, Nevada, and Oregon as well as California itself. This helps me to calculate more accurately weather conditions for the zip code near the borders of California in case monitors inside the border of California are too far from the zip code. In the weather data, there are some outlier for the maximum and minimum temperature. I drop those temperature that were

[21] $PM_{2.5}$ is recorded once every three days.

[22] Deschnes and Greenstone (2011) show that mortality rate increases at the extremes of the temperature.

[23] Some studies also include sun cover and average wind speed. These variables are frequently missing and after I assign them to each zip code, there is no enough observation left to include in the regression.

greater than 130 or less than -30 Celsius. [24] I use measures for the pollutants that correspond with air quality standards in California for the period of study: 1h maximum for ozone, 8h maximum for carbon monoxide, and 1h maximum for nitrogen dioxide. Table 1.11 gives the summary statistics of the pollutants and the weather conditions using the raw data.

Figure 1.5 gives the annual trend of three pollutants for the study period. From 2004 to 2011 there is a reduction in both CO and NO_2 pollution level, but ozone pollution is almost constant around 0.05 ppm.

To assign the daily pollution and the weather conditions to each zip code, if there are monitors within the zip code polygon, I assign the average ozone value of those monitors to the zip code. Figure 1.6 presents the distribution of pollution monitors across California that are mostly concentrated in densely populated areas. If there is not any monitor within the zip code polygon, first I find all the monitors within 20miles radius of the zip code centroid using latitude and longitude of the locations of the monitors. [25] Then, using an inverse distance weighted average of the nearest monitors I calculate the pollution levels and the weather conditions for each zip code according to equation (1.8):

$$P_{zt} = \frac{\sum_m AP_{mt}/(dist_m | dist \leq 20)}{\sum_m 1/(dist_m | dist \leq 20)} \tag{1.8}$$

Where z, t, and m are zip code, day, and monitor, respectively. P is the assigned pollution (weather) level, AP is the actual pollution (weather) that is recorded in the monitor level, and $dist$ is the distance between the monitor and centroid of the zip code. There are two reasons for using the average value instead of the value of

[24] Apparently the outlier numbers come from the sites that the data center does not have sufficient control on the recording process of them. So I drop those outliers from the dataset.

[25] Knittel *et al.* (2011) and Currie and Neidell (2005) use the same strategy to find the nearest monitors to the centroid of the zip codes.

one nearest monitor: to get more accurate values for each zip code polygon, and to preserve sample size in case one nearest monitor has missing value.

Table 1.12 presents the distance between patients and hospitals' zip codes that they are admitted. All the values are in percentages. For example, the distance between residence and the hospital of 34% of the patients was less than 5 miles. Out of this 34%, 4% were children between the ages of 5-19, 35% were adults between the ages of 20-64, and 61% were patients older than 65 years of age. The percentage among different age groups is almost constant for different distance levels. As we can see from Table 1.12 that choosing 20 miles radius, at least for 70% of the observations gives a plausibly accurate measure of pollution and weather conditions based on equation (1.8).

1.5 Estimation Strategy

In this section, I assess whether there is a difference between outdoor time on weekends and weekdays. If this difference exists, I quantify its effect on the estimate of the health impact of ozone.

1.5.1 The Weekend Effect

I use the term of Weekend Effect to remark the fact that there are two significant differences between weekdays and weekends in term of ozone level and the time that people spend outdoors. [26] In the empirical section, first using some summary statistics and running t-test, I want to test whether outdoor leisure time on weekends is different than on weekdays. If it is different I estimate the following equation to

[26] There is another term of Weekend Ozone Effect that comes from the atmospheric science literature. They use this term to the observation that ozone level on weekends is greater than on weekdays in some areas in California, despite the fact that its precursors are lower on weekends.

explain the factors that leads to this observation:

$$outdoor^d_{iwmy} = constant + \alpha income^d_{iwmy} + \beta HHsize^d_{iwmy} + \gamma Lh^d_{iwmy} + \varepsilon^d_{iwmy} \quad (1.9)$$

Where i, w, m, y represent individual, week, month, and year, respectively. Superscript d represents the difference value of a variable between weekend and weekday. [27] So $outdoor^d_{iwmy}$ is the difference of outdoor leisure time between weekend and weekday of individual i at week w, month m, and year y, i.e. $outdoor^d_{iwmy} = outdoor^{weekend}_{iwmy} - outdoor^{weekday}_{iwmy}$. $income$ is the family income, $HHsize$ is the total number of persons in the household of an individual, Lh total hours that an individual usually works per week, and ε^d_{iwmy} is the error term. All of these variables are the difference between weekend and weekday.

The difficulty of estimating this equation is that in the data set that I use, each individual is observed only throughout a single day. So I do not observe an individual's outdoor time on weekends and weekdays to see what factors explain the potential difference between these days. To overcome this issue, I match individuals in the sample by their characteristics, and treat those individuals in each matched group as one person. By doing this process, hypothetically I have information on individuals at different times. So I can calculate the average value of variables of interest among individuals within each matched group. Then I run the regression of equation (1.9).

The more matching criteria you choose, the less number of observation you end up with for each matched group. On the other hand, the less matching criteria leads to more heterogeneity within each group, and it would be hard to believe that individuals within each group are similar. I use individuals' age, sex, marital status, metropolitan status, labor force status, and family income range as a categorical

[27]For weekends I use average value over Saturday and Sunday, and for weekdays I use Monday through Friday

variables to construct the matched sample. [28] Therefore, all the individuals who are in the same cell based on these criteria I assign them a unique identification number, i. Then I calculate the average value of the individuals' outdoor time with the same id, i, who are surveyed on the weekend of week w, month m, and year y to calculate the $outdoor^{weekend}_{iwmy}$. The process is exactly the same for all the other variables in equation (1.9). These classifications help me to partially control for physical and economic factors, and in general individuals' life style in order to obtain more homogeneous groups of people within each group. This homogeneity is useful in order to test whether after controlling some personal characteristics, there is still a difference between weekends and weekdays outdoor leisure. I could also control for residential state of individuals, full/part time student, and hourly/non-hourly labor force status, but there will not be sufficient observation to do the estimation.

For explanatory variables, in order to explain the outdoor time difference between weekends and weekdays, I need to choose those variables that average value among people is meaningful. For example, average income of people is meaningful, but not average value of their residential state number. So, I consider individual's outdoor leisure time, family size, and weekly labor hours. I present the result of estimation of equation (1.9) in section 1.6.1. In the main estimation, in order to calculate the dependent variable I only use the outdoor leisure time. For sensitivity analysis, I repeat the same estimation for another dependent variables. In the second dependent variable, I include all the outdoor time both leisure and work related time. [29]

[28] I use five age categories to match individuals: 15-25, 26-35, 36-45, 46-64, and 65+.

[29] I was also interested in calculating the dependent variable based on only those activities that their average time on weekends and weekdays varies largely. These activities are golfing, playing baseball, fishing, hunting, playing unclassified sports. The problem is that by choosing only these activities there is no enough observation to do the estimation.

1.5.2 Health Effect of Ozone

In order to estimate the health impact of ozone pollution on respiratory disease and asthma hospitalization, I estimate the following equation which controls for the weekend effect:

$$Health_{zt} = constant + \alpha_0 x_{zt} + \alpha_1 x_{zt} wd_t + \sum_{j=1}^{4}[\beta_j x_{zt-j} + \gamma_j x_{zt-j} wd_{t-j} + \mu_j P_{zt-j}] + f_z + g_t + \epsilon_{zt}, \tag{1.10}$$

where $z, t, and \epsilon$ are zip code, day, and error term, respectively. Health is the number of the respiratory disease (asthma) hospitalization that is normalized by the population of zip code. *constant* is constant term, x is ozone level, wd is weekend fixed effect, P include CO, NO2 and weather conditions, f is the zip codes fixed effect, and g is day of the week and month-year fixed effects to account for time and seasonal effects.

This model is a conventional model to estimate the health impact of pollution in the literature (e.g. Neidell (2009), Neidell and Kinney (2010)) except for the interaction of ozone pollution and the weekend fixed effect that accounts for the weekend effect. This interaction helps me to estimate the different health impact of ozone for weekends and weekdays. According to previous studies that usually find the health impact of ozone up to four days after exposure to pollution, I include four lags of ozone and weather variables. Instead of focusing on the impact of each lag separately I calculate the overall impact of ozone on hospitalization, $\sum_j \frac{\partial Health_{zt}}{\partial x_{zt-j}} = \alpha_0 + \alpha_1 wd_t + \sum_{j=1}^{4}[\beta_j + \gamma_j wd_{t-j}]$. Calculating this value leads to seven different values of the health impact of ozone for each day of the week. I.e. because of the accumulation effect of ozone and different health impact of ozone on weekends and weekdays, the overall health impact of ozone will be different for each day of the week.

1.6 Estimation Results

1.6.1 Outdoor Time on Weekends and Weekdays

Table 1.13 presents average outdoor leisure for different classifications. In all the cases the null hypothesis is rejected with very small P-values. This rejection means for all the subgroups, average outdoor leisure on weekends is greater than on weekdays and the difference varies between 10-50 minutes. Besides this difference, from Table 1.13 we can infer another important point. Comparing the weekend-weekday outdoor leisure difference within each groups tells us that this difference varies even between subgroups. For example, the weekend-weekday outdoor leisure difference for male is 45 minutes and for female is 30 minutes. So, this means that controlling for these criteria in the estimation that is provided in the following is important. Because part of the weekend-weekday outdoor leisure difference is explained by these matching criteria.

Table 1.14 represents the results of equation (1.9). The dependent variable in column (1) is the outdoor leisure, and in column (2) is all the outdoor activities including work related activities.

In Table 1.14, constant term in both of the specifications is statistically significant and it is positive. The magnitude of the constant term as compare to the mean (80 minutes) and median (45 minutes) is very large. That means people spend much more time outdoors on weekends than on weekdays. The coefficient of household size is not significant. However the coefficient of weekly labor hours is negative and significant in the first specification. This can happen because people who work more, their work hours is also spread over weekends. So their outdoor leisure on weekends can not be that different from weekdays. As we can see in column (2), if I add the work related outdoor activities the significant coefficient disappears.

It is worth noting that these estimations are after controlling for other demographic characteristics of individuals. This means that even if I control for some physical, economic, and cultural characteristics of individuals, there is still a large difference between outdoor time that people spend on weekends and weekdays. If more continuous information about individuals was available, getting a precise magnitude and the sign of the estimate of each variable would be possible.

Other than the fact that time that people spend outdoors on weekends is significantly different than on weekdays, ozone pollution levels in some locations are also higher on weekends as compared to weekdays. Ozone is formed from volatile organic compounds (VOC) and oxides of nitrogen (NO_X) in the presence of sunlight and heat. To control ozone pollution, the EPA regulates the measurements of VOC and NO_X. VOC and NO_X are mainly released by vehicles, construction equipment, and industrial paints and solvents that are dependent on human activity patterns over the week. Usually, levels of these precursors are lower on weekends than on weekdays and much of this difference is ascribable to mobile sources (e.g. motor vehicle) rather than to stationary and area sources (Blanchard and Tanenbaum (2003)). We expect to see the same trend over a week for ozone pollution. However, in some regions in California such as the South Coast Air Basin, the San Francisco Bay Area Air Basin, and some urban area in the Central Valley, the ozone level is typically higher on weekends than on weekdays and this phenomenon is called Weekend Ozone Effect. [30] The weekend ozone effect has been well documented in the atmospheric science literature, and it usually happens in highly polluted and populated areas despite the fact that the level of precursors of ozone (VOC and

[30] The main reasons is a decrease NO_X emission at weekend but there are several other hypothesis that can be possible explanations for weekend ozone effect: NO_X-reduction, NO_X-timing, carryover near the ground, carryover aloft, increased weekend emissions, aerosol and ultraviolet (UV) radiation, and ozone quenching hypothesis (EPA (2014)). For more information on the weekend ozone effect hypothesis see the California Air EPA June 2003 report on the weekend ozone effect.

NO_X) on weekends are lower than on weekdays (Yarwooda *et al.* (2008), Qin *et al.* (2004), Fujita *et al.* (2003), Blanchard and Tanenbaum (2003)).

In sum, based on American Time Use Survey data, the time that people spend outdoors on weekends is different than on weekdays. So, I expect that health outcomes will be different on weekends than on weekdays even if the ozone level stays unchanged. In addition, there are many research in atmospheric science literature that show the ozone levels on weekends is significantly higher than on weekdays in some regions of California. Since, dose-response is not a linear function of ozone pollution, I expect that the health impact of ozone on weekends is different than on weekdays, even if people spend the same time outdoor every day of the week. In other words, the same amount of change in the ozone levels may cause greater change in health outcomes on weekends as compared to weekdays, because the original level of ozone is different. Both these factors are accelerating each other and if we do not control for the weekend effect, the estimate for the weekend health impact of ozone will be biased upward. To account for these differences, I use the interaction of ozone and the weekend fixed effect in the empirical estimation to distinguish between the coefficient of ozone for weekends and weekdays. However, using this method it is not possible to distinguish between the impact of outdoor time difference and the weekend ozone effect on the estimate.

1.6.2 Health Impact of Ozone

Before presenting the results of equation (1.10), it is worth to discuss some information about the dependent and pollution variables. Based on the discussion of outdoor time and ozone level on weekends and weekdays, I was expecting to observe more hospitalization on weekends than weekdays. But in Table 1.15 for all age categories, Sunday and Saturday have the lowest number of hospitalizations. In terms of percentages, every day on weekend accounts for around 11-12%, and every day on weekday accounts for around 14-16% of the total hospitalizations.

As we can see in Table 1.16, the mean value of O_3 is higher on weekends than weekdays and the opposite is true for CO. The magnitude of medians that are not shown here are different from means, but medians repeat the patterns of the means for both O_3 and CO. The median value for NO2 is different than its mean value, and is higher on weekdays than weekends. So possibly the variation of the hospital admissions in Table 1.15 is mostly driven by CO and NO2.

Figure 1.7 displays the daily variation in respiratory disease hospitalization for age categories. Purple lines separate the seasons: spring, summer, fall, and winter. The hospitalization rate is in its highest level in the spring and the winter, and the lowest rate is in the fall. In this figure I do not control for the population, but Figure 1.8 plots the daily value of respiratory disease hospitalizations after normalizing by the population. In Figure 1.8 the curve for old individuals, which are the most vulnerable age group based on the hospitalization rate, has a declining trend from the spring to the winter but not for two other groups. In both figures there is enough daily variation in hospitalization rate to be explained.

Figure 1.9 presents the daily variation of three pollutants over the year indicating large daily variation in the data. I control for the seasonal fixed effects, CO, NO2, and weather conditions that I include in the main estimation to see if there is still enough variation in ozone pollution. Figure 1.10 provides the residuals after this process, and there is still a sufficient variation in O_3 to explain the health impact. By comparing Figures 1.7 and 1.9, there is a similar patterns over the year between hospitalization, CO and NO2, but the pattern of O_3 is reverse. This seasonal trend should not cause any problem in the estimation, because I use the daily variation to explain the health impact.

In all analyses I only focus on months from March to October, because the ozone level is high in this period of the year and most likely to have the weekend ozone effect issue. Tables 1.17 and 1.18 provide the estimation results for the equation (1.10) without interaction of ozone and weekend fixed effect. For the dependent variable I use respiratory disease hospitalization rate in Table 1.17 and asthma hospitalization rate in Table 1.18. The coefficients have anticipated signs, but the coefficient for individuals ages 65+ is not significant in Table 1.17. For children, estimate of the impact of ozone in column (1) indicates that 1ppm increase in the five-day average ozone leads to 1.23 increase in the respiratory disease hospital admissions for a zip code with 10000 population. Based on Table 1.17 adults ages 20-64 are more vulnerable group to ozone that is counterintuitive. This results can be explained by avoidance behavior. Since old people and children would respond to pollution level, then without controlling for the avoidance behavior we underestimate the health impact of pollution. In Table 1.18 only the coefficient for adults ages 20-64 is significant.

In Tables 1.19 and 1.20 I control for the weekend effect, interaction term, and run the same estimation as in Tables 1.17 and 1.18. As mentioned early, the overall health impact of ozone should vary across days of the week. The impact of ozone on respiratory disease and asthma for adults ages 20-64 on weekdays are significantly greater than weekends. This result indicates that this age group spends less time outdoors on weekends and not controlling for the weekend effect overestimates the health impact of ozone on weekends. The opposite is true for children if we consider the respiratory disease, but not for asthma. Ozone does not have a significant impact on asthma hospitalization of old individuals ages 65+. For the respiratory disease of old people, the estimate is very large positive value on Thursdays and negative value on Sundays. Small or negative number can be explained by the presence of avoidance behavior, but these values are very unusual and the reason should be something else.

Overall, controlling for the weekend effect changes the magnitude of the estimates for days of the week, but the significance of the results depends on individuals' age. Chi-square test for the difference between the estimates confirmed that with the inclusion of weekend effect the coefficients are significantly different for children and seniors but not for adults, Tables 1.19.

1.7 Discussion and Conclusion

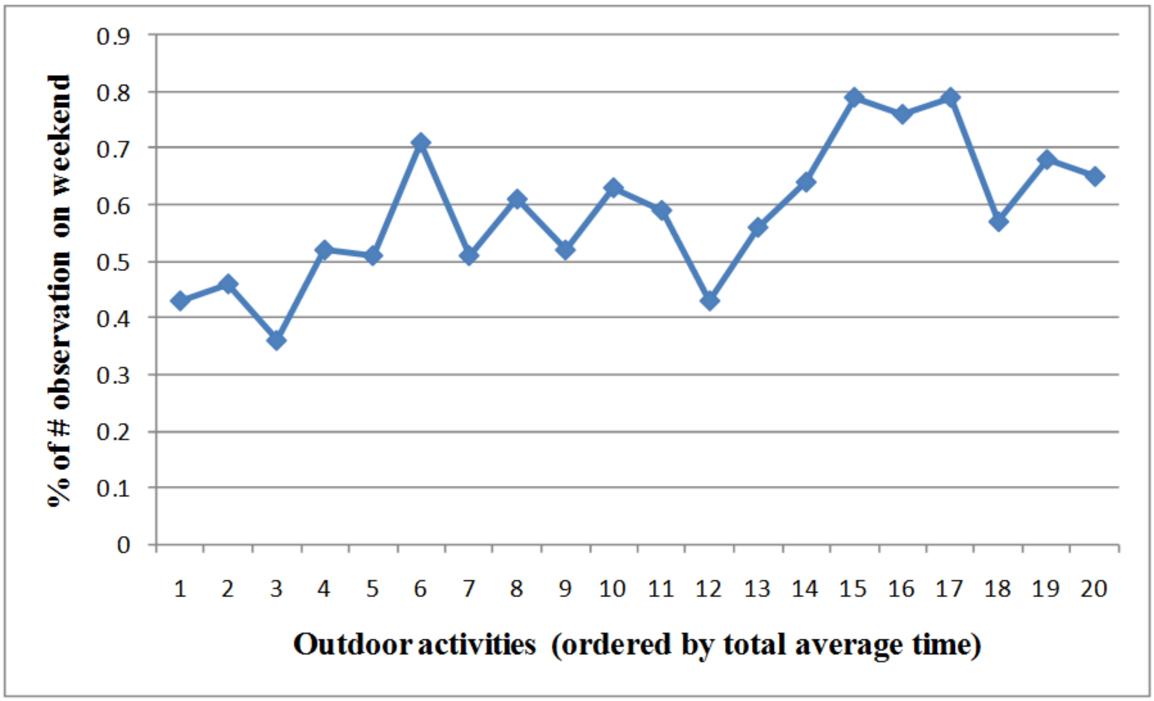
In this study I show that time people spend outdoors on weekends is significantly greater than weekdays. In doing so, I develop a theoretical framework that explains how an individual's shadow value, outdoor and indoor leisure varies over time and how it depends on his characteristics. In the estimation part, I show that after controlling for some of the individuals' characteristics there is still a large differences, around 50 minutes, between weekends and weekdays outdoor time. This difference potentially can be explained by some other personal taste, and money or time constraints that I did not have data on them.

In the second stage of this study, I examine the effect of controlling for the weekend effects in estimating the impact of ozone on respiratory disease and asthma hospital admissions. Results of the estimation indicates that controlling for the weekend effect changes the health impact of ozone for adults ages 20-64. For children, the impact of ozone on respiratory disease is greater on weekends than weekdays, and the effect on asthma varies over day of the week. Ozone does not have a significant impact on asthma hospitalization of old individuals ages 65+.

Controlling for the weekend effect in addition to partially correcting for the measurement error regarding the pollution measure, it gives a precise estimate for each day of the week. These differences in the health impact of ozone can be considered in the pollution policies. For example, there can be a different pollution standards for weekends and weekdays to reach a better improvement in public health.

1.8 Figures

Figure 1.1: Percentage of Activities on Weekends vs Activities



Note: The horizontal line is ordered based on mean time of activities.

Figure 1.2: Annual Respiratory Disease Hospitalization by Age Categories

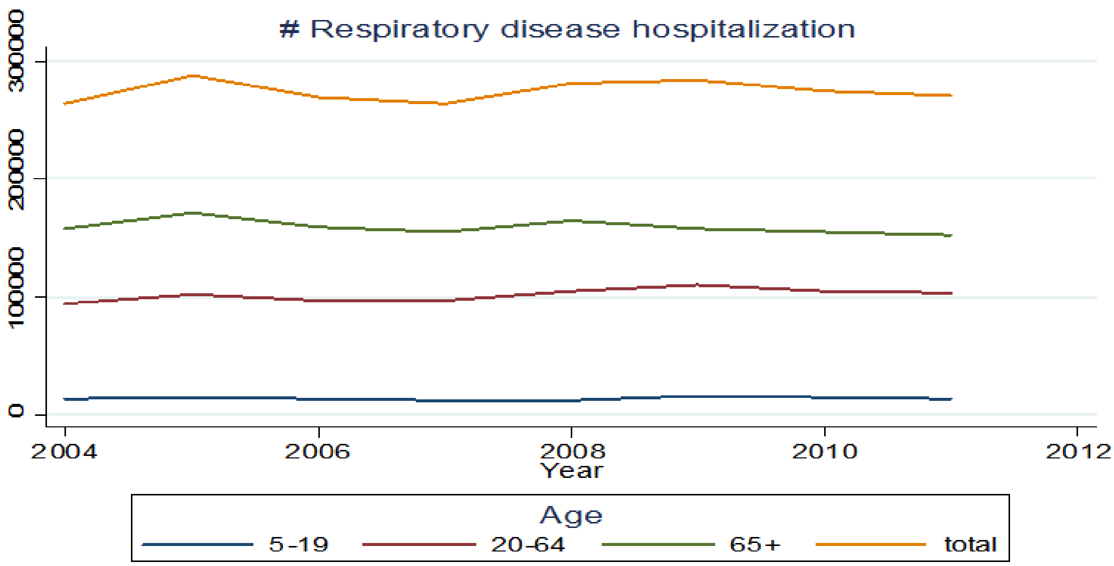


Figure 1.3: Distribution of Respiratory Disease Hospitalization in California

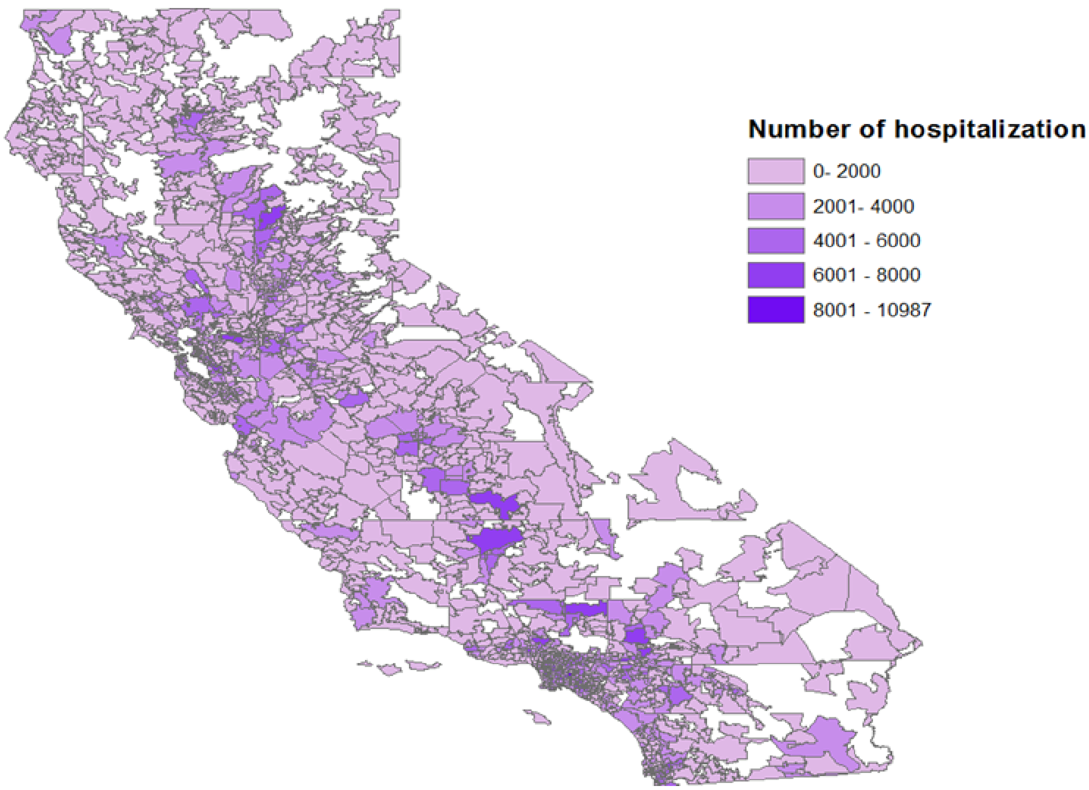


Figure 1.4: Distribution of Population in California

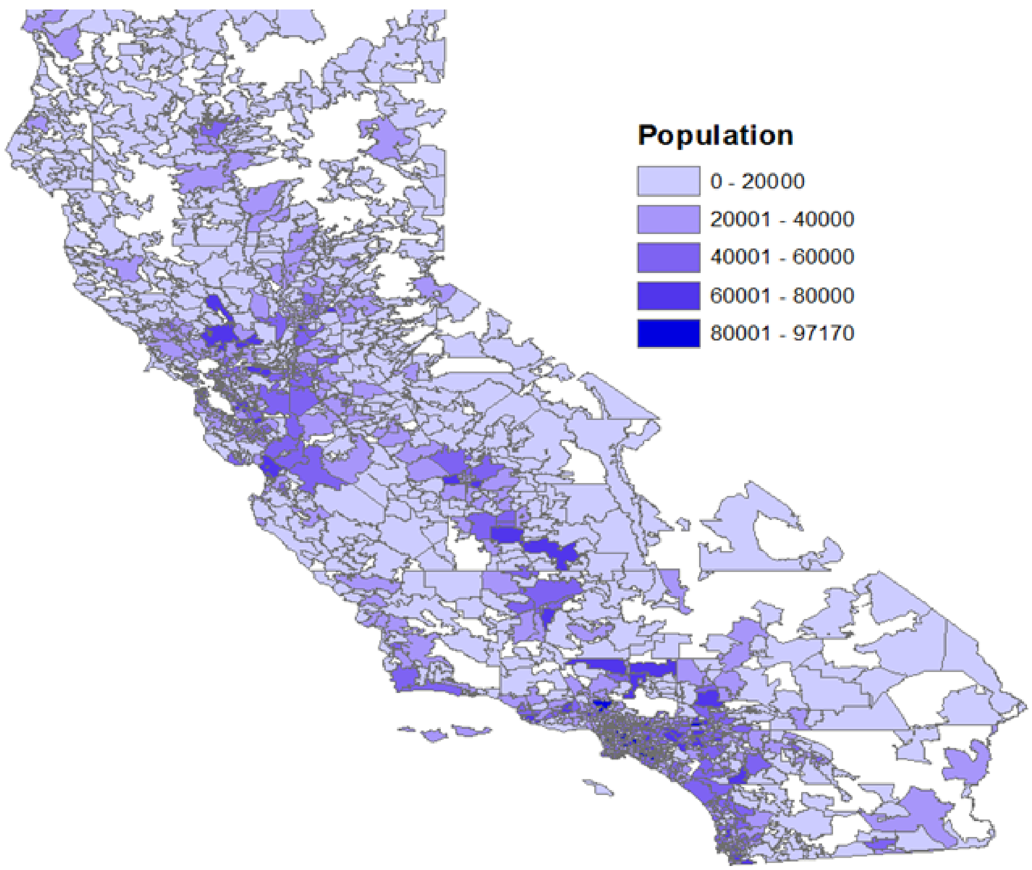


Figure 1.5: Annual Average Level of Pollutants

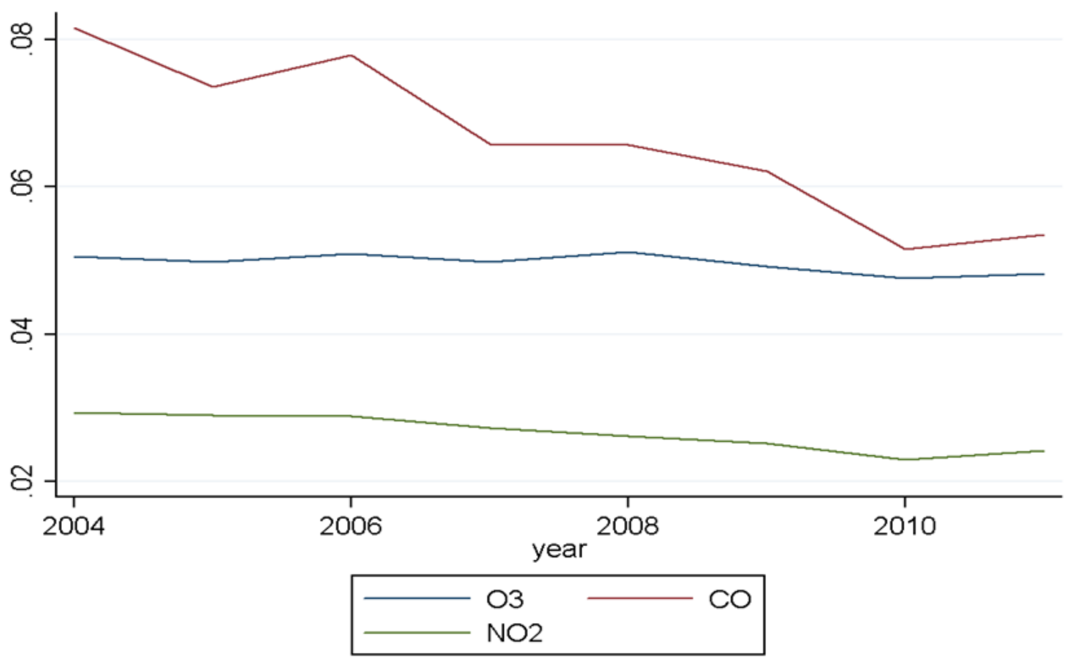


Figure 1.6: Pollution Monitors in California

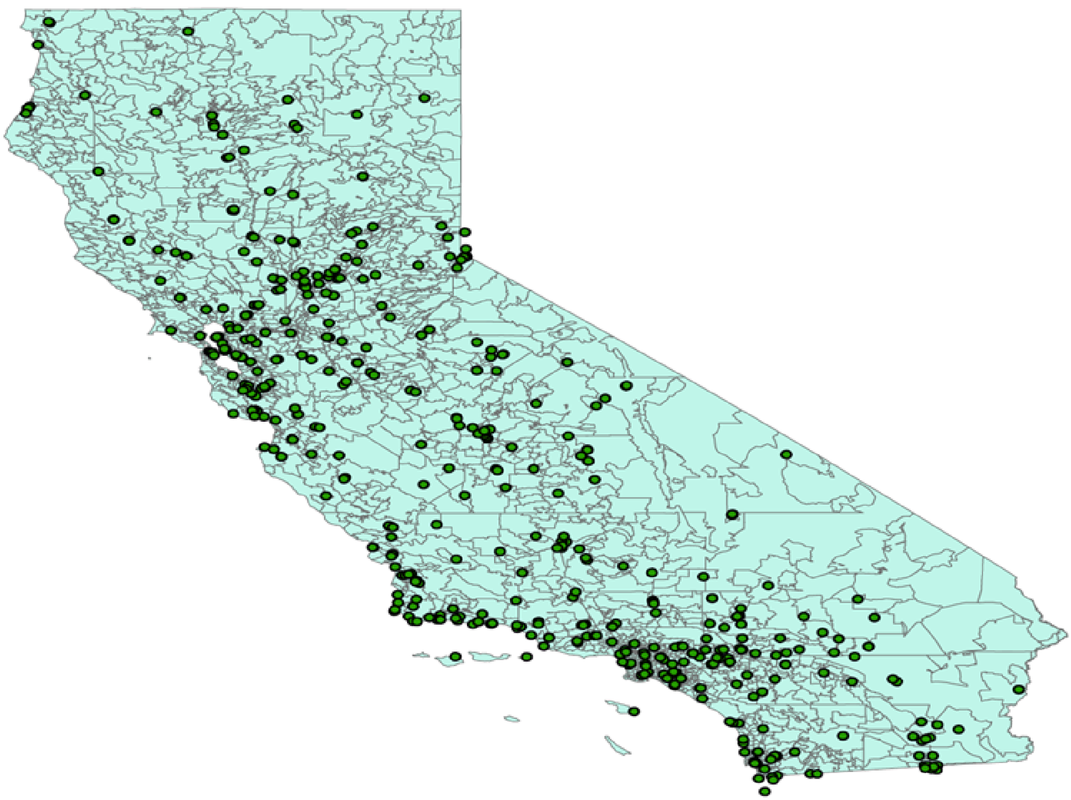


Figure 1.7: Average Daily Respiratory Disease Hospitalizations over the Year

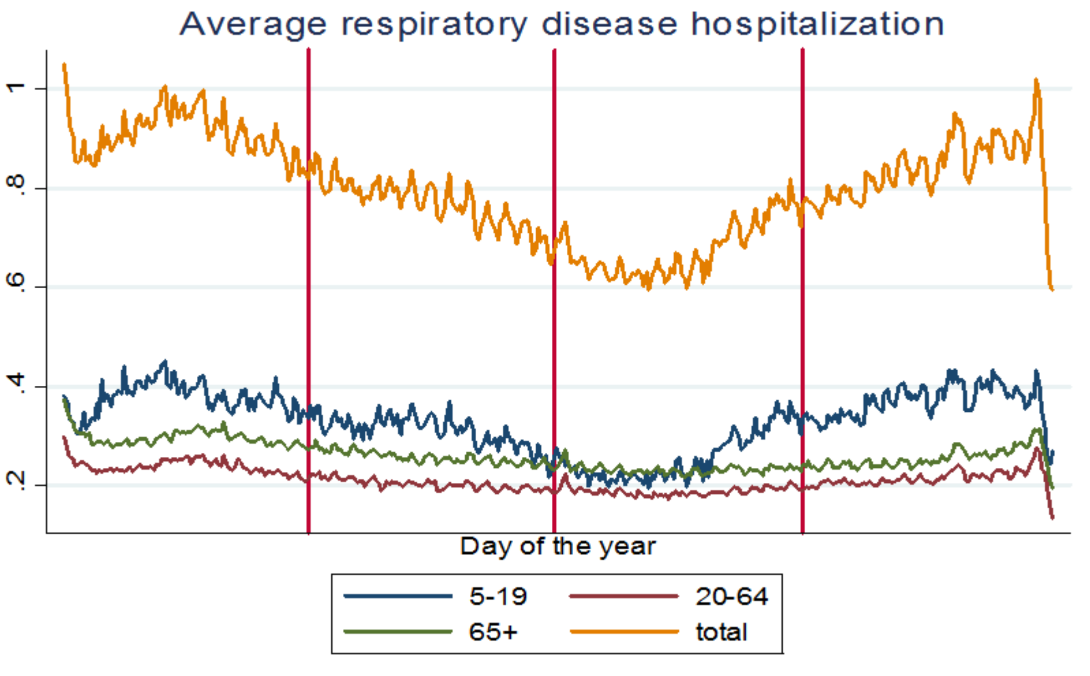


Figure 1.8: Average Daily Respiratory Disease Hospitalizations Normalized by Population over the Year

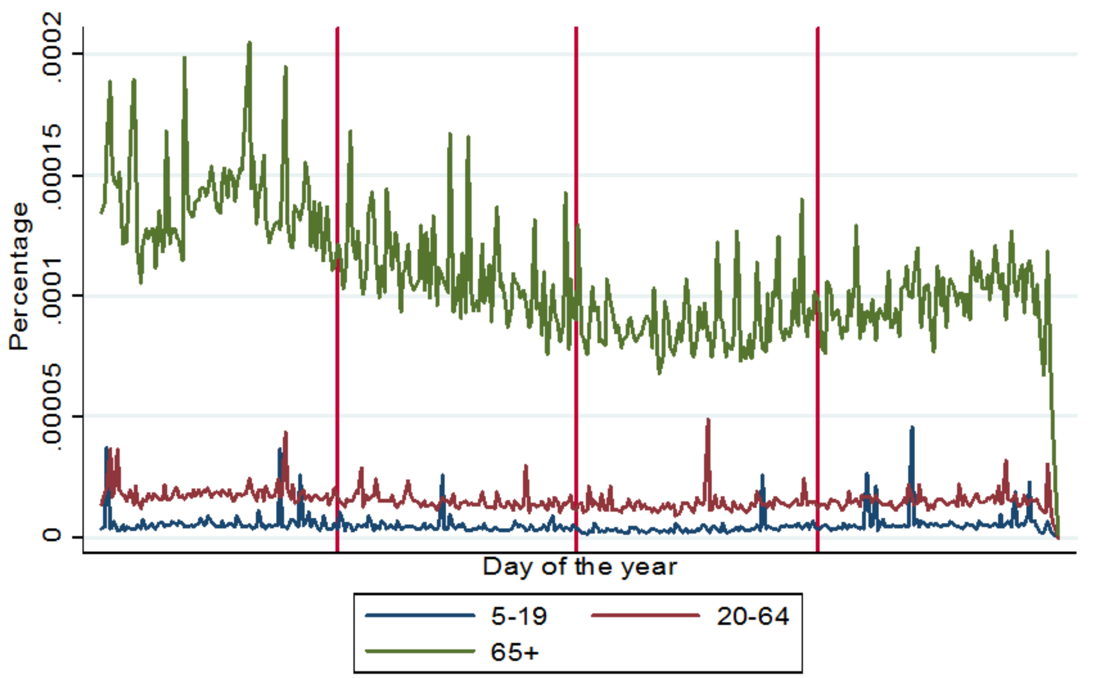


Figure 1.9: Average Daily Pollution Levels over the Year

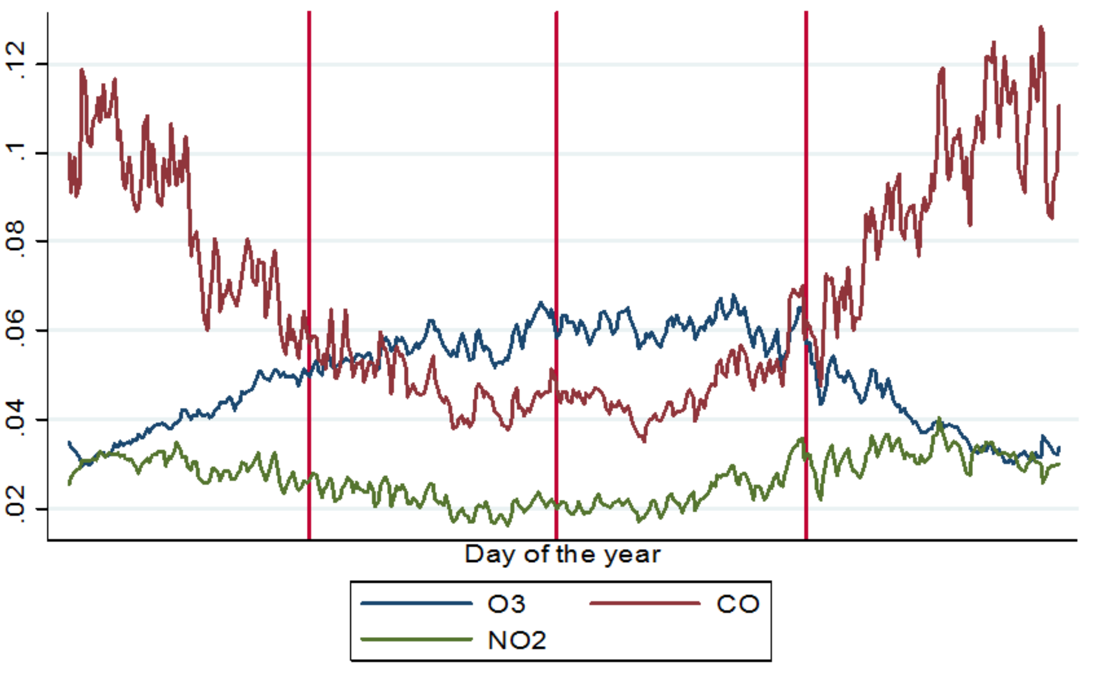


Figure 1.10: Average Daily Ozone Level and Its Residual over the Year

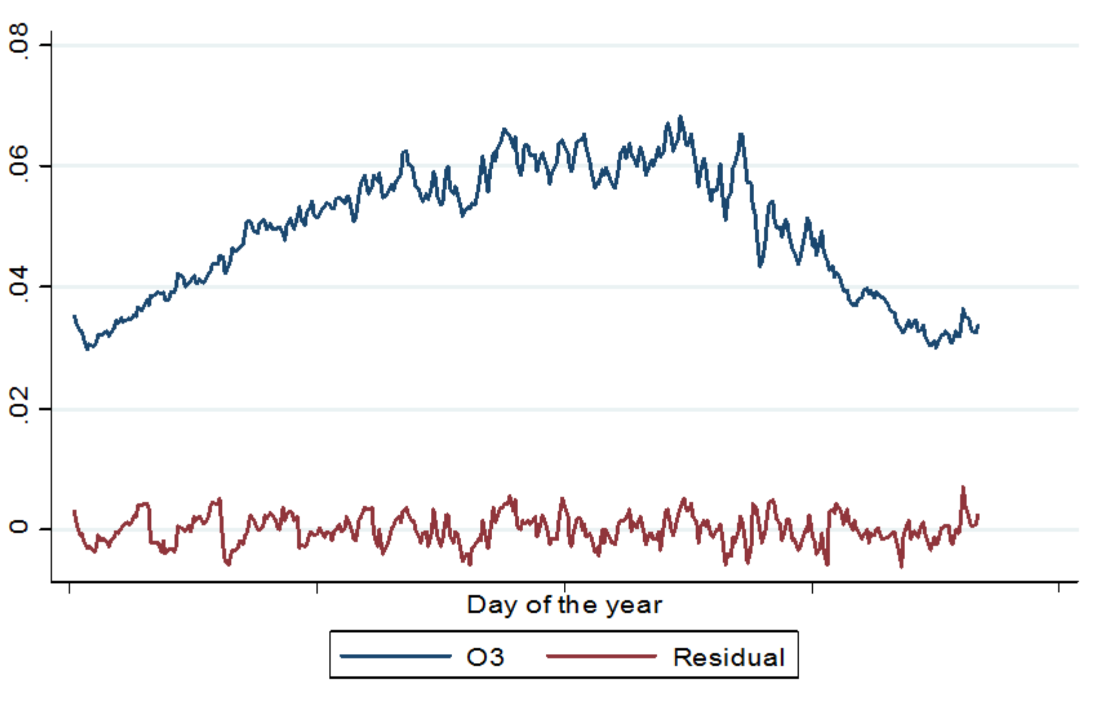
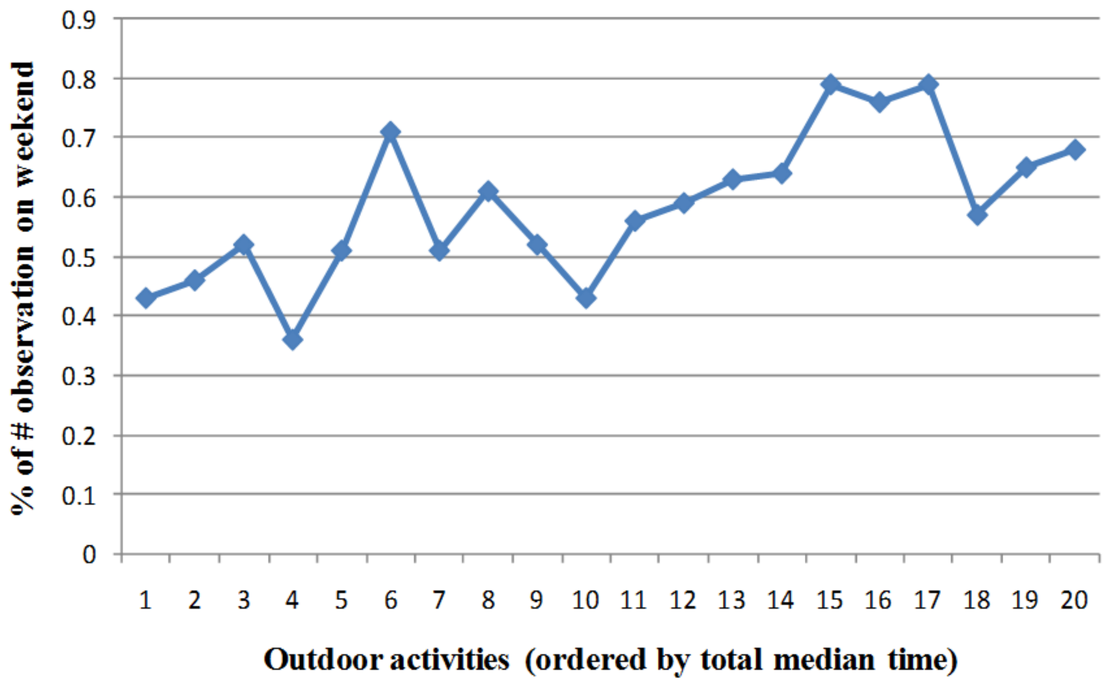
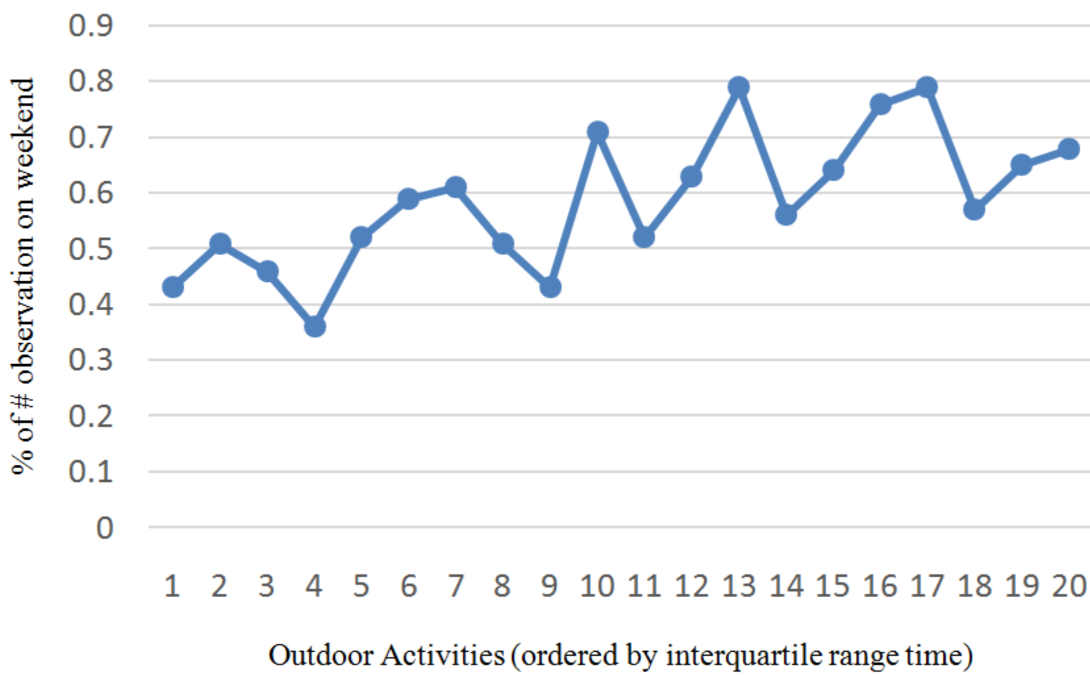


Figure 1.11: Percentage of Activities on Weekends vs Activities



Note: The horizontal line is ordered based on median time of activities.

Figure 1.12: Percentage of Activities on Weekends vs Activities



Note: The horizontal line is ordered based on interquartile time of activities.

Figure 1.13: Annual Respiratory Disease Hospitalization by Age Categories Normalized by the Population of Each Group

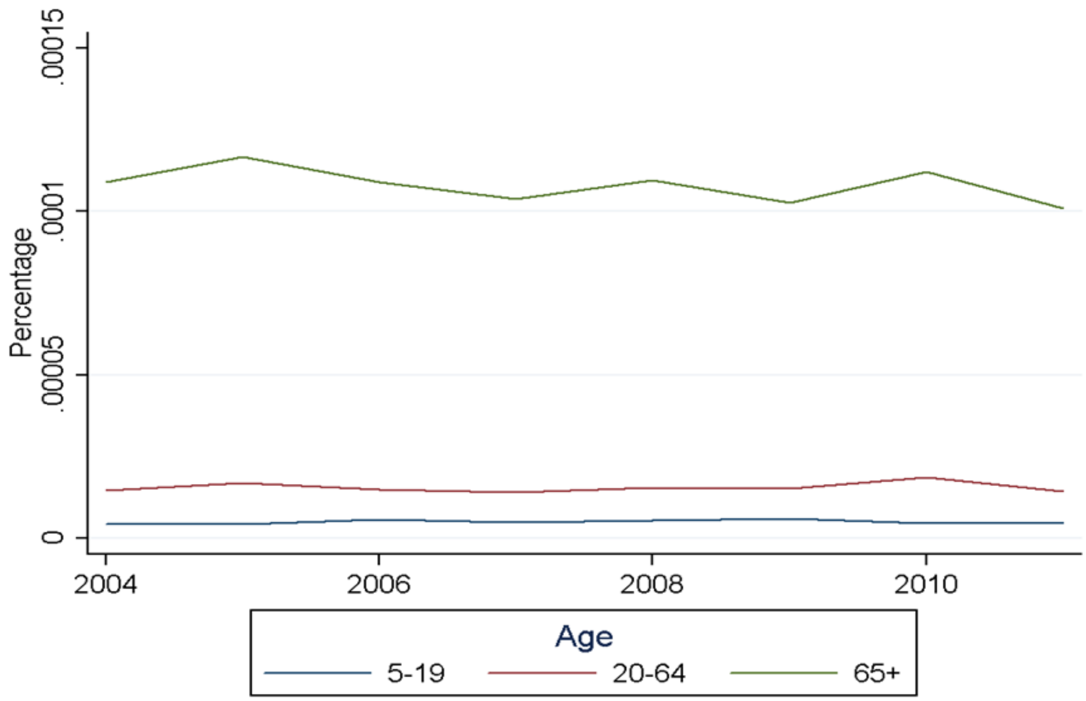


Figure 1.14: Distribution of Respiratory Disease Hospitalization in Los Angeles

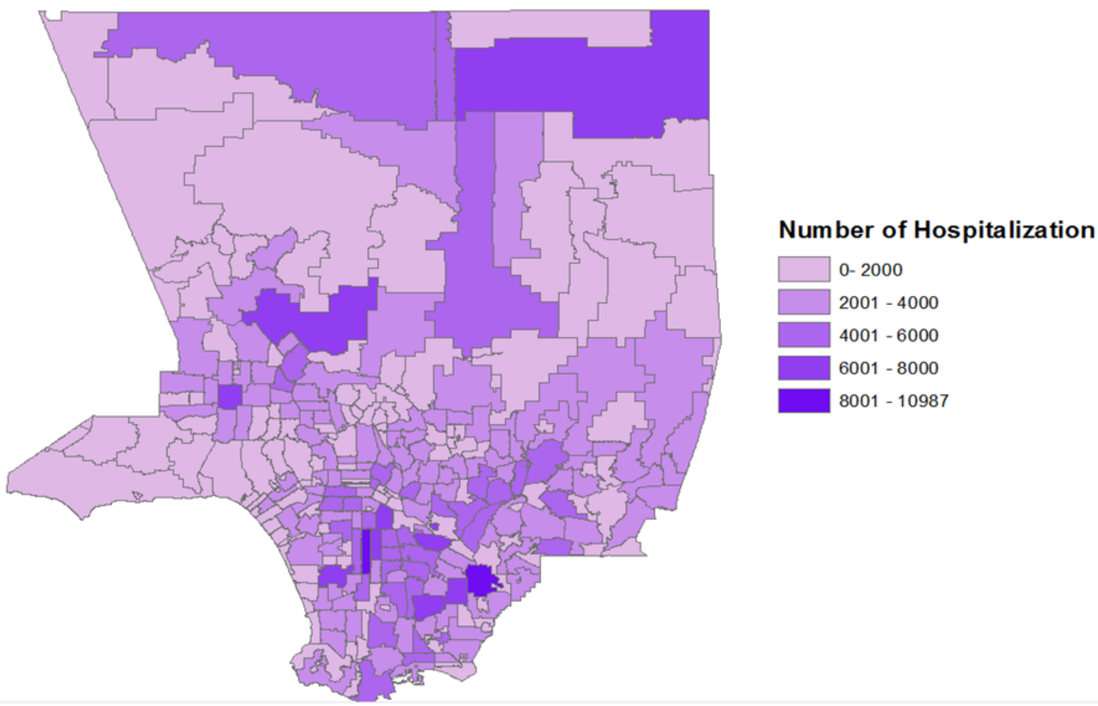
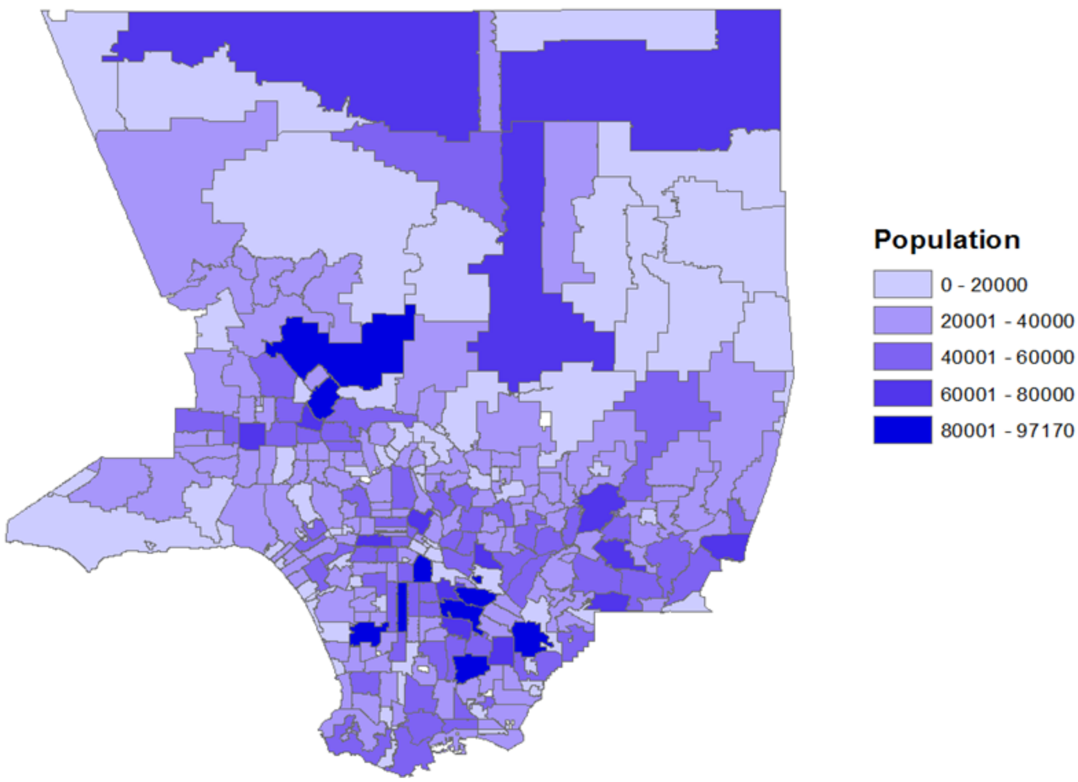


Figure 1.15: Distribution of Population in Los Angeles



1.9 Tables

Table 1.1: Activities and ATUS Codes Used to Define Time Spent Outdoors

Activity Code	Activity	Location	Classification
Any	Any besides working	Outdoors away from home	1
Any	Any	Walking	2
Any	Any	Biking	3
130106	Boating	Anywhere	4
130112	Fishing	Anywhere	5
130136	Yoga	Outdoors away from home	6
130102	Playing baseball	Anywhere	7
130103	Playing basketball	Not at gym	8
130104	Biking	Not at gym or at home	9
130108	Climbing/spelunking/caving	Not at gym	10
130109	Dancing	Outdoors away from home	11
130110	Participating in equestrian sports	Outdoors away from home	12
130113	Playing football	Outdoors away from home	13
130114	Golfing	Anywhere	14
130116	Hiking	Anywhere	15

Table 1.2: Activities and ATUS Codes Used to Define Time Spent Outdoors (Cont'd)

Activity Code	Activity	Location	Classification
130118	Hunting	Anywhere	16
130119	Participating in martial arts	Outdoors away from home	17
130120	Plying racquet sports	Outdoors away from home	18
130121	Participating in rodeo competitions	Not at gym	19
130122	Rollerblading	Anywhere	20
130123	Playing rugby	Anywhere	21
130126	Playing soccer	Not at gym	22
130127	Playing softball	Anywhere	23
130130	Playing volleyball	Not at gym	24
130131	Walking	Not at gym or at home	25
130132	Participating in water sports	Not at gym	26
130124	Running	Outdoors away from home	27
130125	Skiing/ice skating/snowboarding	Outdoors away from home	28

Table 1.3: Activities and ATUS Codes Used to Define Time Spent Outdoors (Cont'd)

Activity Code	Activity	Location	Classification
130128	Using cardiovascular equipment	Outdoors away from home	29
130133	Weightlifting/strength training	Outdoors away from home	30
130134	Working out, unspecified	Outdoors away from home	31
130199	Playing other sports	Outdoors away from home	32
150301	Building house/wildlife site/other structure	Anywhere	33

Table 1.4: Summary Statistics

Category	Outdoor leisure		Household size		Weekly labor hours	
	Mean	Median	Mean	Median	Mean	Median
Age						
15-25	94.6[117.7]	53	3.6[1.5]	4	31.6[15.1]	35
26-35	89.2[113.2]	45	3.2[1.4]	3	41.1[12.2]	40
36-45	86.1[116.2]	45	3.4[1.4]	4	42.1[12.4]	40
46-64	76.5[104.2]	40	2.3[1.3]	2	41.6[12.6]	40
65+	70.7[93.8]	40	2.6[1.5]	2	30.5[16.2]	32
Male	98.6[125]	60	2.8[1.5]	3	43.5[13.3]	40
Female	67.2[91.9]	36	2.7[1.5]	2	36.8[12.8]	40
Married	85[113.1]	45	3.3[1.3]	3	41.2[13.2]	40
Single	78.7[106.2]	40	2.3[1.5]	2	38.7[13.6]	40
Metropolitan	79.1[103.9]	45	2.8[1.5]	3	40[13.2]	40
Non-metropolitan	97.1[137.3]	45	2.7[1.5]	2	40.5[14.7]	40
Employed	82.2[112]	45	2.9[1.4]	3	40.1[13.4]	40
Unemployed	81.1[105.7]	45	2.7[1.5]	2	No Obs.	No Obs.

Notes: Values in the brackets are the standard deviation.

Table 1.5: Summary Statistics (Cont'd)

	Mean	Median	Mean	Median	Mean	Median
Category	Outdoor leisure		Household size		Weekly labor hours	
Age						
Hourly labor force	81.7[114.3]	40	2.9[1.5]	3	38.1[11.7]	40
Non-hourly labor force	81[106.2]	45	2.8[1.4]	3	44.1[11.3]	40
Full time student	86.6[105.5]	46	3.7[1.5]	4	25.5[14.7]	24
Part time student	74.6[102.6]	40	3.1[1.5]	3	38.1[12.5]	40

Notes: Values in the brackets are the standard deviation.

Table 1.6: Summary Statistics

Activity	Mean (minutes)	Median (minutes)	Interquartile range	#Obs	Weekday mean	Weekend mean	% of Obs on weekend
Walking	14.7	5	12	33579	13.5	16.3	0.43
Biking	36.7	15	35	1825	27.6	47.2	0.46
Working out, unspecified	38.4	30	45	33	40.9	34.1	0.36
Any besides working	48.5	20	50	9032	34.3	61.4	0.52
Running	55.1	45	30	609	48.8	61.2	0.51
Playing other sports	86.2	60	76	78	76.4	90.4	0.71
Playing basketball	104.1	90	75	381	102.8	105.4	0.51
Participating in water sports	106.3	90	65	1213	97.3	112	0.61
Playing volleyball	107	90	85	88	105.2	108.7	0.52
Playing soccer	110.1	120	85	182	104	113.6	0.63

Table 1.7: Summary Statistics (Cont'd)

Activity	Mean (minutes)	Median (minutes)	Interquartile range	#Obs	Weekday mean	Weekend mean	% of Obs on weekend
Plying racquet sports	114.8	119	57	49	99.3	125.5	0.59
Playing softball	115.5	100	75	111	114.3	117.1	0.43
Playing baseball	120.3	100	120	125	104.3	132.9	0.56
Rollerblading	120.6	120	120	89	101.6	131.3	0.64
Hiking	125.5	120	90	182	96.1	133.2	0.79
Skiing/ice skating/snowboarding	138.1	120	120	46	176.3	126.2	0.76
Boating	156.1	120	159	219	176.4	150.7	0.79
Golfing	189.1	180	167.5	476	179.9	195.9	0.57
Fishing	239.3	205	200	395	211.7	252.5	0.68
Hunting	246.7	197	197.5	216	216.6	262.7	0.65

Table 1.8: Summary Statistics: Hospitalization Rate (Person/Zip Code/Day)

	Age 5-19	Age 20-64	Age 65+
Respiratory diseases (mean)	.014	.104	.163
Standard deviation	(.122)	(.365)	(.477)
Range	[0,4]	[0,10]	[0,16]
Asthma	.005	.014	.008
Standard deviation	(.073)	(.123)	(.094)
Range	[0,3]	[0,4]	[0,3]
# Observation	7766676	7766676	7766676
Population / zip code (mean)	4669	12703	2342
Standard deviation	(5139)	(12486)	(2248)
Range	[0,30536]	[0,59850]	[0,11318]
# Observation	1656	1656	1656

Table 1.9: Summary Statistics: Asthma Hospitalization Rate Among Zip Codes with Different Population During 2004-2011

Population range	# zip codes in each range of population	Total # hospitalization of zip codes in 2004-2011	# hospitalization /zipcode in each range of population	# hospitalization /zipcode/day in each range of population
0-5000	1448	50519	34.88	0.008
5001-10000	694	43274	62.35	0.019
10001-15000	331	21333	64.45	0.021
15001-20000	260	24158	92.91	0.031
20001-25000	164	17854	108.86	0.037
25001-30000	129	18446	142.99	0.048
30001-35000	75	13410	178.8	0.061
35001-40000	51	10338	202.70	0.069
40001-45000	30	7076	235.86	0.080

Table 1.10: Summary Statistics: Asthma Hospitalization Rate Among Zip Codes with Different Population During 2004-2011 (Cont'd)

Population range	# zip codes in each range of population	Total # hospitalization of zip codes in 2004-2011	# hospitalization /zipcode in each range of population	# hospitalization /zipcode/day in each range of population
45001-50000	11	3277	297.90	0.101
50001-55000	6	1764	294	0.100
55001-59850	5	2168	433.6	0.148

Notes: Column 4 is the total number of asthma admission during 2004-2011 per zip code in the specific population range. Column 5 is the average daily asthma admission during 2004-2011 for each zip code in the specific population range.

Table 1.11: Summary Statistics: Pollution and Weather

Variables	Mean	Standard deviation	Range	#Observation
Ozone max 1h (ppm)	0.05	0.02	[0,.44]	5391324
Carbon Monoxide max 8h (tenths of ppm)	0.70	0.71	[0,45.69]	4199740
Nitrogen dioxide max 1h (ppm)	0.03	0.02	[0,.21]	4805116
Precipitation (mm)	19.91	101.14	[0,2000]	4472764
Maximum temperature (Celsius)	161.07	81.99	[-30,130]	3731256
Minimum temperature (Celsius)	57.27	82.22	[-30,130]	4586985
Average sun cover sunrise to sunset (%)	67.00	32.04	[0,100]	729
Average daily wind speed (meters per second)	28.12	16.52	[0,1542]	403547
Maximum relative humidity (%)	80.58	19.52	[3,100]	4788305
California Standard				
Ozone max 1h (ppm)	0.09 since 2005			
Carbon Monoxide max 8h (ppm)	9 since 1989			
Nitrogen dioxide max 1h (ppm	0.18 since 2007			

Table 1.12: Summary Statistics: Percentage of the Respiratory Disease Hospitalization

Distance between patient's and hospital zip code	Age 5-19	Age 20-64	Age 65+	Total
Any distance	5	37	58	100
$<$ 5 miles	4	35	61	34
$<$ 10 miles	4	36	60	48
$<$ 20 miles	4	37	59	68
$<$ 30 miles	4	37	59	80
$<$ 40 miles	5	37	58	86
$<$ 50 miles	5	37	58	89%

Notes: Numbers in each cell presents the percentage of the respiratory disease hospitalization of each age group in the whole dataset for the years 2004-2011. Total number of respiratory disease hospitalization was 2,470,707 but for 56,258 observations there is not enough information to calculate the distance between patient's and hospital's zip code. So each of these percentages are calculated after dropping those observations without distance values.

Table 1.13: Summary Statistics: Outdoor Time

	Outdoor leisure			
Categories	Weekday	Weekend	P-value	#Observation
Age				
15-25	74.1[1.8]	117.2[3.0]	.00	4357
26-35	57.6 [1.6]	106.9[2.6]	.00	4840
36-45	61.3[1.5]	110.3[2.4]	.00	5864
46-64	57.7[1.2]	95.1[1.8]	.00	8126
65+	66.4[1.8]	75.7[2.3]	.00	4060
Male	75.1[1.2]	120.8[1.7]	.00	12597
Female	52.5[.7]	83.3[1.3]	.00	14650
Married	63.3[1.0]	106.5[1.6]	.00	12921
Single	61.9[.9]	96.5[1.4]	.00	14326
Metropolitan	60.7[.7]	98.1[1.1]	.00	23118
Non-metropolitan	73.7[2.4]	118.9[3.5]	.00	3927
Employed	56.5[.8]	106.4[1.4]	.00	16619
Unemployed	71.2[1.1]	92.5[1.7]	.00	10628

Notes: Values in the brackets are the standard deviation. P-value is determined from two-sample t-test, where null hypothesis is: $mean_{weekday} - mean_{weekend} \geq 0$.

Table 1.14: OLS: Determinants of Outdoor Time Difference Between Weekends and Weekdays

	Dependent var	
	outdoor leisure	total outdoor time
Household size	-3.39	0.94
	(4.30)	(4.64)
Weekly labor hours	-0.70*	-0.34
	(0.40)	(0.43)
Intercept	54.23***	51.06***
	(5.93)	(6.39)
N	617	620

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1.15: Summary Statistics: Number of Respiratory Disease Hospitalization for the Years 2004-2011

Day of the week	Age 5-19	Age 20-64	Age 65+	Total
Saturday	13,117	95,813	153,440	262,370
Sunday	13,938	96,423	150,316	260,677
Monday	18,439	131,433	204,473	354,345
Tuesday	17,369	126,332	195,780	339,481
Wednesday	15,883	122,829	188,992	327,704
Thursday	15,570	119,856	188,710	324,136
Friday	15,523	118,406	189,239	323,168
Total	109,839	811,092	1,270,950	2,191,881

Table 1.16: Summary Statistics: Pollution Levels for the Years 2004-2011

Day of the week	Ozone	CO	NO_2
Saturday	0.05	0.67	0.02
Sunday	0.05	0.62	0.02
Monday	0.04	0.67	0.02
Tuesday	0.04	0.69	0.02
Wednesday	0.04	0.70	0.02
Thursday	0.04	0.70	0.02
Friday	0.04	0.71	0.02

Table 1.17: Regression of Ozone on Respiratory Disease Hospitalization by Age, Without Controlling for the Weekend Effect

	Age 5-19	Age 20-64	Age 65+
Ozone (sum of lags)	1.23*	2.01***	0.22
	(0.73)	(0.02)	(11.24)
N	761150	761150	761150

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Coefficients from estimation are multiplied by a factor of 10^4. Standard errors are based on the robust variance estimator that are clustered by zip codes.

Table 1.18: Regression of Ozone on Asthma Hospitalization by Age, Without Controlling for the Weekend Effect

	Age 5-19	Age 20-64	Age 65+
Ozone (sum of lags)	1.50	0.83***	1.44
	(0.99)	(0.45)	(2.28)
N	445343	445343	445343

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Coefficients from estimation are multiplied by a factor of 10^4. Standard errors are based on the robust variance estimator that are clustered by zip codes.

Table 1.19: Regression of Ozone on Respiratory Disease Hospitalization by Age, with Controlling for the Weekend Effect

	Age 5-19	Age 20-64	Age 65+
Ozone (sum of lags)			
Saturday	2.77***	1.20***	-5.44
	(0.09)	(0.02)	(10.10)
Sunday	3.91***	.77	-21.25**
	(.67)	(1.44)	(9.75)
Monday	1.53	3.45	2.95
	(1.06)	(2.17)	(9.89)
Tuesday	2.55***	4.81***	-4.18
	(.34)	(.04)	(28.31)
Wednesday	.79	1.97***	-2.73
	(0.73)	(0.32)	(22.97)
Thursday	-1.36**	1.03***	20.20***
	(0.58)	(0.43)	(4.91)
Friday	.19	1.00	3.48
	(1.64)	(0.73)	(2.44)
χ^2	9.24	3.66	23.55
P-value	0.09	0.59	0.00
N	761150	761150	761150

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Coefficients from estimation are multiplied by a factor of 10^4. Standard errors are based on the robust variance estimator that are clustered by zip codes.

Table 1.20: Regression of Ozone on Asthma Hospitalization by Age, with Controlling for the Weekend Effect

	Age 5-19	Age 20-64	Age 65+
Ozone (sum of lags)			
Saturday	3.80***	1.32	1.61
	(0.71)	(0.85)	(3.18)
Sunday	2.48***	1.05	-0.81
	(.15)	(.99)	(4.85)
Monday	1.76	.49***	1.67
	(1.28)	(.11)	(2.04)
Tuesday	3.60***	.29***	.12
	(.45)	(.05)	(1.81)
Wednesday	.27	.59***	-.75
	(0.84)	(0.25)	(3.51)
Thursday	-.24	1.05**	3.22
	(1.11)	(0.56)	(2.07)
Friday	.43	1.08***	3.91
	(.78)	(0.36)	(2.85)
χ^2	4.47	5.35	8.62
P-value	0.48	0.37	0.13
N	445343	445343	445343

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Coefficients from estimation are multiplied by a factor of 10^4. Standard errors are based on the robust variance estimator that are clustered by zip codes.

Chapter 2

EFFECT OF CHILDHOOD EXPOSURE TO OZONE ON CHILDREN'S SKILL

2.1 Introduction

A notable number of studies recently examined the contemporaneous impact of pollution on public health [1] . The public burden of pollution forms one of the main bases for environmental policies. However, less is known about the long term impact of pollution on the formation of human capital. A few recent studies illustrate the long term impact of pollution on human capital and labor outcomes, yet there is still a lot to be learned in this area. The major goal of this study is to estimate the effect of exposure to ozone pollution on children's skill.

So far, the literature that examined the effect of pollution on children's skill mostly has focused on the most hazardous pollutants such as fine particles ($PM_{2.5}$). Almost nothing is known about the effect of ozone on children's skill, and this is what I focus on in this chapter. In addition to estimating the effect of ozone, I also control for the exposure to pollution from early childhood until teenage years, which is another novelty of this study. Previous studies that have examined the effect of pollution on cognitive/non-cognitive skills mostly focused on the effect of the early childhood exposure to pollution. While early childhood is an important stage of the child development process and a vulnerable period of children to be exposed to pollution, by focusing on this stage of life we do not learn about the rest of the child development process which can be critical in terms of child development and thus, important for policy.

[1] See, for instance Currie and Neidell (2005), Currie and Walker (2009), Currie *et al.* (2009), Vahedi (2013)

A well-known challenge with isolating the causal effect of the exposure to pollution levels on children's skills is an omitted variable bias. It is likely that neighborhood pollution levels are correlated with other neighborhood characteristics, such as crime rate or school quality. Without controlling for these neighborhood characteristics, the estimated effect of pollution will likely be biased. To deal with this issue, I use the instrumental variable that is introduced by Chay and Greenstone (2005). The instrumental variables they propose isolates exogenous variation in pollution generated by law enforcement and can be used to estimate the causal effect of pollution on children's skills. To estimate the model empirically, I combine the data from two sources. The Panel Study of Income Dynamics (PSID) provides a rich panel of a nationally representative sample of households. The Child Development Supplement (CDS), added to the original survey in 1997, that specifically focuses on children and provides rich data to study human capital formation. From the PSID and CDS I use information on children and their family characteristics along with the households' time investments in their children. To measure the exposure to pollution, I collected pollution data recorded via the monitors throughout the country from the US Environmental Protection Agency (EPA).

The OLS estimates suffer from omitted variable bias due to the endogenous nature of pollution measure. Based on health studies we expect that the effect of ozone on children's skill is negative or at least zero. OLS estimate of the effect is positive which is counterintuitive, even though it is statistically insignificant. To mitigate the omitted variable bias issue I use the IV estimation. This method accounts for the endogenous nature of pollution variable and the effect becomes negative. However, the effect of ozone on children's skill is marginally insignificant. This insignificant effect of ozone on children's skill is due to smaller sample and the loss of precision in the first stage of IV estimation.

The remainder of the chapter is organized as follows. Section 2 discusses the related studies. Data description is in section 3. Section 4 presents the econometric model. Section 5 discusses the estimation results. Section 6 discusses and concludes.

2.2 Prior Literature

The study of the long term impact of early childhood exposure to pollution has been hampered by the lack of appropriate data. Oftentimes the data does not have sufficient information about the households or the children. Even rich data does not solve the selection issue, and finding an exogenous variation in pollution becomes a challenge. To deal with these limitations, researchers have used quasi-experimental design or instrumental variable approach. For example, Sanders (2011) studies the impact of prenatal exposure to total suspended particulate (TSP) on educational achievement in high school. He utilizes county-level variation in timing and magnitude of sudden change in TSP levels that happened in response to industrial recession in the early 1980s. He claims that the dramatic change in TSP levels is strongly correlated with the industrial and manufacturing production. He uses this relationship to construct an instrumental variable (IV) for TSP. Specifically, Sanders (2011) defines an IV for TSPs as a relative share of county-level employment in manufacturing. The IV estimate is relatively larger than its Ordinary Least Square (OLS) counterpart: one standard deviation reduction in TSPs leads to a 6 percent increase in high school math scores when using the instrument, compared to the OLS estimate of only about 2 percent.

Almond *et al.* (2009) study school performance of children in Sweden who had been affected in utero during the Chernobyl disaster. The authors focus on children's achievement in the final year of compulsory school (age 16) and performance in high school (age 19). While being far from Chernobyl, Sweden did receive 5% of the nuclear fallout due to wind and weather conditions. Moreover, weather conditions generated a large variation in the amount of fallout among the affected regions in Sweden. This incident provided a natural experiment to study the impact of exposure to radiation on school performance of children. The study finds that the cohort affected by the fallout perform significantly worse in the final year of compulsory school and particularly in math. They also have a lower rate of high school graduation and lower GPA conditional on graduation.

Bharadwaj *et al.* (2014) examine the impact of exposure to air pollution during gestational trimesters on the educational performance in 4th grade in Santiago, Chile. Using sibling comparison and air quality alerts, the authors account for sorting and avoidance behavior. The idea is that by using sibling comparison they can control for factors that are correlated with pollution levels (through residential choice) and a child's educational achievement. For example, parents' income and their education can be important determinants of residential choice and have a direct impact on a child's educational performance. If people respond to air quality alerts, by controlling for these alerts, the authors take into account the subjects' avoidance behavior. The authors find a significant negative influence of pollution on math and language skills.

Evens *et al.* (2015) study the impact of lead concentration in whole blood (B-Pb) on educational performance. They examine the impact of blood lead on 3rd grade Illinois Standard Achievement Tests (ISAT) scores in Chicago public schools. After controlling for family income, demographics, and low birth weight or preterm-birth, the authors find that even the low blood lead levels [2] has a significant impact on educational performance. They find that $5\mu g/dL$ increase in

[2] B-Pb of $< 10\mu g/dL$

B-Pb in early childhood is associated with 32% increase in the risk of failing reading and math tests. Consistent with the previous studies, the results demonstrate that the impact of lead exposure is non-linear and it is steeper at lower levels.

All these studies focus on the effect of exposure to pollution in utero or during very early childhood on later educational outcome. Even though children are vulnerable to pollution during this period, but it is not the only period that is crucial in child development process. After this early stage, children still remain vulnerable to pollution, and the magnitude of the effect may vary by age and it is important to learn about this period as well. In the current study I control for exposure to pollution even as late as 15 years of age and estimate its effect on children's skill. Further, all these studies have focused on the most hazardous pollutants because of their sever negative effect that is found in the health and epidemiological studies. We do not know about the effect of ozone on educational outcomes which is the focus of the current study.

There are other studies that investigate the effect of childhood exposure to pollution on other type of adulthood outcomes such as wage and criminal activities. Isen *et al.* (2014) is one of the first studies that link childhood exposure to pollution directly to labor outcomes. The authors use the drastic change in TSP due to implementing the 1970 Clean Air Act Amendment (CAAA) to address the impact of childhood exposure to pollution on labor outcomes. Using the data from the Longitudinal Employer Household Dynamics File (LEHD) they compare the labor outcomes of those who were born right before the CAAA implementation with those who were born right after the policy implementation in counties that experienced a sharp change in TSP levels. The study finds that 10 unit decline in TSPs in the year of birth is correlated with 1% decrease in annual earnings of individuals in their late thirties. A back of the envelope calculation suggests that there is roughly \$6.5 billion lifetime earning gain for the entire cohort that were affected by CAAA.

Reyes (2007) studies the impact of childhood exposure to lead on criminal activities in adulthood. It has been previously shown that lead exposure has a negative impact on the development of the central nervous system and brain. Higher lead level is associated with behavioral disorders, such as aggressiveness, hyperactivity, and lack of emotional control. Reyes (2007) uses a variation of lead pollution levels among states over time due to removing lead from gasoline under the CAAA. She links the sharp drop of crime in 1990s to the decline of lead in the late 1970s and early 1980s to find that the phase-out of lead from gasoline explains 56% of the decline in violent crime in 1990s.

2.3 Data

Primary source of the data for this study is the Panel Study of Income Dynamics (PSID) and three waves of Child Development Supplement (CDS-I, CDS-II, and CDS-III). I also use pollution data from the US Environmental Protection Agency (EPA).

2.3.1 PSID and CDS

The PSID is a nationally representative longitudinal survey of the US individuals and families in which these individuals reside [3] . It provides a wide range of information on families and individuals. Since 1968 the PSID has collected data on family composition changes, housing and food expenditures, marriage and

[3] The PSID survey initially started in 1968 with a nationally representative sample of households. However, after the first wave of the survey in the following years, children from a household in the main sample who left the family and formed their own family were added as new households to the survey. Adding the new generation of the households into the survey and also dropouts from the main sample made the current sample unrepresentative of the national population. In order to fix this issue, PSID contains household's weights.

fertility histories, employment, income, health, consumption, wealth, and time spent on housework. The original PSID survey mainly focuses on households and particularly the head of the households, and then on spouses. From the main PSID survey, I use demographic information about the parents of children.

While PSID collects some information about children in the household, this information is quite limited. Starting in 1997, PSID added a new component, the CDS, that specifically focuses on children and collects detailed information on them. So far three waves has been administered: in 1997, 2002, and 2007. The data include but not limited to general school achievement information; the CDS also administers the subset of standard tests to assess academic skills of children. These tests include mathematics and language skills among other content areas. For this study I use the Letter-Word (LW) test scores as a measure of children's skill. The LW test is a subset of Woodcock-Johnson Revised (WJ-R) test of achievement that measures the symbolic learning and reading identification abilities of children at ages between 3 to 17. I use the raw scores for the LW test that is well-suited for examining changes in a child's performance on a WJ-R sub-test over time. [4]

The CDS also collects time use data on children for two days during a week: one weekday and one weekend day. Subjects fill out (with their caregiver if they are too young) a detailed time diary during these days. They provide information on what they have done (type of activities), where they have done these activities (location of activities), starting and ending time of activities (duration of activities), and who was with them during the activities over 24 hours. I use the time diary information to extract the time that children spend on developing their skills, either alone (e.g.

[4] The PSID also reports the standardized scores of the LW test that are normalized using a child's raw score, his age, and other children's scores in his age category. The standard scores are useful for cross-sectional comparison between different age groups. However, it is not useful to study changes in a child's performance over time. For further detail on the LW test see`https://psidonline.isr.umich.edu/Guide/default.aspx`

time at school or working alone on home works) or with their parents (e.g. studying with parents). Table 3.1 lists the variables that I use in the analyses, years of data, and their sources.

I focus on the PSID and CDS surveys that are administered between the years of 1997 and 2007. My target sample are children between 3 and 15 years old who have at least one LW test score. Even though the test is given to children between ages 3 and 18, I observe a dramatic unexplained drop in study time for children older than 16 and I drop those observations from the sample. I further restrict the sample based on the income data. I drop observations from families who had weekly income below \$100 as a minimum income required for a sustainable living in the absence of savings in the model. I also drop the high income observations with annual income above \$150000. These restrictions based on income remove 4% of the observations. [5]

Table 3.2 provides summary statistics of the data from the PSID and CDS. All the time variables are calculated in hours per week units and the family income is weekly income in 2000 dollars. Table 3.3 presents demographics variables for the CDS sample started at 1997.

Figure 3.1 shows the average LW test scores for every age group of children. The average test score increases at a declining rate by children's age. This observation is a crude measure of improvement in children's skill as they grow up. Figure 3.2 provides the average time that parents actively spend with their children and the time that children spend on their education alone for every age group. I define the parental time with their child "active" if at least one of the parents actively engaged in the activities that the child performs. Further, the education time alone is the time spent on school related activities such as time spent at school or time spent on homework

[5] Some of the families have zero or negative income where negative income corresponds to business or farm losses. There are only 22 observations with zero or negative income which is less than 1% of all the observations with non-missing income.

at home that the child does on his own. As demonstrated in figure 3.2, as children grow up they spend less time with their parents and more on school related tasks.

2.3.2 Pollution

The ideal air pollution data for this study would be the exact measure of pollution that a child inhales. Unfortunately, such detailed and precise measure of exposure to pollution is not available unless it is recorded in a lab experiment. In case of the United States, researchers normally use the measure of pollution that is recorded by the EPA via monitors throughout the country. Figure 3.3 shows the actual locations of monitors nationwide that are placed by the EPA to record variety of pollutants.

Among the multiple alternatives of pollution measures I use ground level ozone to control for environmental hazards because of its availability for the entire period of the CDS sample. Bad ozone is the one that is found in the Earth's lower atmosphere, near the ground level. The ground level ozone is not generated directly. It is formed from the chemical reactions between oxides of nitrogen (NOx) and volatile organic compounds (VOC) in the presence of sunlight and heat. NOx and VOC are emitted from different sources such as cars, power plants, industrial boilers, and refineries into the air. Previous studies have shown that ozone has negative effects on human health, specially on children and elderly. [6]

Using the centroid of households' census block and latitude and longitude of the monitors I find a nearest monitor to every household's location. Assigning a monitor that is very far from a household and may not be true representative of the pollution levels that the household is exposed to, can lead to measurement error. For that reason, I only keep those households that are within 20 miles of the nearest pollution monitor. Since the PSID survey after year 1997 is run biannually, I only observe

[6] For instance Currie and Neidell (2005) and Neidell (2004).

households' location every other year. So, I do not observe what pollution levels a household faces between the two surveys. In order to obtain more information from pollution data, if a household lives in the same census block in two consecutive surveys, I assume it was living in the same census block in-between.

Since the main data from the PSID is collected on yearly basis, the pollution data should match that annual pattern. Instead of reporting an annual measure collected from the monitors, the EPA generates so-called "Design Value". The design value is a statistic that the EPA generates to describe the air quality status at a particular location relative to the National Ambient Air Quality Standards (NAAQS). Based on the design value, the EPA determines if a particular monitor or, at a more aggregate level, a county is in attainment status or not. The monitor or county with the design value above the predetermined threshold is considered to be in non-attainment and these with the design value below the threshold are considered to be in attainment status. If a county is in non-attainment status, it is required to lower the pollution levels below the designated threshold. Under the Clean Air Act Amendment (CAAA) every year the EPA assigns a county attainment/non-attainment status.

In order to facilitate the interpretation of the results, I normalize the design value by its standard deviation. Figure 3.5 shows the distribution of the pollution for the pooled data. Figure 3.4 gives an idea about the variation of the pollution levels across the country in 2000 where the darker colors represent higher level of pollution.

2.4 Econometric Model

This section presents the empirical model to estimate the effect of childhood exposure to ozone on children's skill. I first use the OLS model to estimate the effect of ozone on children's skill

$$\theta_{i,t} = \eta_1 \theta_{i,t-5} + \eta_2 \tau_{i,pt} + \eta_3 \tau_{i,ct} + \eta_4 x_{i,t} + \eta_5 age_{i,t} + W_{i,t} + \epsilon_{i,t} \tag{2.1}$$

where i stands for child and t for time; θ is the child's skill; $\tau_{i,pt}$ is parents' time with the child; $\tau_{i,ct}$ is the child's time investment alone; $Age_{i,t}$ is the set of age dummies for the child; $W_{i,t}$ is a vector of other control variables and includes household income and parents' free time; and ϵ is the error term. Since the CDS data is only available for 1997, 2002, and 2007, the time lag in my model is five years instead of one. In other words, to control for initial stock of skills, I use a test score that is five years apart from the current score. $\tau_{i,pt}$ and $\tau_{i,ct}$ are measuring the investments by the child's parents and himself. Parents' free time is the total time available in a week minus sleeping time and labor hours.

Next, to mitigate these omitted variable biases I use instrumental variable approach. Following Chay and Greenstone (2005) and Bento *et al.* (2013) I use attainment status of a residence as an instrument for pollution levels. Under the CAAA if a county is in non-attainment status, it is required to lower the pollution levels below the designated threshold. Therefore, the attainment status of a residence can be used as an instrumental variable for pollution levels. Similar to Bento *et al.* (2013) I use the attainment status of monitors instead of counties. In real life, there is a variation in pollution levels among monitors even within a county. The EPA assigns attainment status of a county based on the pollution levels of the worst monitor within a county. As a result, local regulators tend to focus on dirtier monitors to coordinate with the EPA's standard. Therefore, most of the change happens in dirtier monitors, not all the monitors within a county. Hence, using the attainment status of monitors captures the pollution levels change better than county attainment status. Further, I use the attainment status with one year lag as the instrumental variable, since if a county is in non-attainment status, it is required to bring the pollution levels below the designated threshold in the upcoming year.

First and second stage equations for the IV analysis are given by

$$\theta_{i,t} = \eta_1\theta_{i,t-5} + \eta_2\tau_{i,pt} + \eta_3\tau_{i,ct} + \eta_4\hat{x}_{i,t} + \eta_5 age_{i,t} + W_{i,t} + \epsilon_{i,t}$$

$$x_{i,t} = \varrho_1 z_{i,t} + \Pi Y_{i,t} + \xi_{i,t} \tag{2.2}$$

Where the first equation is the same as the OLS regression model. In the second equation $z_{i,t}$ is the attainment status dummy of the monitor near child i at year t; $Y_{i,t}$ is the remainder of the control variables from the first equation; and $\xi_{i,t}$ is the error term.

2.5 Estimation Results

This section presents estimation results of equations (2.1) and (2.2). In the second column of table 2.4 I use only pollution level and children's age as independent variables. In third column I add the investment variables, and in the last column I estimate the full model. Not surprisingly, stock of children's skill, children's time investments on education alone, and household's income are positively correlated with the children's skill and are statistically significant. In the full model, Column 4, the coefficient on the lagged test score implies that children who scored one point higher on previous test, on average score 0.192 points higher on consecutive test. Children time investment on education is positively related to test scores: one more hour per week is associated with 0.0325 points increase in test score. Increasing household weekly income by 100 dollars is associated with 0.3 points higher test score. To control for the unobserved time-invariant characteristics of individuals, I use first difference approach. Table 2.5 presents the results of the estimation. As expected, the standard deviations of the estimates are large, and the effect of ozone is not statistically significant.

In all the specification the coefficient for pollution levels is positive and insignificant. This positive coefficient becomes much smaller in the last column when I control for the investment variables, household's income, and parents' labor hours. The OLS coefficient however does not represent a causal effect of pollution on test scores because of the omitted variable bias. As mentioned above, one potential source of the omitted variable bias are unobserved neighborhood characteristics that are correlated with both pollution levels and test scores, such as crime rate or school quality.

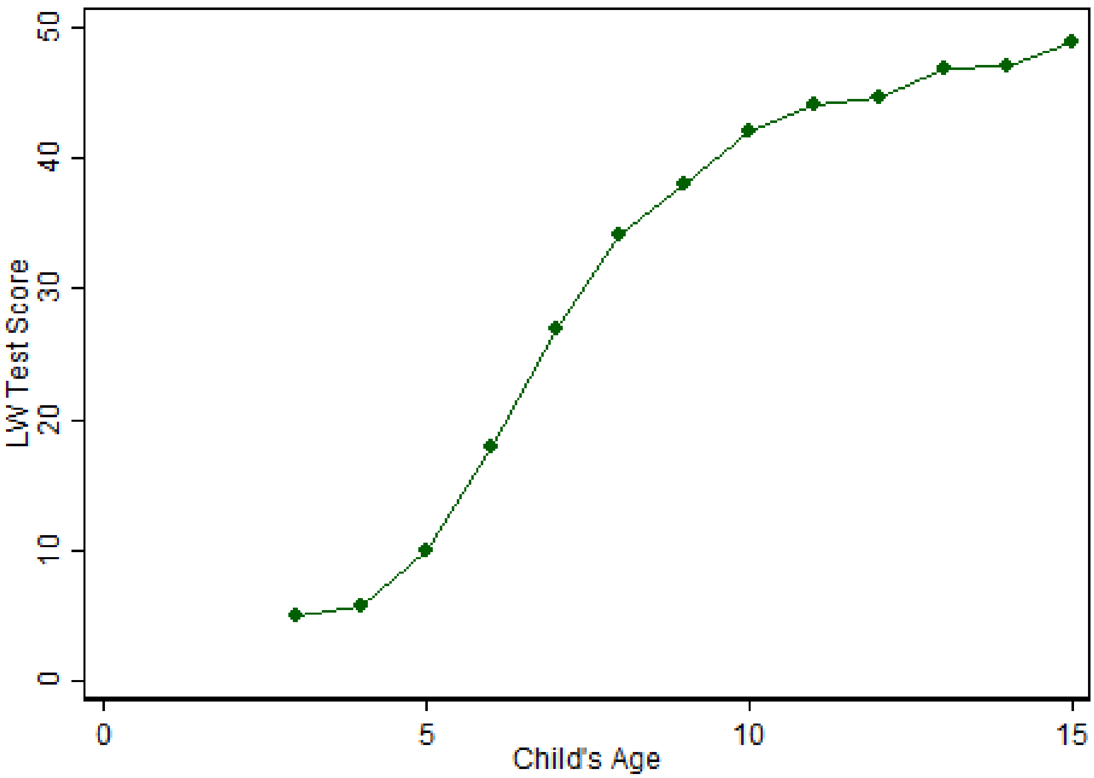
Instrumental variable estimates presented in Table 2.6 plausibly identify the causal effect of pollution on children's skill. The first stage result supports the relevance of attainment status as an instrumental variable. In the second stage the coefficient of pollution is negative but statistically insignificant. One percent increase in pollution levels lowers the children's test score by 0.3 percent. As evidenced by the results, instrumental variable partially solves the omitted variable bias - the estimate of the pollution effect has the expected sign, but suffers from low precision. Large standard error of the estimates in part can be due to small sample. Related, standard errors of the IV estimates are always larger than the OLS estimates due to the loss of precision in the first stage.

2.6 Discussion and Conclusion

In this study I estimated the effect of exposure to ozone on children's skill. I estimated this effect both using OLS and IV estimation to control for endogeneity of pollution measure. I use a panel data from the Panel Study of Income Dynamics and Child Development Supplement for test scores, time investment, and demographics of children and their family. I merge these data sets with the measures of ozone index from the Environmental Protection Agency.

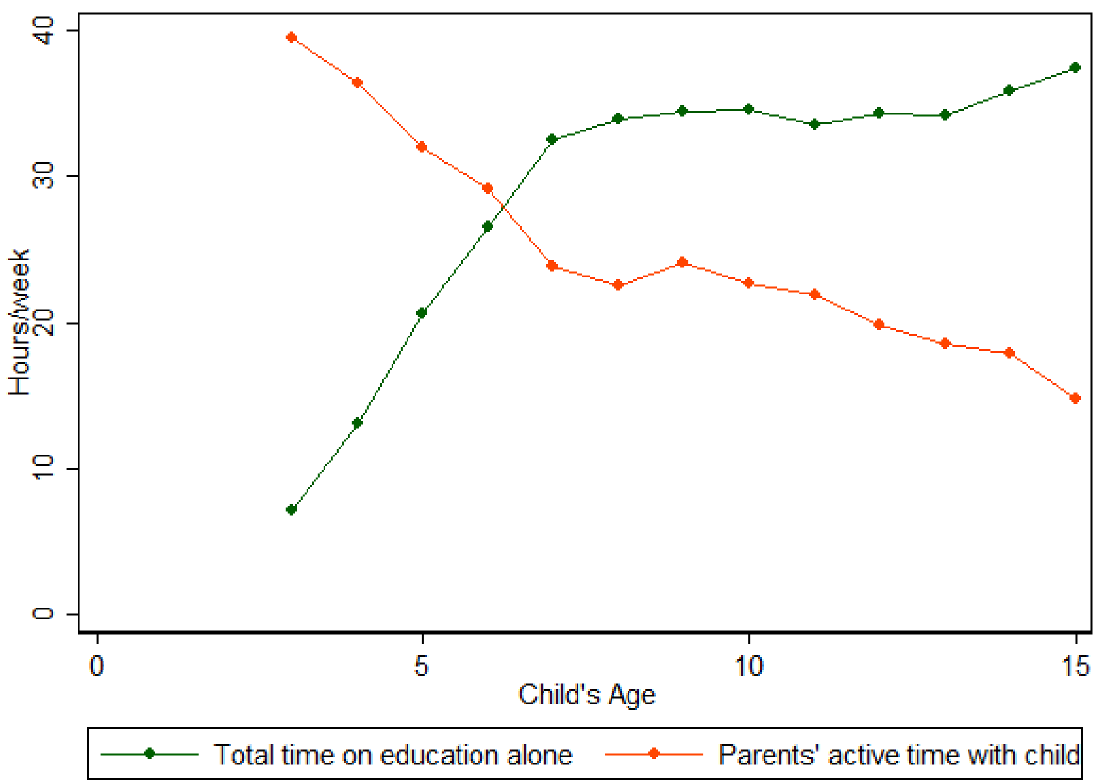
Because of the endogeneity problem effect of ozone on children's skill from OLS estimation is counterintuitive in sign and not significant. Instrumental variable approach mostly corrects for the endogeneity problem, and results of IV estimation demonstrate the negative effect of ozone on children's skill. However, the coefficient is imprecisely estimated. The low precision of the IV estimate is because of both small sample and the fact that IV estimation in general produces large standard errors due to the loss of precision in the first stage. In the next chapter I use structural estimation approach to improve on this issue and to provide a framework to analyze policy counterfactuals.

Figure 2.1: Average LW Test Scores by Age



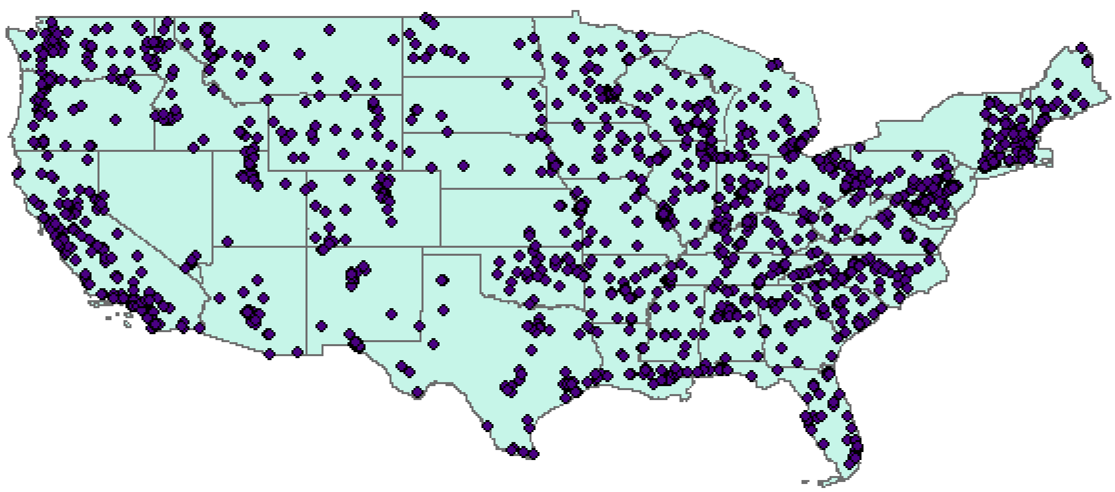
Source: The LW test scores comes from the PSID-CDS.

Figure 2.2: Average Child's Time on Education Alone



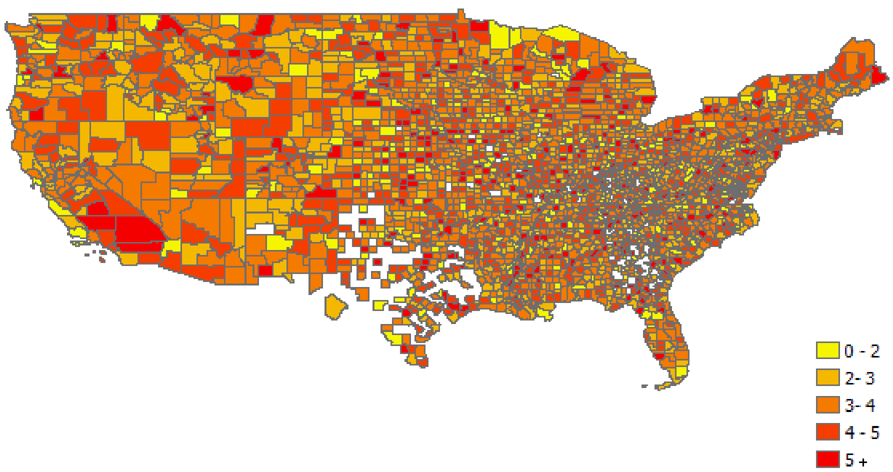
Source: The time diary information comes from the CDS.

Figure 2.3: Pollution Monitors' Location



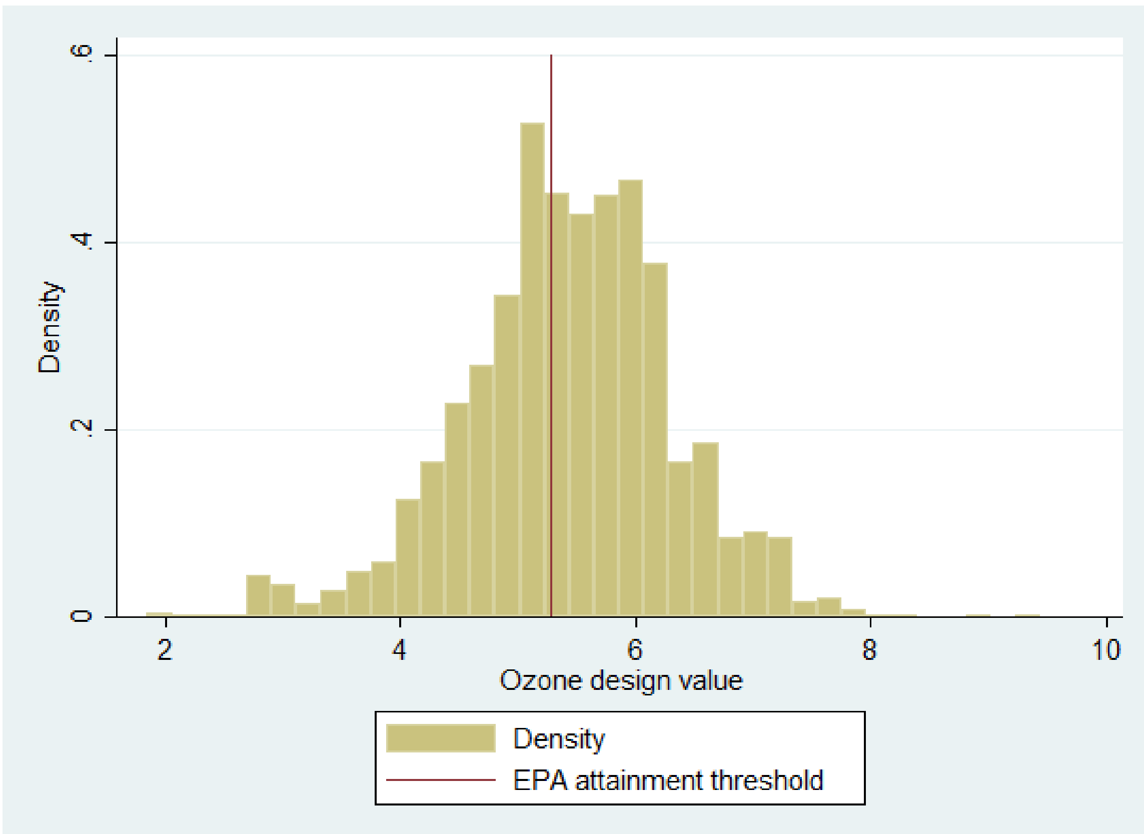
Source: The exact geographical location of the monitors is from the EPA.

Figure 2.4: Ozone Average by County for Year 2000



Source: Ozone data comes from the EPA.

Figure 2.5: Pollution Distribution



Source: Ozone data comes from the EPA.

2.8 Tables

Table 2.1: Data Sample

	Used variable from the data	Years	Source
I_t	Annual family income	1996,98,2000,02,04,06	PSID
h_t	Parents' labor hours	1996,98,2000,02,04,06	PSID
θ_t	Letter-Word score	1997,2002,07	CDS
τ_t^p	Active time parents spend with child	1997,2002,07	CDS
τ_t^c	Time that child spend at school study alone	1997,2002,07	CDS
x_t	Pollution	1997-2007	EPA

Table 2.2: Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Family income	971.98	600.29	100.06	2882.64	5302
LW Test Score	35.98	14.71	1	57	1940
Total time on education alone	30.97	15.9	0	88.5	1899
Parents' active time with child	23.74	14.9	0	143.5	1899
Ozone design value	5.43	0.89	1.85	9.45	4086

Notes: This table shows the sample's characteristics at the beginning of the study, year 1997.

Table 2.3: Summary Statistics for Sample at 1997

Variable	Mean	Std. Dev.	N
Mothers education (years)	14	1.9	854
Family size	2.2	0.9	958
Mothers age at first birth	24.28	5.74	855
Family income ($/week)	899.39	589.11	851

Notes: This table shows the sample's characteristics at the beginning of the study, year 1997.

Table 2.4: OLS Results. Dependent Variable: LW Test Score

	(1)	(2)	(3)
Ozone design value	0.231	0.244	0.0688
	(0.325)	(0.307)	(0.304)
Lagged test score		0.175***	0.192***
		(0.0234)	(0.0231)
Parents' active time with child		0.0219	0.00635
		(0.0144)	(0.0135)
Child's time on education alone		0.0347**	0.0325**
		(0.0108)	(0.0104)
Household income			0.00251***
			(0.000371)
Parents' free time			0.0289
			(0.0225)
Intercept	4.433*	-0.429	-1.643
	(1.957)	(2.205)	(2.834)
N	1940	1940	1940

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

All the specifications control for child's age.

Table 2.5: First Difference Estimate. Dependent Variable: LW Test Score

Independent Variables	Coefficient
Ozone design value	1.548
	(1.818)
Parents' active time with child	-0.052
	(0.066)
Child's time on education alone	0.123*
	(0.062)
Household income	0.002
	(0.002)
Parents' free time	-0.014
	(0.121)
Intercept	17.580***
	(1.132)
N	761

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

All the variables represent first differences.

Table 2.6: IV Estimates of Pollution on LW Test Score

First stage	Ozone design value
Attainment status	-1.415***
	(.0532)
Second stage	LW test score
Ozone design value	-0.273
	(0.391)
Lagged test score	0.193***
	(0.0229)
Parents' active time with child	0.00472
	(0.0135)
Child's time on education alone	0.0329**
	(0.0104)
Household income	0.00258***
	(0.000372)
Parents' free time	0.02960
	(0.0227)
Intercept	38.19***
	(2.953)
N	1940

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

It is controlled for child's age in the regression.

Chapter 3

CHILDREN'S SKILL FORMATION: THE ROLE OF CHILDHOOD EXPOSURE TO POLLUTION

3.1 Introduction

In previous chapter I estimated the effect of exposure to ozone on children's skill using instrumental variables approach. In general, the reduced form model is helpful to estimate the effect of exposure to ozone on children's skills in the presence of endogeneity and selection. However, such model does not account for the potential behavioral response of individuals. As a result, we cannot uncover the underlying mechanism of the pollution effect, nor can we estimate the policy counterfactuals. In this chapter, I develop a structural model that links children's exposure to pollution to their educational outcomes. I model an optimization problem of altruistic households that value their children's skill level and contribute to the development of their children through monetary and time investments. To the best of my knowledge, this is the first study that attempts to estimate the skill formation technology taking into account exposure to pollution.

Using the theoretical model to address this question has multiple benefits. This model is well suited to study the mechanism of pollution effects on children's skill as well as to conduct counterfactual analyses. I jointly estimate the parameters of the model which makes it feasible to identify the effect of the unobserved monetary investment. I also control for the exposure to pollution from early childhood until the teens, which is another novelty of this study. Previous studies that have examined the effect of pollution on cognitive and non-cognitive skills mostly focused on the effect of the early childhood exposure to pollution. While early childhood is an important stage of the child development process and a vulnerable period of children to be exposed to pollution, by focusing on this stage of life we do not learn about the rest of the skill development process which can be critical in terms of child development and thus, important for policy.

A well-known challenge with isolating the causal effect of the exposure to pollution levels on children's skills is an omitted variable bias. It is likely that neighborhood pollution levels are correlated with other neighborhood characteristics, such as crime rate or school quality. Without controlling for these neighborhood characteristics, the estimated effect of pollution will likely be biased. To deal with this issue I use the instrumental variable that is introduced by Chay and Greenstone (2005). This instrumental variable isolates exogenous variation in pollution generated by law enforcement and can be used to estimate the causal effect of pollution on children's skills. To estimate the model empirically, I combine data from two sources. The Panel Study of Dynamics (PSID) provides a rich panel on a nationally representative sample of households. Since 1997 the Child Development Supplement (CDS) has been added to the original survey that specifically focuses on children and provides rich data to study human capital formation. From the PSID and CDS I collect information on children and their family characteristics along with the households' time investments in their children. To measure the exposure to pollution, I collected pollution data recorded via the monitors throughout the country from the US Environmental Protection Agency (EPA).

Previous studies have focused either on parental investment in a children's skill (estimates of a child's skill formation technology in child development literature) or on the effect of pollution on these skills (environmental studies). To the best of my knowledge, the structural model I develop in this study is the first model that combines two approaches to estimate the effect of pollution on children's skills while accounting for parents' investment. Omitted variable bias explained above is not completely solved in the OLS estimation. While the instrumental variables strategy can potentially deal with the bias, it does not allow for the examination of the mechanism of the pollution effect, nor it allows to conduct counterfactual analysis. In this study, I control for the omitted variables bias using an instrumental variable, and my structural model enables me to run counterfactual analysis. In addition, environmental studies that use quasi-experimental design to estimate the pollution effect, usually focus on a particular age group of children. In this study, I estimate the heterogeneous effect of pollution on children's skills at different ages.

Consistent with previous studies, my results demonstrate the negative effect of pollution on children's skills and that effect varies by age. I find that on average a one standard deviation decline in pollution levels that children are exposed to, during the development process, increases their test scores by 0.07 standard deviation. This number is 0.10 standard deviation for 3 year olds and 0.04 standard deviation for 14 year old children. This result implies that children are more vulnerable to pollution in early childhood, which is consistent with findings in health and epidemiological literature.

Using the estimation results, I run two counterfactual experiments: pollution reduction, and income transfer to the households. In the first experiment, I estimate the effect of reducing the stream of pollution that children are exposed to by one standard deviation at every age of the development process. As a result of this change, the average test scores among all age groups increases by 0.09 standard deviation. If I were to ignore the households' behavioral response to the change in pollution levels in the model, this effect would go up as high as 0.13 standard deviation. Larger effects when the behavioral channel is shut down, indicate that households compensate in response to pollution change.

In an income transfer experiment, households receive a transfer of \$600 per week at every age of children's development period which corresponds to one standard deviation of income. This transfer on average increases the test score by 0.12 standard deviation. The effect of an income transfer accrues over time and leads to a larger effect for 15 year old children due to cumulative effect through stock of a child's skill. If households invest the entire income transfer on children's skills, the effect would be larger. However, the transfer is split up between consumption and monetary investment on children. To compare the results of the two experiments, the effect of one standard deviation reduction in pollution levels on test scores is the same as the effect of a weekly income transfer of \$120, or \$6240 per year. In addition to a pollution effect, the results also emphasize the importance of monetary and time investments on children's skills.

The remainder of the chapter is organized as follows. Section 2 explains the theoretical model. Data description is in section 3. Section 4 discusses the estimation issues and identification strategy. Section 5 presents the estimation results. Section 6 presents counterfactual analysis. Section 7 discusses and concludes.

3.2 Model

In order to estimate and predict the effect of pollution on skill formation and long-term outcomes, I build a conceptual model that is closely related to Cunha and Heckman (2007) and Boca *et al.* (2014). In the following sections I explain the structure of the model, present its solution, and demonstrate the mechanisms through which pollution affects children's skill formation.

3.2.1 Timing and Household's Preference

Economy consists of homogeneous households, where each child is born with initial stock of skills denoted by θ_0. Time t is modeled as discrete with finite horizon. The child development process takes $M+1$ periods, i.e. the timing is $t = 0, 1, \ldots, M$. At the beginning of every period a household knows its child's skill level, θ_t, household's income, I_t, parents' labor hours, h_t, and pollution level, x_t. The unitary household optimally allocates its time between leisure and investment on child's skill. The leisure time includes parents' leisure, l_{pt}, and child's leisure time, l_{ct}. Time investment on child's skill includes time that parents spend with their child, τ_{pt}, and time that children invest in their skill alone, τ_{ct}, e.g. school time [1]. Further, a household allocates its income on consumption of a composite good, c_t, and monetary investment on child's skill, e_t. There is no labor choice in the model. In every period, parents' labor hours and household's income is exogenously realized. However, these stochastic processes are correlated through the initial condition and I will elaborate on them in section 3.2.4.

Every period the household receives utility from consumption, parent's leisure time, child's leisure time, and child's skills level. I assume a simple Cobb-Douglas functional form for the household's utility function [2]. The household's preference is represented by

$$U(c_t, l_{pt}, l_{ct}, \theta_t) = \alpha_1 \ln c_t + \alpha_2 \ln l_{pt} + \alpha_3 \ln l_{ct} + \alpha_4 \ln \theta_t \quad t = 1, \ldots, M-1, \tag{3.1}$$

[1] Boca *et al.* (2014) divide the time spending with children into two groups of active and passive time. In this study I only use one category of parental time. I include any parental time such that at least one of the parents are actively engaged in doing a activity with the child as productive time.

[2] Alternatively, I could have used more general form of the Constant Elasticity of Substitution (CES) model. While the CES allows complementarity between variables, it is not the focus of this study. Also, Cobb-Douglas functional form simplifies the model.

I normalize the input weights to add up to one, $\sum_j \alpha_j = 1$, and $\alpha_j > 0$ for $j = 1, \ldots, 4$. At period M when the child leaves the household, in addition to the flow utility similar to previous periods, the household receives utility from the stock of its child's skill at the end of the period M. This extra utility is $\beta\varphi \ln \theta_{M+1}$, where β is the discount factor and the household weights the last period's skill by the factor of φ. So, the household's utility in the last period is

$$U(c_M, l_{pM}, l_{cM}, \theta_M) = \alpha_1 \ln c_M + \alpha_2 \ln l_{pM} + \alpha_3 \ln l_{cM} + \alpha_4 \ln \theta_M + \beta\varphi \ln \theta_{M+1}, \quad (3.2)$$

3.2.2 Children's Skill Formation Technology

A household faces a trade-off when making investment in its child's skills: investing in children, it has less resources for consumption and leisure, but at the same time, this investment boost its child's skill and generated higher utility derived from the increased child's skill. I model skill technology as the Cobb-Douglas form with some modification: to allow the decision variables of a household to potentially be correlated with the pollution levels, I interact investments with the pollution variable. Child's skill, θ, evolves over time according to the following technology [3] :

$$\ln \theta_{t+1} = \ln R_t + \delta_{1,t} \ln \theta_t + \delta_{2,t} \ln \tilde{e}_t + \delta_{3,t} \ln \tilde{\tau}_{pt} + \delta_{4,t} \ln \tilde{\tau}_{ct} + \delta_{5,t} \ln x_t + \ln u_t \quad (3.3)$$

s.t.

$$\tilde{e}_t = e_t + p_1 e_t x_t + p_2,$$

$$\tilde{\tau}_{pt} = \tau_{pt} + q_1 \tau_{pt} x_t + q_2,$$

$$\tilde{\tau}_{ct} = \tau_{ct} + r_1 \tau_{ct} x_t + r_2,$$

$$t \in 0, 1, \ldots, M$$

[3]In principal θ can include cognitive and non-cognitive skills/abilities. However, in this study I assume a one-dimensional skill.

A child's skill at the end of period t, θ_{t+1}, depends on total factor productivity (TFP), R_t, child's stock of skill at the beginning of the period, θ_t, monetary investments, e_t, parental time with the child, τ_{pt}, child's alone time investment on skill alone, τ_{ct}, pollution levels that a child is exposed to, x_t, and error term, u_t. In the equation (3.3) and in the constraints δ's, p's, q's, and r's are the production function parameters. The simple interactions between pollution levels and the decision variables add the complementarity between them both in the level and growth of children's skill. More importantly, it makes the model more general and allows the optimal decision variables to potentially depend on the pollution levels. Without the interaction terms the model forces the decision variables to be independent of pollution levels. By making the model more general via adding this interaction, the pollution can either directly affect the child's skill, or indirectly by affecting the household's investment behavior. Later I explain these two channels in more detail. The error term in the technology function includes any other factors that are not controlled in the model. These factors are uncorrelated with the rest of the variables in the model, even though they might directly affect children's skill.

In equation (3.3) entails complementarity and substitutability between all the inputs. For example, marginal effect of any investment variable is affected by the level of pollution that a child is exposed to, or the parental time investment can be substituted by monetary investment. In equation (3.3) the marginal effect of the stock of children's skill is called self-productivity of the technology function and is defined as $\frac{\partial \theta_{t+1}}{\partial \theta_t}$. If this derivative is positive it means that starting a period with high skill level leads to accumulating more of skill with the same levels of investments. The self-productivity effect is the dynamic component of the the model that enables accumulation of child's skill over time. This dynamic component carries the effect of investments and various shocks from one period to consecutive periods. Stronger connection between periods through the self-productivity leads to higher persistence of these effects.

3.2.3 Household's Dynamic Problem

The household's value function is

$$V_t(\theta_t, S_t) = \max_{\substack{\tau_{pt},\tau_{ct},l_{pt},\\ l_{ct},c_t,e_t}} U(c_t, l_{pt}, l_{ct}, \theta_t) + \beta E[V_{t+1}(\theta_{t+1}, S_{t+1}|S_t)] \tag{3.4}$$

subject to

$$S_t = (I_t, h_t, x_t, u_t)$$

$$\tau_{pt} + l_{pt} = T - L_t$$

$$\tau_{pt} + \tau_{ct} + l_{ct} = T$$

$$c_t + e_t = b_t$$

where T is the total time available. S_t is the vector of state variables that includes household's income, I_t, parents' labor hours, L_t, pollution levels, x_t, and the error term, u_t, and it is realized at the beginning of the period t before making decision. The second constraint implies that parents spend the time that is left after work either with their child or on leisure. The third constraint describes the allocation of child's own time on spending his time with parents, on investing on skill alone, or on leisure. The last constraint is the household's budget constraint: the household spends its income either on consumption of the composite good or investing on its child. [4] Labor hours, income, and pollution levels are exogenous and stochastic and E is the conditional expectations operator with respect to I_{t+1}, h_{t+1}, x_{t+1}, and u_{t+1}.

[4]Every period a household exogenously receives non-labor income. Having the labor choice allows more flexibility in household's decision process and add another trade off between working and spending time with child. It can also affect policy analysis as well. However, in the current version of the study for simplicity I abstract from the labor choice.

In the last period of child development, M, the household maximizes the value function that only depends on the flow utility as following:

$$V_M(\theta_M, S_M) = \max_{\substack{\tau_{pM},\tau_{cM},l_{pM},\\ l_{cM},c_M,e_M}} \alpha_1 \ln c_M + \alpha_2 \ln l_{pM} + \alpha_3 \ln l_{cM} + \alpha_4 \ln \theta_M + \beta\varphi \ln \theta_{M+1} \quad (3.5)$$

There is no uncertainty in the last period of child development and the household faces the similar set of constraints as previous periods.

3.2.4 Heterogeneity and Sources of Uncertainty

Households in the model differ from each other in five dimensions which are the sources of heterogeneity in the model: child's initial skill level, household's income, parents' labor hours, pollution levels, and the technology error term. Besides these values, all the households are homogeneous in terms of preferences, skill formation function, and the law of motion for income, parents' labor hours, and pollution processes. Initial values for child's skill level, household's income, parents' labor hours, and pollution levels are drawn jointly from a normal distribution:

$$(\theta_0, I_0, h_0, x_0) \sim LogNormal(\mathbf{M}, \Sigma),$$

Child's skill evolves over time according to equation (3.3). Household's income, parents' labor hours, pollution levels, and the error terms are the sources of the uncertainty in the model. Amount of income, labor hours, and pollution levels that the household faces in each period are determined by stochastic processes modeled as follows:

$$\ln \mathrm{I}_{t+1} = \omega_0 + \omega_1 \ln \mathrm{I}_t + \varepsilon_{t+1}, \quad \varepsilon_t \overset{iid}{\sim} Normal(0, \sigma_\varepsilon)$$

$$\ln h_{t+1} = \rho_0 + \rho_1 \ln h_t + \epsilon_{t+1}, \quad \epsilon_t \overset{iid}{\sim} Normal(0, \sigma_\epsilon)$$

$$\ln x_{t+1} = \gamma_0 + \gamma_1 \ln x_t + \gamma_2 z_t + \xi_{t+1}, \quad \xi_t \sim Normal(0, \sigma_\xi)$$

where z_t is the instrumental variable for pollution. Pollution levels that the household faces every period comes from an exogenous process. The underlying assumption is that the location choice is not part of the household decision variables. This implies that the model does not control for the residential sorting. If pollution levels is positively correlated with other characteristics of the neighborhood that a household lives and negatively affect child's skill, then my estimation of pollution effect on child's skill will be an upper bound. For example, crime rate can be higher in neighborhoods with high levels of pollution. Children in these neighborhood may perform poorly not only because of exposure to high pollution levels, but also negative effect of high crime rate. Since, I do not allow households to choose their location and therefore pollution levels, I will attribute all the negative effects of neighborhood only to pollution. The opposite is true if the neighborhood characteristics that are correlated with pollution levels have positive effect on children's test score, such as school quality.

To control for the omitted variables bias I use the instrumental variable that is introduced in Chay and Greenstone (2005). The authors use the attainment status of counties that is assigned by the EPA based on the pollution levels of the counties. If a county is assigned to be in non-attainment status it has to lower the pollution levels for the next year by the law. The idea is that pollution changes exogenously by the law enforcement and if the household composition of the county does not change within a year this exogenous change in pollution levels can be used to estimate the causal effect of pollution on children's test score. To account for this exogenous change in pollution due to government policy I include the attainment status, z_t, in the pollution process. Using the instrumental variable in the pollution process helps to satisfy the orthogonality assumption between the technology error, u_t, term and the pollution levels, x_t, in the technology function.

In the model I abstract from parents' labor decision because the main focus of this study is child development and not individuals labor decision. Therefore, labor hours and household's income are exogenously determined. One way to look at this issue is to drop the labor hours from the model and simply assume that every parents have T hours to spend on leisure or child development. However, this may cause bias in estimating the effect of pollution on child's skill if the labor hours is correlated with pollution and child's skill. [5] Instead, without complicating the model and endogenizing labor choice, I make the model richer by including the exogenous labor hours process in the model. As long as the exogenous process of labor hours, to a reasonable extent, captures the reality in the data I can be confident that this assumption will not bias the estimation.

3.2.5 Model Solution

Given the functional forms of the utility function and the children's skill formation technology, there is a closed form solution for the optimal decision variables [6] . The optimal value for the monetary investment in child's skill is

$$e_t^* = \frac{\beta\delta_{3,t}(1+p_1x_t)A_tI_t - \alpha_1 p_2}{(1+p_1x_t)(\alpha_1+\beta\delta_{3,t}A_t)}, \tag{3.6}$$

The monetary investment is directly related to income level and negatively related to the relative weight of consumption in the utility function, α_1. If any of p_1 and p_2 take value of zero, e_t^* will be independent of x_t, i.e. monetary investment on child's skill will not be correlated with the pollution levels.

[5] This is true for household income, too. One could assign the same level of income for every household and remove the income process from the model.

[6] I allow the corner solution in my model, however because of the Cobb-Douglas form of the preferences, the decision variables will not take their maximum possible values, i.e. $e_t^* < b_t$ and $\tau_{pt}^*, \tau_{ct}^* < T$. If any of the decision variables take their maximum possible values, then one of the elements in the preference function is zero and the utility function is not defined at zero. However, the decision variables can be zeros.

The optimal value of parental time, τ_{pt}, is a solution of the quadratic equation of $a\tau_{pt}^2 + b\tau_{pt} + c = 0$ where a, b, and c are functions of the model parameters and exogenous variables. [7] The model therefore can have multiple solutions for the time investments. By definition, τ_{pt}^* has to be positive. However, there is no analytical solution of the ranges for the structural parameters that guarantee value of τ_{pt}^* to be positive. In the numerical solution I make sure that τ_{pt}^* takes positive values and choose the answer that gives the highest utility level in case of multiple solutions. If any of q_1 and q_2 take value of zero, τ_{pt}^* will be independent of x_t. The same holds for the values of r_1 and r_2 and therefore τ_{ct}^*. Hence, if $p_1p_2 = q_1q_2 = r_1r_2 = 0$, then this model will nest the simple model in which decision variables are independent of pollution levels and excludes the endogeneity of these variables with respect to pollution.

3.2.6 Pollution Impacts

Potentially, pollution can affect a child's skill through two channels. In the first channel, pollution affects a child's skill directly through health. Health effects of pollution can be mild (headache and tiredness), or severe (such as asthma attack and long lasting brain damage from exposure to pollution in early childhood). In turn, poor health can affect a child's productivity. For example, student's poor performance at the school due to lack of attention or tiredness. Severe health shocks such as nervous system damage can have a persistent and long term impact on a child's performance. [8] I will refer to this channel as direct impact.

In the second channel, pollution affects a child's skill indirectly through a household's investment. In response to changes in pollution levels, household can

[7] See 3.10.

[8] Reyes (2007)

adjust their investment behavior accordingly. This adjustment can either mitigate the negative effect of pollution or exacerbate it depending on the direction of the adjustment. I will refer to this channel as indirect impact. Equation (3.7) demonstrates both effects of pollution. For a general form of the skill formation technology $\theta_{t+1} = f(\theta_t, e_t, \tau_{pt}, \tau_{ct}, x_t)$ the the partial derivative of θ_{t+1} with respect to x_t is

$$\frac{\partial \theta_{t+1}}{\partial x_t} = \underbrace{\frac{\partial f_t}{\partial e_t^*}\frac{\partial e_t^*}{\partial x_t} + \frac{\partial f_t}{\partial \tau_{pt}^*}\frac{\partial \tau_{pt}^*}{\partial x_t} + \frac{\partial f_t}{\partial \tau_{ct}^*}\frac{\partial \tau_{ct}^*}{\partial x_t}}_{IndirectImpact} + \underbrace{\frac{\partial f_t}{\partial x_t}}_{DirectImpact} \tag{3.7}$$

The asterisks indicate the optimal values of the decision variables. The three terms on the right hand side of equation (3.7) represent the indirect effect of pollution. In the direct impact, pollution affects a child's skill in the absence of any type of behavioral response. Majority of the epidemiological and health studies focus on the direct impact of pollution.

Combining equations (3.7) and (3.3) together, the direct and indirect impact of pollution on a child's skill for the specific case can be expressed as:

$$\text{Direct Impact} = \frac{\partial f_t}{\partial x_t} = \theta_{t+1}[\frac{\delta_{3,t} p_1 e_t^*}{\tilde{e}_t^*} + \frac{\delta_{4,t} q_1 \tau_{pt}^*}{\widetilde{\tau}_{pt}^*} + \frac{\delta_{5,t} r_1 \tau_{ct}^*}{\widetilde{\tau}_{ct}^*} + \frac{\delta_{6,t}}{x_t}], \tag{3.8}$$

$$\text{Indirect Impact} = \theta_{t+1}[\frac{\delta_{3,t}(1+p_1 x_t)}{\tilde{e}_t^*}\frac{\partial e_t^*}{\partial x_t} + \frac{\delta_{4,t}(1+q_1 x_t)}{\widetilde{\tau}_{pt}^*}\frac{\partial \tau_{pt}^*}{\partial x_t} + \frac{\delta_{5,t}(1+r_1 x_t)}{\widetilde{\tau}_{ct}^*}\frac{\partial \tau_{ct}^*}{\partial x_t}] \tag{3.9}$$

According to equation (3.7) the net impact of pollution on a child's skill is the sum of direct and indirect impacts. The direction of the net effect (negative or positive) depends on the sign and magnitude of the parameters. Predictions of the theory are ambiguous with respect to pollution effect on a child's skill. To fix ideas, let's assume that there is no behavioral response with respect to pollution levels. In the model this assumption holds if $p_1 p_2 = q_1 q_2 = r_1 r_2 = 0$ and this will lead to $\frac{\partial e_t^*}{\partial x_t} = \frac{\partial \tau_{pt}^*}{\partial x_t} = \frac{\partial \tau_{ct}^*}{\partial x_t} = 0$. Substituting these values in the equations (3.8) and (3.9) leads to

$$\frac{\partial \theta_{t+1}}{\partial x_t} = \text{Direct Impact} + \text{Indirect Impact} = \theta_{t+1} \frac{\delta_{6,t}}{x_t} \tag{3.10}$$

Thus, the sign of the net pollution effect only depends on the sign of $\delta_{6,t}$. Based on the previous epidemiological and health studies we expect $\delta_{6,t}$ to be negative. In other words, when households do not adjust their behavior, pollution will negatively affect children's health. However, the assumption of $p_1 p_2 = q_1 q_2 = r_1 r_2 = 0$ is a specific case of the general model and, in fact, these parameters can take any non-zero values. That is, households can react to the pollution levels and accordingly choose the values of the decision variables and by that means either mitigate or even intensify the negative effect of pollution depending on the values of the parameters.

Apart from non-behavioral response, there are two other possible cases: compensatory and reinforcing responses. The compensatory response is when the household mitigates the negative effect of pollution. In this case, pollution still negatively affects the child's skill but to a lesser degree than in the case of non-behavioral response, i.e. $\theta_{t+1} \frac{\delta_{6,t}}{x_t} < \frac{\partial \theta_{t+1}}{\partial x_t}$. The reinforcing response happens when the return to investment in child's skill is too low because of the high pollution levels. In this case the household lowers the investment. Therefore, the negative net effect of pollution in the presence of reinforcing responses is larger than the direct impact alone. Theoretically, the reinforcing behavioral responses mean $\frac{\partial \theta_{t+1}}{\partial x_t} < \theta_{t+1} \frac{\delta_{6,t}}{x_t} < 0$. The range of the parameters that predicts the compensatory or reinforcing response can not be derived analytically.

Since the model does not predict a definite outcome, question about how pollution affects children's skill and how this effect is neutralized (or intensified) by households becomes an empirical question. The estimation results will determine which one of the roles households have: non-responsive to pollution levels, compensatory, or reinforcing role.

3.3 Data

Primary source of the data for this study is the Panel Study of Income Dynamics (PSID) and three waves of Child Development Supplement (CDS-I, CDS-II, and CDS-III). I also use pollution data from the US Environmental Protection Agency (EPA).

3.3.1 PSID and CDS

The PSID is a nationally representative longitudinal survey of the US individuals and families in which these individuals reside [9] . It provides a wide range of information on families and individuals. Since 1968 the PSID has collected data on family composition changes, housing and food expenditures, marriage and fertility histories, employment, income, health, consumption, wealth, and time spent on housework. The original PSID survey mainly focuses on households and particularly the head of the households, and then on spouses. From the main PSID survey, I use demographic information about the parents of children.

While PSID collects some information about children in the household, this information is quite limited. Starting in 1997, PSID added a new component, the CDS, that specifically focuses on children and collects detailed information on them. So far three waves has been administered: in 1997, 2002, and 2007. The data

[9] The PSID survey initially started in 1968 with a nationally representative sample of households. However, after the first wave of the survey in the following years, children from a household in the main sample who left the family and formed their own family were added as new households to the survey. Adding the new generation of the households into the survey and also dropouts from the main sample made the current sample unrepresentative of the national population. In order to fix this issue, PSID contains household's weights.

include but not limited to general school achievement information; the CDS also administers the subset of standard tests to assess academic skills of children. These tests include mathematics and language skills among other content areas. For this study I use the Letter-Word (LW) test scores as a measure of children's skill. The LW test is a subset of Woodcock-Johnson Revised (WJ-R) test of achievement that measures the symbolic learning and reading identification abilities of children at ages between 3 to 17. I use the raw scores for the LW test that is well-suited for examining changes in a child's performance on a WJ-R sub-test over time. [10]

The CDS also collects time use data on children for two days during a week: one weekday and one weekend day. Subjects fill out (with their caregiver if they are too young) a detailed time diary during these days. They provide information on what they have done (type of activities), where they have done (location of activities), starting and ending time of activities (duration of activities), and who was with them during the activities over 24 hours. I use the time diary information to extract the time that children spend on developing their skills, either alone (e.g. time at school or working alone on home works) or with their parents (e.g. studying with parents). Table 3.1 lists the variables that I use, years of data, and their sources.

I focus on the PSID and CDS surveys that are administered between the years of 1997 and 2007. For the sample I use for the estimation, I keep all the children who have at least one LW test score. I focus on children between the age of 3 and 15. Even though the administrated test is given to children between ages 3 and 18, I observe a dramatic unexplained drop in study time of children older than 16 and I drop those observations from the sample. I further restrict the sample based on the

[10]The PSID also .reports the standardized scores of the LW test that are normalized using a child's raw score, his age, and other children's scores in his age category. The standard scores are useful for cross-sectional comparison between different age groups. However, it is not useful to study changes in a child's performance over time. For further detail on the LW test see `https://psidonline.isr.umich.edu/Guide/default.aspx`

income data. I drop observations from families who had weekly income below \$100 as a minimum income required for a sustainable living in the absence of savings in the model. I also drop the high income observations with annual income above \$150000. These restrictions based on income remove 4% of the observations [11] .

Table 3.2 provides summary statistics of the data from the PSID and CDS. All the time variables are calculated in hours per week units and the family income is weekly income in 2000 dollars. Table 3.3 presents demographics variables for the CDS sample started at 1997.

Figure 3.1 shows the average LW test scores for every age group of children. The average test score increases at a declining rate by children's age. This observation is a crude measure of improvement in children's skill as they grow up. Figure 3.2 provides the average time that parents actively spend with their children and the time that children spend on their education alone for every age group. I define the parental time with their child "active" if at least one of the parents actively engaged in the activities that the child performs. Further, the education time alone is the time spent on school related activities such as time spent at school or time spent on homework at home that the child does on his own. As demonstrated in figure 3.2, as children grow up they spend less time with their parents and more on school related tasks.

3.3.2 Pollution

The ideal air pollution data for this study would be the exact measure of pollution that a child inhales. Unfortunately, such detailed and precise measure of exposure to pollution is not available unless it is recorded in a lab experiment. In case of the United States, researchers normally use the measure of pollution that is recorded by

[11] Some of the families have zero or negative income where negative income corresponds to business or farm losses. There are only 22 observations with zero or negative income which is less than 1% of all the observations with non-missing income.

the EPA via monitors throughout the country. Figure 3.3 shows the actual locations of monitors nationwide that are placed by the EPA to record variety of pollutants.

Among the multiple alternatives of pollution measures I use ground level ozone to control for environmental hazards because of its availability for the entire period of the CDS sample. Bad ozone is the one that is found in the Earth's lower atmosphere, near the ground level. The ground level ozone is not generated directly. It is formed from the chemical reactions between oxides of nitrogen (NOx) and volatile organic compounds (VOC) in the presence of sunlight and heat. NOx and VOC are emitted from different sources such as cars, power plants, industrial boilers, and refineries into the air. Previous studies have shown that ozone has negative effects on human health specially on children and elderly. [12]

Using the centroid of households' census block and latitude and longitude of the monitors I find a nearest monitor to every household's location. Assigning a monitor that is very far from a household and it may not be true representative of the pollution levels that the household is exposed to, can lead to measurement error. To avoid the measurement error I only keep those households that are within 20 miles of the nearest pollution monitor. Since the PSID survey after year 1997 is run biannually, I only observe households' location every other year. So, I do not observe what pollution levels a household faces between two surveys. In order to obtain more information from pollution data, if a household lives in the same census block in two consecutive surveys, I assume it was living in the same census block in-between.

Since the main data from the PSID is collected on yearly basis, the pollution data should match that annual pattern. Instead of reporting an annual measure collected from the monitors, the EPA generates so-called "Design Value". The design value is a statistic that the EPA generates to describe the air quality status at a particular

[12] For instance Currie and Neidell (2005) and Neidell (2004).

location relative to the National Ambient Air Quality Standards (NAAQS). Based on the design value, the EPA determines if a particular monitor or, at a more aggregate level, a county is in attainment status or not. The monitor or county with the design value above the predetermined threshold is considered to be in non-attainment and these with the design value below the threshold are considered to be in attainment status. If a county is in non-attainment status, it is required to lower the pollution levels below the designated threshold. Under the Clean Air Act Amendment (CAAA) every year the EPA assigns a county attainment/non-attainment status.

In order to facilitate the interpretation of the results, I normalize the design value by its standard deviation. Figure 3.5 shows the distribution of the pollution for the pooled data. Figure 3.4 gives an idea about the variation of the pollution levels across the country in 2000 where the darker colors represent higher level of pollution.

3.4 Estimation Method

In the following I present the assumption underlying the model specification and explain the procedure to identify the parameters of the model and at the end I explain the estimator that I use in this study.

3.4.1 Parameters

Each time period is equal to one year and child development process lasts until age 15, $M = 15$. Preferences after age 15 are captured in final utility function. In equation (3.3) I allow the TFP and the production parameters , $\delta_{k,t}$'s, to vary by a child's age. In order to reduce the number of parameters to estimate, $6 \times M$, I assume a linear form as following:

$$R_t = \exp(\lambda_{0,1} + \lambda_{0,2} t) \qquad t = 1, \ldots, M,$$

$$\delta_{k,t} = \lambda_{k,1} + \lambda_{k,2} t \qquad k = 1, \ldots, 5; \quad t = 1, \ldots, M, \tag{3.11}$$

Using this linear form, the total number of technology parameters to be estimated reduces to only 12. [13]

Because of the normalization of the preference parameters, I only need to estimate three of α_k's using the following mapping:

$$\begin{aligned} \alpha_1 &= D^{-1} exp(\upsilon_1), \\ \alpha_2 &= D^{-1} exp(\upsilon_2) \\ \alpha_3 &= D^{-1} exp(\upsilon_3), \\ \alpha_4 &= D^{-1}, \end{aligned} \tag{3.12}$$

where $D^{-1} = 1 + \sum_{l=1}^{3} exp(\upsilon_l)$. Instead of estimating α_j's directly, I estimate three υ_l's. All the three υ_l's can take any real numbers, however, the mapping guarantees that $0 \leq \alpha_j \leq 1$ and α_j's sum to one. Another preference parameter to be estimated is the child's skill weighting parameter in the final period, φ.

Initial values for child's skill level, household's income, parents' labor hours, and pollution levels are drawn jointly from a log-normal distribution with average and covariance matrix of $(\mathbf{M}, \Sigma)$. After the initial conditions are realized, the child's skill evolves over time according to the technology function. During the development periods household's income, parents' labor hours, and pollution levels are drawn from stochastic and exogenous processes described in the model section. Six parameters - $\omega_0, \omega_1, \rho_0, \rho_1, \gamma_1$, and γ_2 - and the variances of the error terms of these processes are unknown and need to be estimated.

[13] Two λ's for each of the five δ's and two for R_t.

3.4.2 *Identification*

This section lays out the identification strategy of the model parameters. I jointly estimate the preference and technology parameters within the model. I estimate the parameters of the initial condition and the exogenous processes of household income, parents' labor hours, and pollution levels outside the model using directly the actual data.

If a child starts a period with a skill level θ_t, then the skill at the end of the period t is:

$$\begin{aligned}\ln\theta_{t+1} = \;& (\lambda_{0,1}+\lambda_{0,2}t)+(\lambda_{1,1}+\lambda_{1,2}t)\ln\theta_t \\ & +(\lambda_{2,1}+\lambda_{2,2}t)\ln\tilde{e}_t+(\lambda_{3,1}+\lambda_{3,2}t)\ln\tilde{\tau}_{pt} \\ & +(\lambda_{4,1}+\lambda_{4,2}t)\ln\tilde{\tau}_{ct}+(\lambda_{5,1}+\lambda_{5,2}t)\ln x_t+\ln u_t \\ & = f(\theta_t, e_t, \tau_{pt}, \tau_{ct}, x_t; \Lambda)+\ln u_t,\end{aligned} \tag{3.13}$$

Λ is the vector of all the parameters in the technology function that includes λ's and complementary parameters, p's, q's, and r's. Under the assumption that error term is independently distributed across children and over time the production parameters λ's can be recovered using the non-linear least squares (NLS) approach under the standard full rank condition where the objective function is

$$\hat{\Lambda}_{NLS} = \underset{\Lambda}{\operatorname{argmin}} \sum_{n=1}^{N} (\ln\theta_{t+1} - f(\theta_t, e_t, \tau_{pt}, \tau_{ct}, x_t; \Lambda))^2, \tag{3.14}$$

In order to the full rank condition hold, households should have different levels of investments and values for the exogenous variables. This condition is routinely satisfied in the actual data. Additionally, given that the $\delta_{i,t}$'s are a linear function of children's age, a necessary and sufficient condition for the sample is to contain children of at least two different ages. This condition trivially is met in the actual data. Hence, for a given set of preference parameters the technology parameters can be identified. One remaining concern is identifying the parameters of unobserved monetary investment. However, in the model monetary investment is the only variable that is directly related to observable household income. [14] I substitute optimal solution of monetary investment which is a function of income into the model and back out the technology parameters for a given set of preference parameters.

Next, I identify preference parameters. I take the value for the discount factor, β, as given and set it to conventional value 0.96. I need to identify υ's and the final period's preference parameter, φ. The time investments, τ^*_{pt} and τ^*_{ct}, are functions of φ and υ's. [15] For a given set of technology parameters the time investments can be inverted to recover the preference parameters. Empirically, the sufficient condition to identify φ and υ's is that I observe at least two households with different levels of time investments and their children are in different ages. [16] Hence, converting actual measure of time investments yields values of preference parameters that are homogeneous across the households. Basically, if there are enough variation in time investments and children's ages, I can recover preference parameters.

I estimate the average and covariance matrices of the log-normal distribution of the initial condition outside the model directly from the actual data. Lastly, I need to identify the parameters of the law of motions for income, parents' labor hours, and pollution. Since the law of motions for income, parents' labor hours, and pollution levels are homogeneous across the households, I can use household longitudinal data to identify the parameters of these process outside of the model. If I have income, parents' labor hours, and pollution data for at least three periods, I can identify the parameters of these three law of motions. Since this condition holds in the actual data I recover those parameters.

[14] $e^*_t = \frac{\beta\delta_{3,t}(1+p_1x_t)A_tI_t-\alpha_1p_2}{(1+p_1x_t)(\alpha_1+\beta\delta_{3,t}A_t)}$

[15] See 3.11 for more detail.

[16] This condition also holds if I observe at least one household in the data at two different points in time.

3.4.3 Estimator

I use the Method of Simulated Moments (MSM) to estimate the parameters of the model. Denote M_N to be the vector of sample moments that summarizes the relationships between variables where N indicates the number of children in the data sample. The goal of the estimation is to identify a set of parameters of the model that generates a vector of moments, $\tilde{M}_S$, as close to true moments, M_N, as possible.

The empirical process starts at year 1997. In the first period, I draw a vector of household's income, parents' labor hours, pollution levels and initial child's skill at the age of 3 for S number of households. For a given set of model's parameters I solve the household's decision problem for 1997 that yields the household's optimal value of consumption, monetary and time investments on the child, parents' and child's leisure at 1997 and the child's skill in 1998. In year 1998 I draw from the distribution of shocks to household's income, parents' labor hours, and pollution levels. Using the exogenous processes for these variables, I calculate their values for 1998. Simulating the child's skill for 1998 from the household's decision at 1997 and knowing the household's income, parents' labor hours, and pollution levels I solve the household's decision problem as before. Repeating this process yields the path of optimal decision variables of the household's and child's skill from 1997 to 2007. I repeat the same process for all the S households. While I observe children's skills in 1997, 2002, and 2007, I can simulate these values for all years from 1997 to 2007 using the data generating process (DGP).

For a given set of parameters and using the DGP, I simulate a sample of S households and calculate the moments of the sample, $\tilde{M}_S$, that is analogous to M_N in the actual data sample. The problem is to find a set of parameters, Ψ, that minimizes the following function:

$$\hat{\Psi}_{N,S,W} = \underset{\Psi}{\operatorname{argmin}}(\tilde{M}_S - M_N)W(\tilde{M}_S - M_N)$$

where W is the inverse of the covariance matrix of data moments, M_N, that I estimate it by re-sampling the actual data. Moments that I use in the estimation are average, standard deviation, and the correlation between test score, parents' active time with children, and child's alone time on education at each child age group: 3-6, 7-10, and 11-15 years old. I also use the correlation between test score, parents' active time with children, child's alone time on education, pollution instrumental variable, households' income, and parents' labor hours at each child age group.

In the set of moments the orthogonality condition between the instrumental variable and the error term is $E(z_t u_t) = 0$. If the pollution levels was not endogenous, then it would be uncorrelated with the error term, i.e. $E(x_t u_t) = 0$. However, because of the omitted variables this condition does not hold anymore and instead the error term is orthogonal to the instrumental variable. Hence, I use the moment condition of $E(z_t u_t) = 0$ instead of $E(x_t u_t) = 0$.

3.5 Estimation Results

In this section, I present estimated parameters of the structural model and within the sample fit.

3.5.1 Preference Parameters

All the households have the same time invariant preferences; thus, I only need to estimate one set of preference parameters. The estimated preference parameters are presented in table 3.4 panel A. The transformed preference parameters, υ's, are difficult to interpret. I instead present the original preference parameters, α's.

Since the preference parameters in the flow utility are normalized, α's are the weights that households place on each input in their utility function. The weight that an average household attach to parental leisure and child leisure is almost the same, 0.19. These weights for consumption good and child's skill are 0.26 and 0.35, respectively. This means that in terms of flow utility households value one unit of child's skill more than one unit of any other input in its preference. The value of child's skill is actually greater because of the weighted child skill in the last period of the child development in the utility function of the household. Another measure to compare the importance of different "good" in the household's utility is to compare the household's willingness to pay for one additional unit of the "goods". Units of parents' and child's leisure time is hours per week and the units of child's skill is questions correct on the LW test. Households' average willingness to pay for parental leisure, child leisure, and child's skill are \$10, \$9, and \$57, respectively. Household's willingness to pay for different "goods" also reveals the high value of child's skill for the household.

The estimated scaling factor for children's skill in the last period, φ, is 9.06. In order to provide some insight on the value of the scaling factor, if the model's assumption was that the household live infinitely and it values its child's final skill in all the future periods, the implied value of φ would be $\frac{1}{1-\beta} = 25$. So, this implies the household has a high valuation child's skill.

3.5.2 Skill Formation Technology Parameters

There are two groups of technology parameters: time invariant, and time variant parameters. Time invariant technology parameters, p's, q's, and r's , are the complementary parameters between investment variables and pollution levels. Table 3.4 panel B provides the estimated values of these parameters. Standard errors are calculated using the bootstrap method. The time variant technology parameters are the TFP, R_t, and the elasticity of child's skill with respect to inputs, δ's, that vary by child's age. These parameters are presented in figures 3.6-3.8.

Recall that $\frac{\partial e_t^*}{\partial x_t} = \frac{\alpha_1 p_1 p_2}{(1+p_1 x_t)^2(\alpha_1+\beta\delta_{3,t}A_t)}$; since p_1 and p_2 are both positive, the monetary investment is increasing in pollution levels. If pollution goes up, household will spend more on child's skill. Hence, the household has a compensatory role regarding monetary investment on child. It means that the household who lives in a polluted area invests more money on its child to compensate for the negative effect of pollution.

Based on the signs of these parameters it is not very straightforward how parental time and pollution levels are related. However, for the given point estimate results and average values for the exogenous variables of the model I can calculate the change in parental time in response to change in pollution levels. Parental time is increasing in pollution levels, so the household has compensatory role regarding parental investment on child. Thus, the child who lives in a polluted neighborhood, receives more parental time to compensate for the negative effect of pollution. For the given average pollution levels, 10% higher pollution on average leads to 0.005% more parental time. The opposite is true for the child's educational time alone. Meaning that a child who lives in a polluted area spend less time on education alone as compare to a child in a better neighborhood. I must point out that p_1, q_2, and r_1 are not statistically significant.

Figures 3.6-3.8 show the value of δ_t's by child age over the child development horizon. In figure 3.6 parameters of $\delta_{1,t}$, and $\delta_{2,t}$ represent the elasticity of child's skill with respect to stock of child's skill and monetary investment, respectively. The self-productivity parameter, $\delta_{1,t}$, is downward sloping. This pattern represents the diminishing marginal return of stock of child's skill. In other words, having a higher skill to improve child's skill is more important in early childhood than later.

Because of the interaction terms of investment variables with pollution levels, $\delta_{2,t}$-$\delta_{4,t}$ represent the elasticities of the effective decision variables ($\tilde{e}_t$, $\tilde{\tau}_{pt}$, and $\tilde{\tau}_{ct}$), i.e. the original decision variables interacted with pollution levels. As it was expected figures 3.6 and 3.7 show that $\delta_{2,t}$, $\delta_{3,t}$, and $\delta_{4,t}$ are positive. In addition, the elasticity of parental time interacted with pollution is decreasing as child ages, and the elasticity of educational time alone and monetary investment interacted with pollution become more productive as child ages. Decreasing pattern of $\delta_{1,t}$ and $\delta_{3,t}$ is in line with the previous findings of child development literature that emphasize the importance of early childhood intervention. [17] Stock of child's skill and parental time are more productive early on in child's development process. On the other hand, increasing pattern of $\delta_{2,t}$ means that monetary investment becomes more productive as the child grow older. Further, getting more productive at studying alone that includes school time for older children may lead to the increasing pattern of $\delta_{4,t}$.

Figure 3.8 shows that the coefficient of pollution levels is negative and is increasing by child's age. Without any interaction in the model, δ_5 represents the elasticity of child's skill with respect to pollution. Increasing pattern means that younger children are more vulnerable to environmental hazards than their older peers. However, more intuitive way to explain the effect of pollution is the direct effect of pollution that is given in equation (3.8). The direct effect is the measure of interest that shows how child's skill will be affected by pollution in the absence of compensatory investment by the households. The direct effect varies by age and it depends on income level, parents' free time, pollution levels, and value of child's skill. The computed direct impact for 3 years old child is about -1.44. This means that for an average child

[17]For example Heckman (2008) and Cunha *et al.* (2006)

of age 3 year, increasing pollution levels by one standard deviation reduces child's test score by 0.10 standard deviation. This value for 14 years old children is about 0.04 of standard deviation. On average the direct effect across all the ages is about 0.07 standard deviation for one standard deviation increase in pollution levels. The decreasing pattern, absolute value, of the negative effect of pollution supports the fact that younger children are more vulnerable to environmental hazards than their older peers.

3.5.3 Within Sample Fit

This section presents the within sample fit results for three main variables of the LW test score, parental time with child, and child's educational alone time in tables 3.5-3.6 and figures 3.10-3.12. Tables 3.5 and 3.6 show the results for three age categories and the numbers are the average values of the variables for every age group. Figures 3.10-3.12 present similar results as the tables with finer categories of child age. Figures demonstrate a good fit of the model to the data.

The estimated model is able to predict the increasing and concave pattern of the LW test score data. Parental time average in the data has a sharp decline from age 3 to 8, and after age 8 it declines with a slower rate. The model is fitting the data well in terms of these slopes before and after the age of 8. Children's alone time on education is increasing up to age 7, and after they start the school the time growth slows down and almost stays constant. The model is able to track this pattern for all the age groups before and after the age of 7.

3.6 Counterfactual Analysis

In this section I consider two counterfactual policy experiments using the results of the point estimates. The first policy is an environmental policy that exogenously reduces the pollution levels. The second policy is the income transfer to households. The counterfactual exercise will provide the magnitude and direction of the household response to these policy changes and the effect on children's skill.

3.6.1 Experiment 1: Pollution Decline

As with any policy intervention, it is critical to have a credible estimate of the costs and benefits of environmental policy for its successful implementation. In this section I examine the impact of policy that reduces pollution level using the point estimates results.

Figure 3.13 shows the result of the experiment. The baseline is the simulated test scores from the model for the estimated parameters. In the experiment, I reduce the stream of the pollution that children are exposed to by one standard deviation at every age of the development process. In other words, children now live in a neighborhood with better air quality as opposed to what they used to live. Reduction in pollution levels leads to increase in children's test score for all the ages. This improvement comes from multiple channels. First channel is the direct effect of pollution that has a negative effect on children's skill. Among the decision variables children's alone time on education, τ_{ct}, increases in response to pollution decline and positively affects the child's skill. These factors, reinforced by self-productivity, $\frac{\partial\theta_{t+1}}{\partial\theta_t}$, magnify the improvement in child's skills brought by the reduction in pollution level. However, both monetary investment and parental time with the child are positively correlated with pollution levels. This means that in response to pollution reduction the household will enjoy from more consumption good and parents' leisure than investing on the child. This response will dampen the positive effect of pollution reduction on the child's skill.

It is also worth examining how the experiment result would change if I were to ignore the household's investment responses to pollution decline. To do so, I simply calculate the children's skill using the skill formation technology with the new pollution levels and using investment values from the baseline model. This method, therefore, shuts down the investment response channel because it assumes that investment variables stay unchanged after the pollution decline. The result of this exercise together with the previous ones is presented in figure 3.13. In the full model average test score among all the age groups increase by 0.09 standard deviation in response to one standard deviation decline in pollution levels, as compared to a 0.13 standard deviation increase in the average test score when the investment response is ignored. Smaller effect in the full model as compare to the case of shutting down the response channel means that compensatory effect of monetary investment and parents' time with the child is stronger than the reinforcing effect of the child's alone time on education.

Since the productivity of inputs in the skill formation technology is age dependent and the effects accumulate over time through the self-productivity, the effect of pollution on children's skill varies by age. Figures 3.14-3.16 show the distribution of the changes in test score due to pollution reduction for three age categories in the full model with investment responses. The pattern for the case without investment responses are qualitatively similar.

3.6.2 Experiment 2: Income Transfer

Income transfer is a conventional policy solution to improve children's skills especially among the disadvantaged families. Additional income transfer may influence the investment decisions of the households and, thus, child's skill. In income transfer counterfactual experiment households receive transfer of one standard deviation of income that is roughly \$600 per week at every age of children. Unlike the pollution reduction counterfactual, household's income affects child's skill only through the monetary investment that is positively correlated with income levels.

Figure 3.17 demonstrates the result of income transfer on children's skill. On average, one standard deviation income transfer increases the test score by 0.12 standard deviation. Because of cumulative effect through stock of child's skill, the effect of income transfer is magnified for older children and it leads to larger gap for 15 years old children.

To compare the results of two experiments figure 3.18 presents the effect of two experiments in one frame. The horizontal axis is the amount of weekly transfer to the households and the vertical axis is the LW test score. The flat curve presents the average test score from one standard deviation reduction in pollution levels without income transfer. The upward slopping curve represents the average test score for different amount of income transfers. The figure shows that reducing pollution levels by one standard deviation leads to the same average test score as a weekly income transfer of \$120, i.e. \$6240 per year. Figure 3.19 presents similar result but for the average utility level instead of average test score. The figure shows that reducing pollution levels by one standard deviation leads to the same average household utility as a weekly income transfer of around \$30, i.e. \$1560 per year. The fact that smaller income transfer equates utility levels as compare to \$120 that equates the LW test scores under two experiment comes from the preference and the technology function. Pollution does not directly affect the household's utility but through the child's skill. Hence, pollution reduction increases the household's utility through improving the child's skill. However, since the income directly enters into the household's utility through the consumption good the small transfer of income can lead to the same level of utility as pollution reduction. The income transfer of \$30 will not lead to a large improvement in the child's test score because it will be split up between consumption and monetary investment on the child. If the goal is to improve the child's test score equal to pollution reduction, the amount of transfer has to be large enough such that overflow of transfer into monetary investment is enough to improve the test score equal to pollution reduction. Related, Chay and Greenstone (2005) find that the elasticity of housing values with respect to particulates concentrations ranges from -0.20 to -0.35 at an average housing price of \$40290.

3.7 Discussion and Conclusion

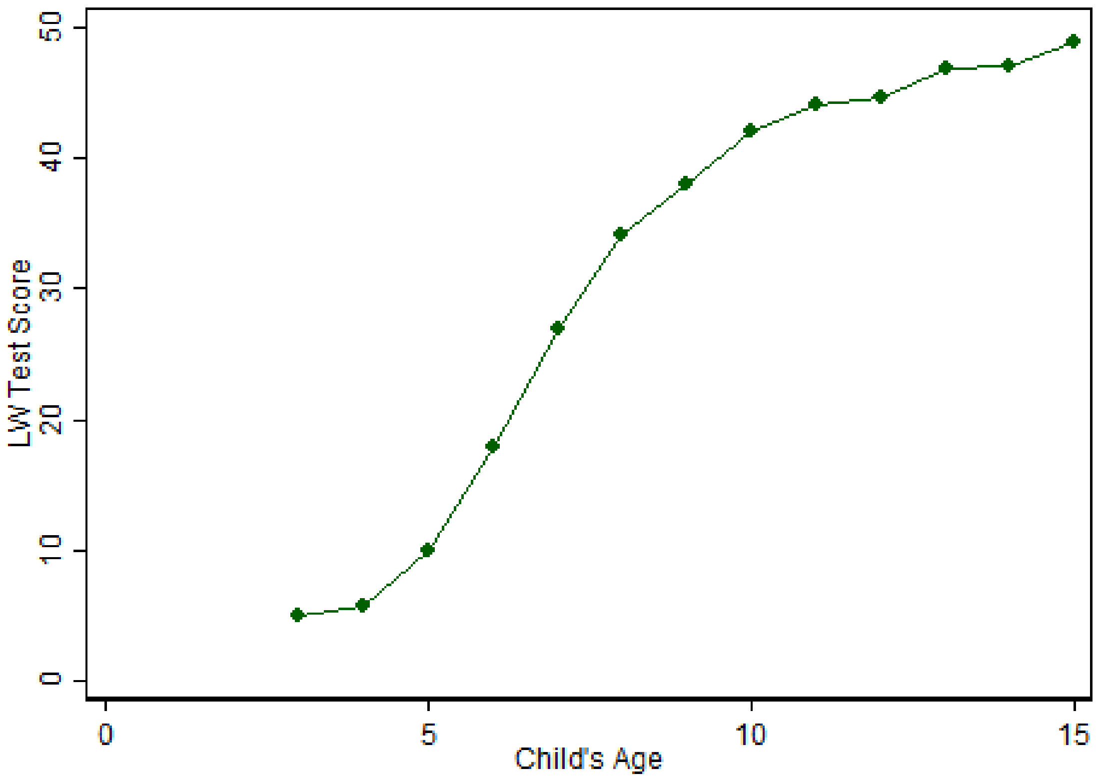
In this chapter I estimated the skill formation technology of children while including pollution levels as one of the determinants of children's skill. I used panel data from the Panel Study of Income Dynamics and Child Development Supplement as a source of data on test scores, time use, and demographics of children and their family. I merged these data sets with the measures of pollution from the Environmental Protection Agency: I used ozone as a control for pollution exposure of children because it is one of the most hazardous pollutants to human health.

My estimation results show that pollution has negative effect on the LW test scores. I find that increasing the pollution levels by one standard deviation decreases the test scores by 0.07 standard deviation on average. This average effect, however, masks significant heterogeneity among different age groups. Thus, test score of 3 years old would go down by 0.10 standard deviation in response to one standard deviation increase in pollution levels, while for the 14 years old children this effect is nearly twice as small - 0.04 standard deviation. This finding implies that children are more vulnerable to pollution in early childhood. The magnitude of the effect I find is significant compared to other policies of improving children's test scores estimated in the literature. For example, Krueger (1999) shows that reducing class size by 7-8 students in a class with 20 students, improves the students standardized test scores by 0.22 standard deviation.

The estimated effects above only represent the direct impacts of pollution without accounting for its indirect impact and the dynamic nature of the model. In the presence of behavioral response the net effect of pollution can be different than the direct impact alone; moreover, the net effect can be accumulated over time and create multiplier effect. Thus, the benefit of continuously living in a neighborhood with one standard deviation lower pollution levels from the age of 3 to 15, translates into 0.19 standard deviation higher test score at age 15.

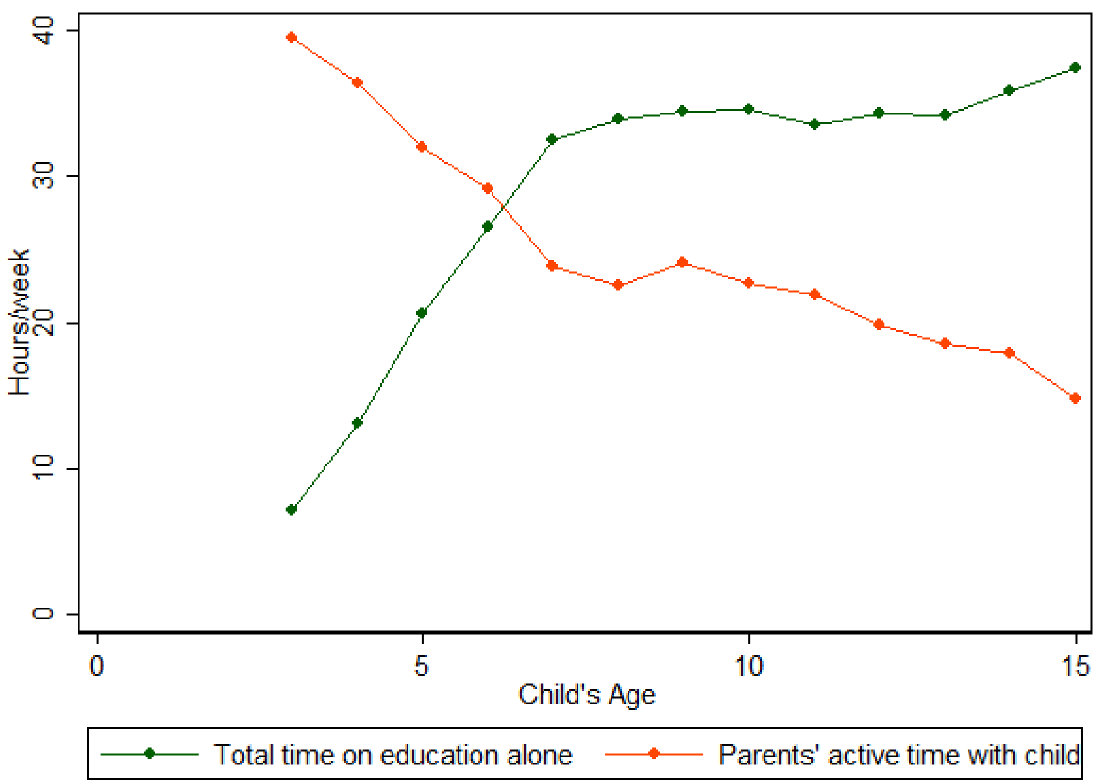
I consider income transfer as a policy that could potentially mitigate the negative effect of pollution on child's skill. I find that an annual income transfer of $31,000 (an equivalent of one standard deviation household income) on average increases the test scores by 0.12 standard deviation. To compare the magnitude of this effect, Dahl and Lochner (2012) find that $1000 increase in family income leads to 0.06 standard deviation increase in combined math and reading scores. In a relate study, Neilson and Zimmerman (2014) estimated that $77,000 per student in school construction expenditure results in 0.21 standard deviation gain in reading score.

Figure 3.1: Average LW Test Scores by Age



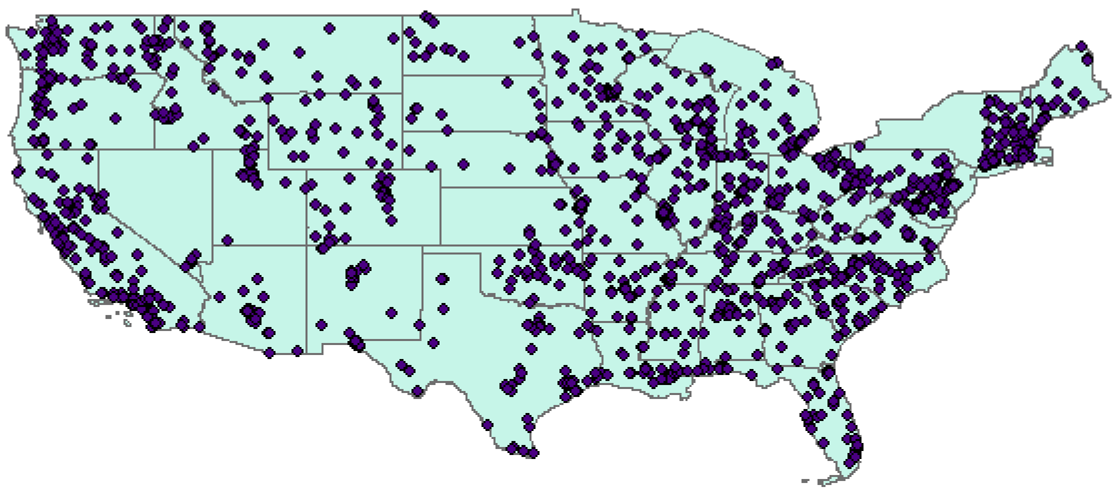
Source: The LW test scores comes from the PSID-CDS.

Figure 3.2: Average Child's Time on Education Alone



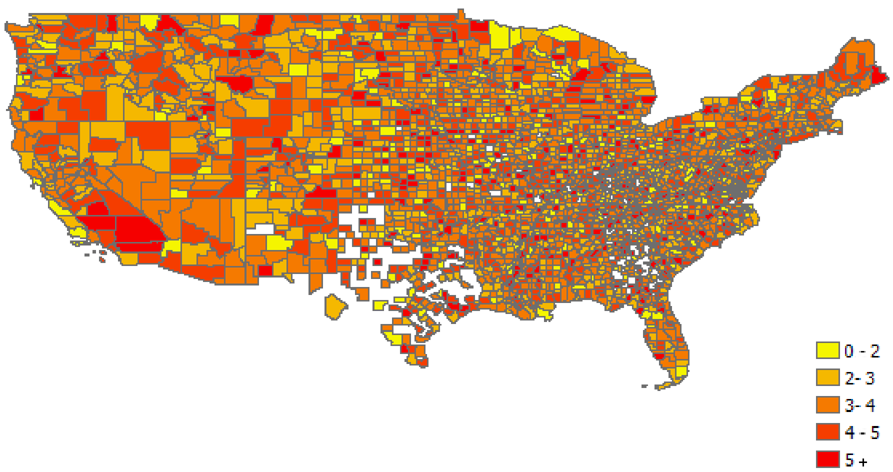
Source: The time diary information comes from the CDS.

Figure 3.3: Pollution Monitors' Location



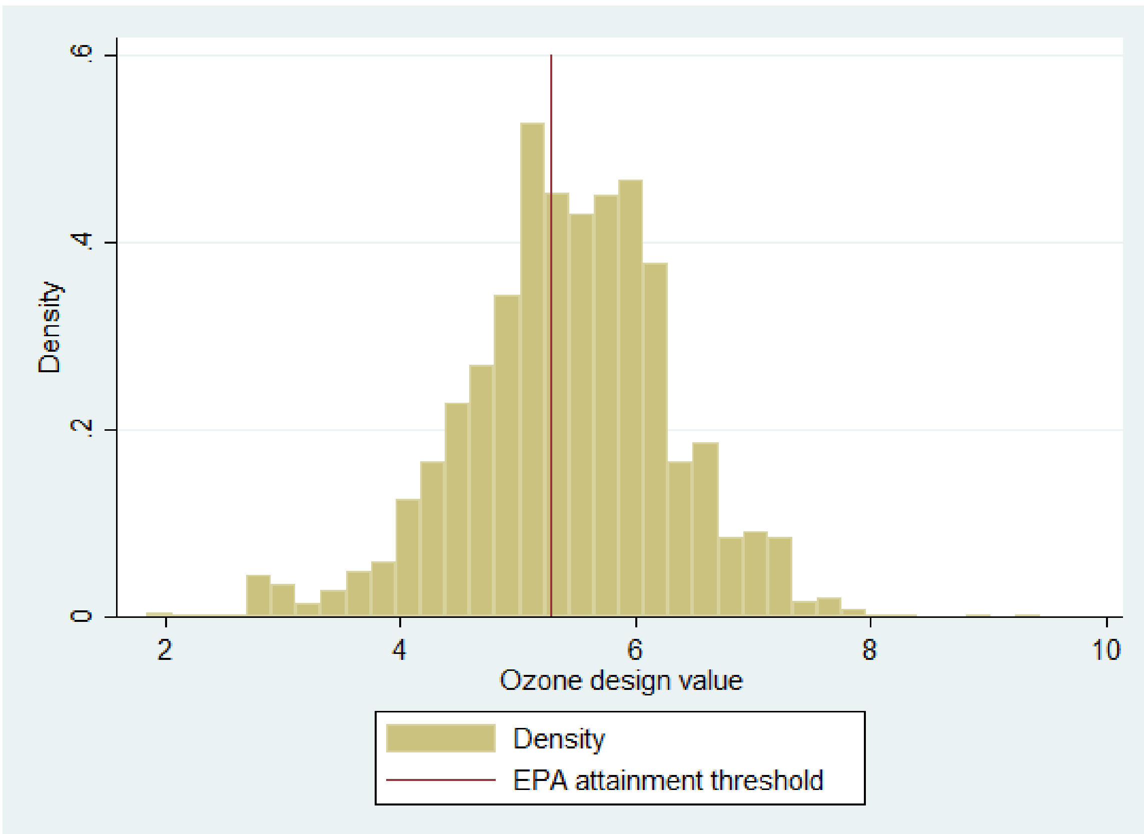
Source: The exact geographical location of the monitors is from the EPA.

Figure 3.4: Ozone Average by County for Year 2000



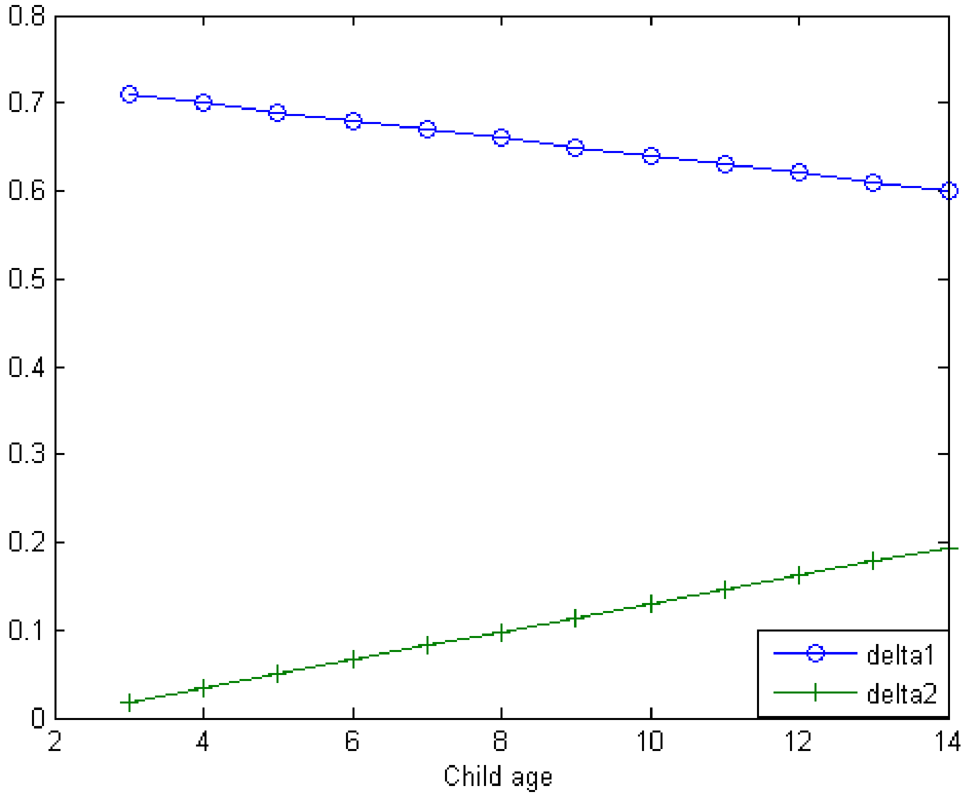
Source: Ozone data comes from the EPA.

Figure 3.5: Pollution Distribution



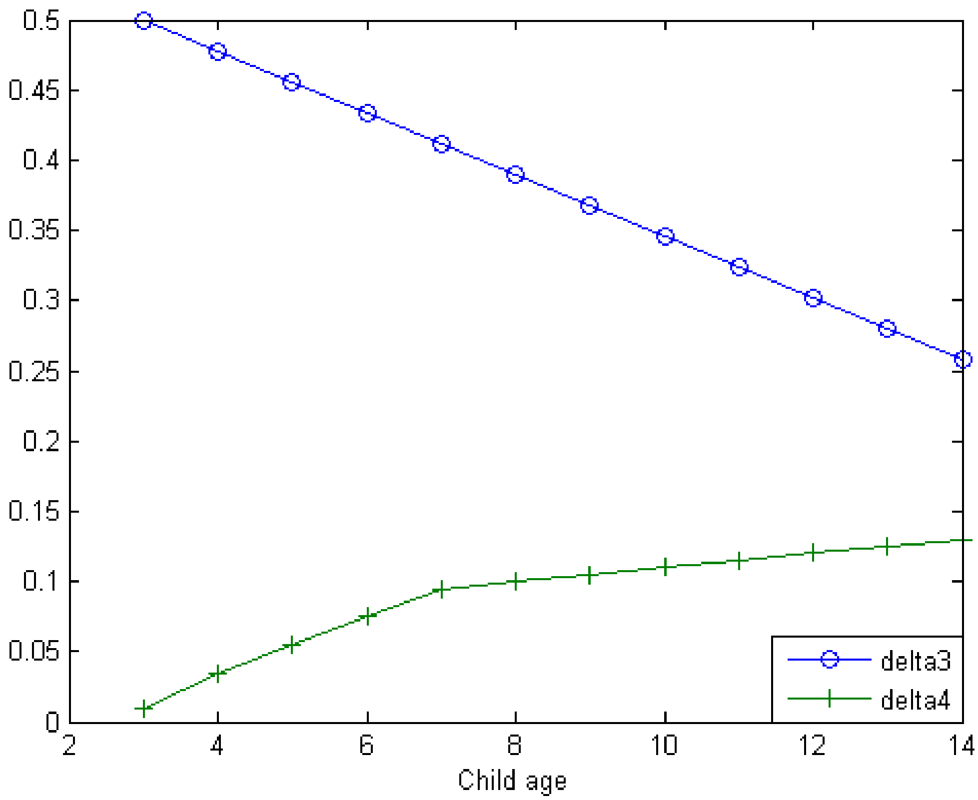
Source: Ozone data comes from the EPA.

Figure 3.6: Elasticities by Child's Age



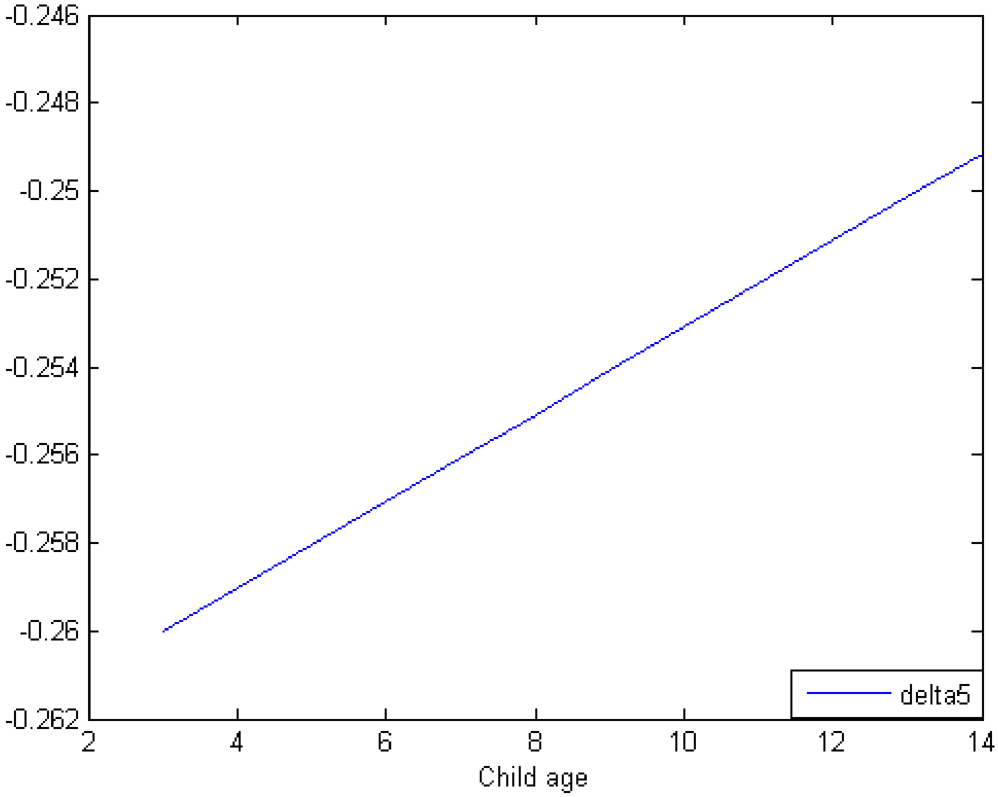
Note: δ_1 and δ_2 are the elasticity of θ_{t+1} with respect to θ_t and $\tilde{e}_t$, respectively.

Figure 3.7: Elasticities by Child's Age



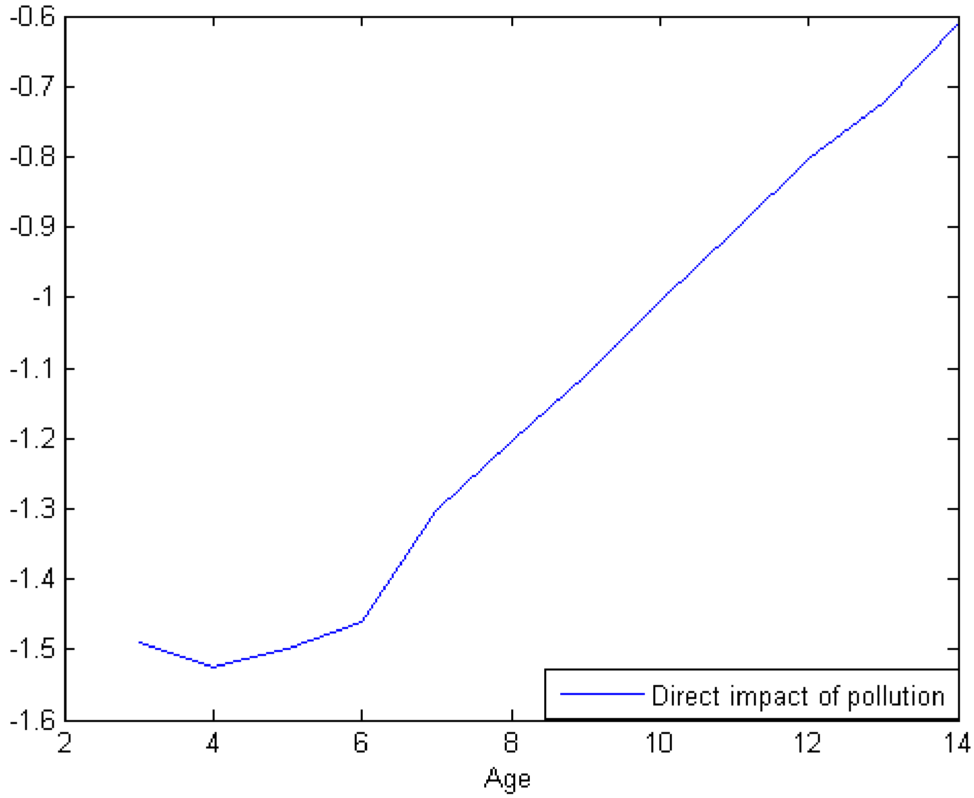
Note: δ_3, and δ_4 are the elasticity of θ_{t+1} with respect to $\tilde{\tau}_{p,t}$, and $\tilde{\tau}_{c,t}$, respectively.

Figure 3.8: $\delta_{5,t}$ by Child's Age



Note: δ_5 is the power of x_t in the skill technology function.

Figure 3.9: Direct Impact of Pollution on Test Scores



Note: This graphs shows the direct effect of pollution on the LW test score for given values of exogenous variables in their mean level. To fix the idea, one standard deviation increase in pollution levels roughly reduces the LW test score by 1 score for a child at age 14. This number is around 3.5 for an 8 years old child.

Figure 3.10: LW Test Scores by Child's Age in the True and Simulated Data

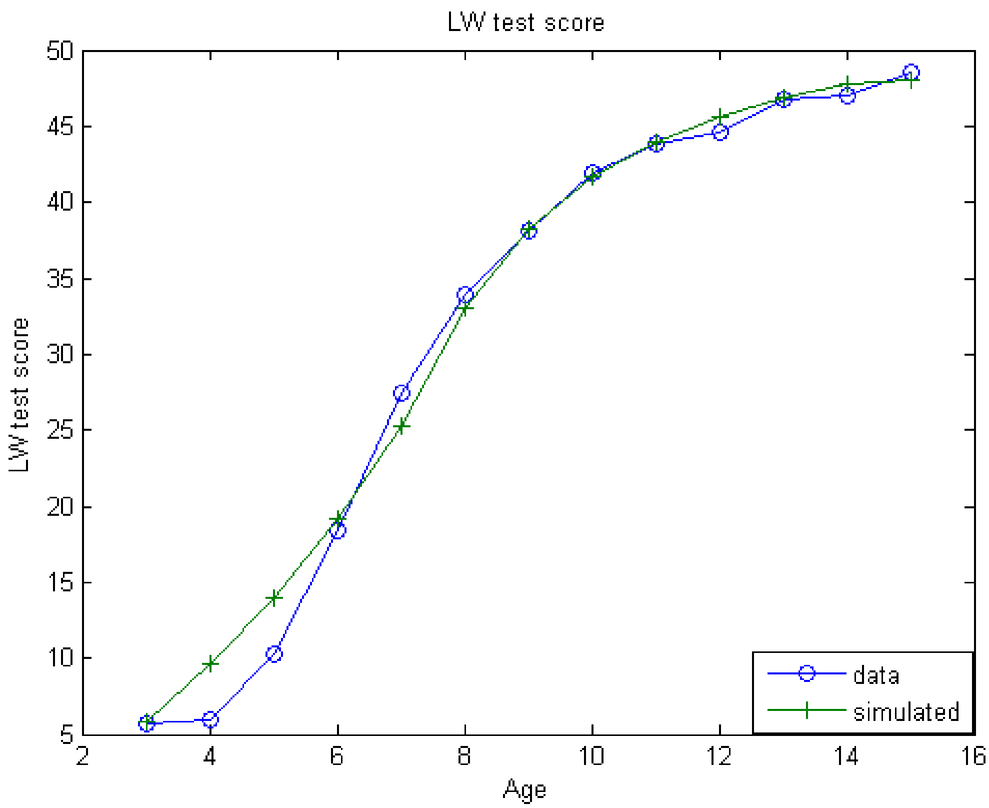


Figure 3.11: Parental Time by Child's Age in the True and Simulated Data

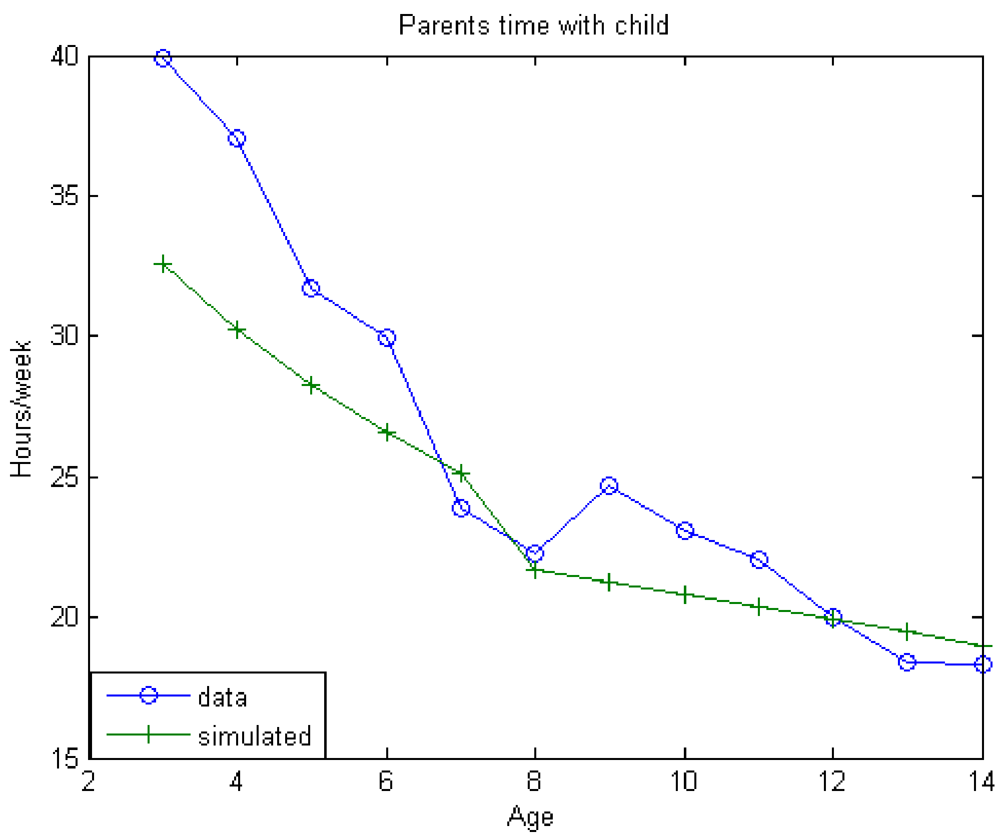


Figure 3.12: Education Time Alone by Child's Age in the True and Simulated Data

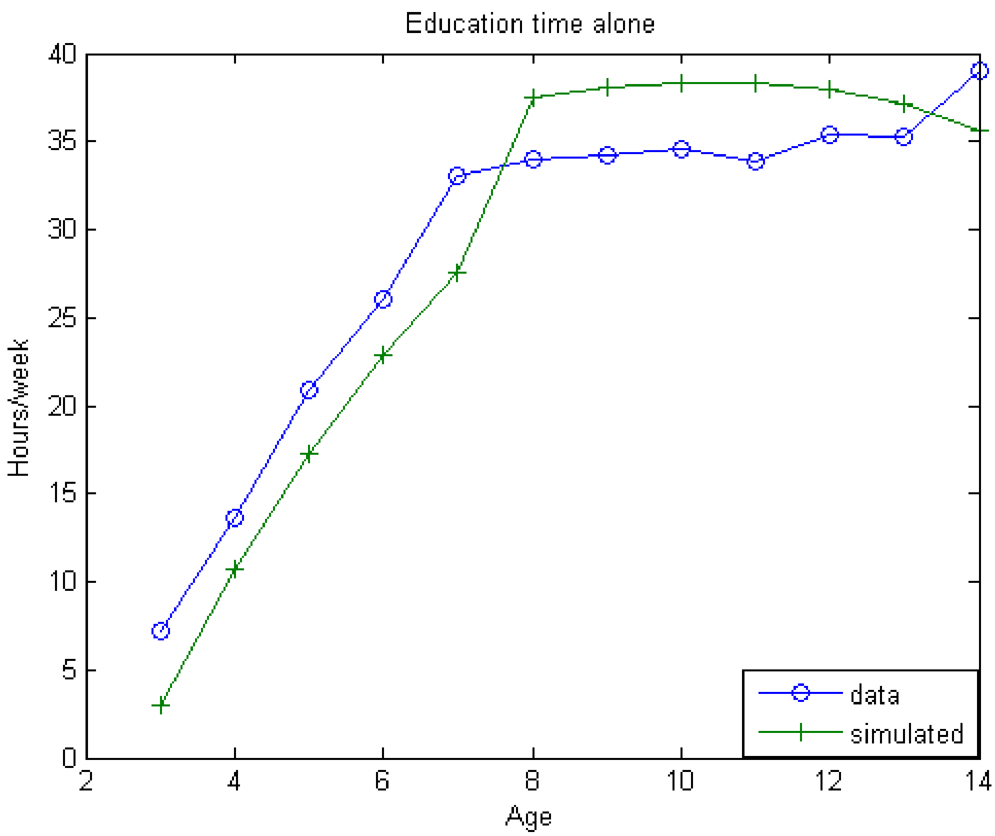
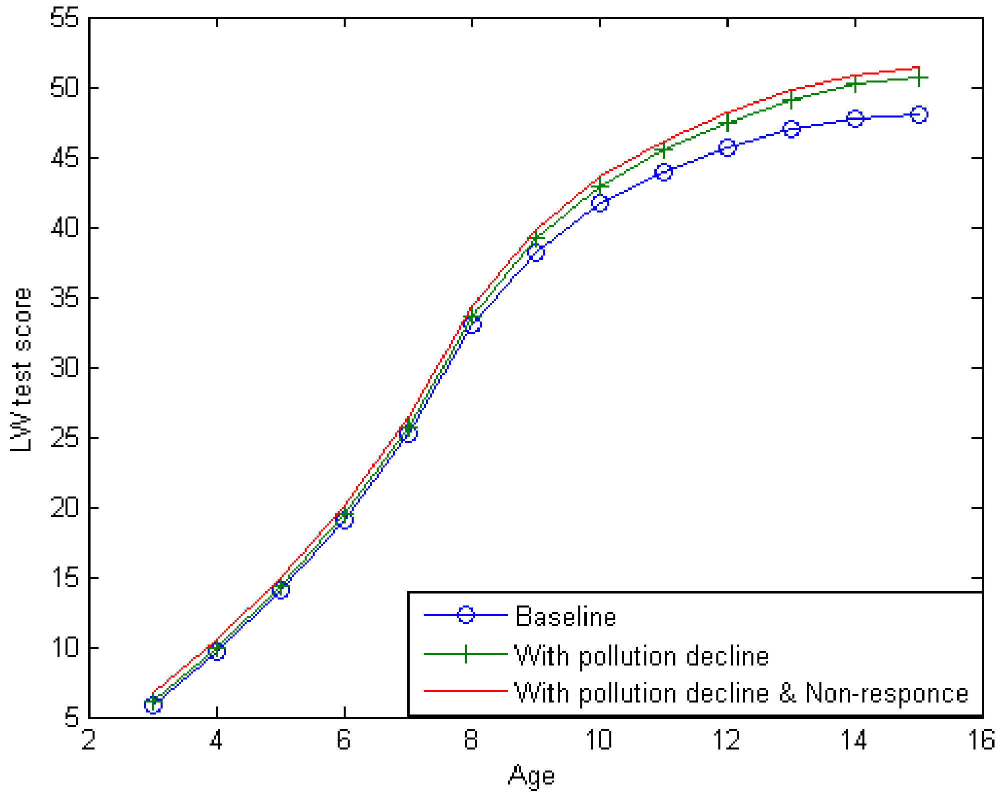


Figure 3.13: Average Child's Test Scores



Note: The baseline is the simulated data using the point estimate results. The second curve is the simulation results from one standard deviation reduction of pollution level for all the ages. The third curve is the simulation results from one standard deviation reduction of pollution level for all the ages without investment responses.

Figure 3.14: LW Test Score Change Due to Pollution Decline. Age 3-6

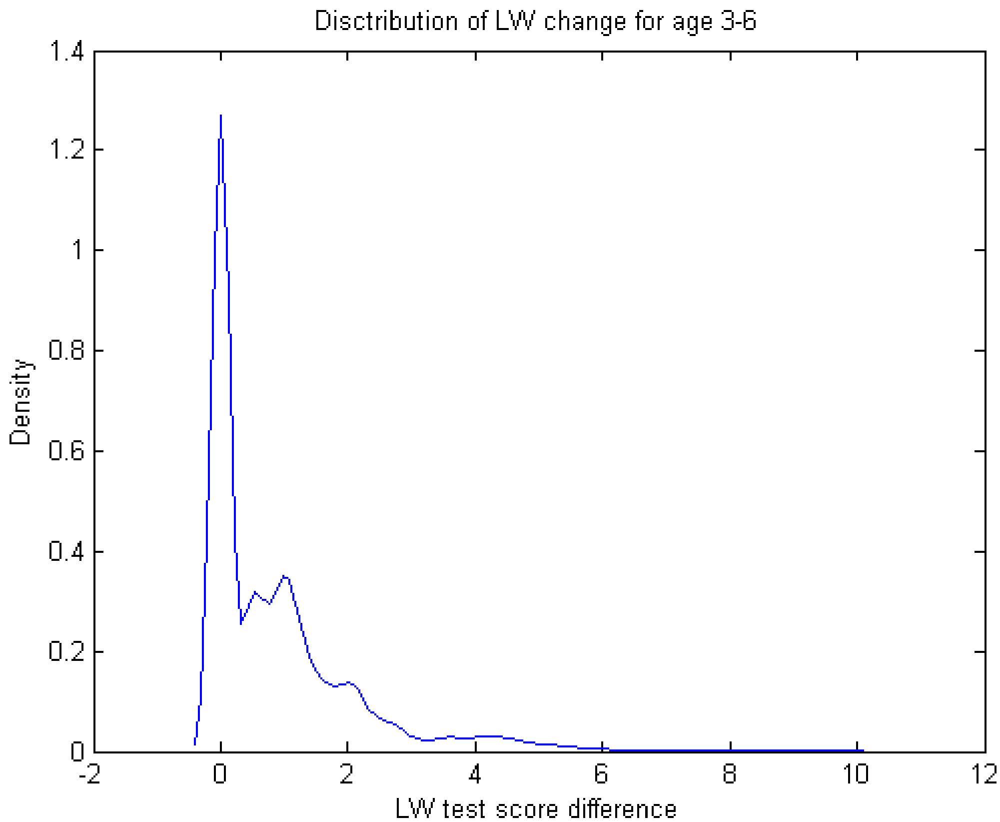


Figure 3.15: LW Test Score Change Due to Pollution Decline. Age 7-10

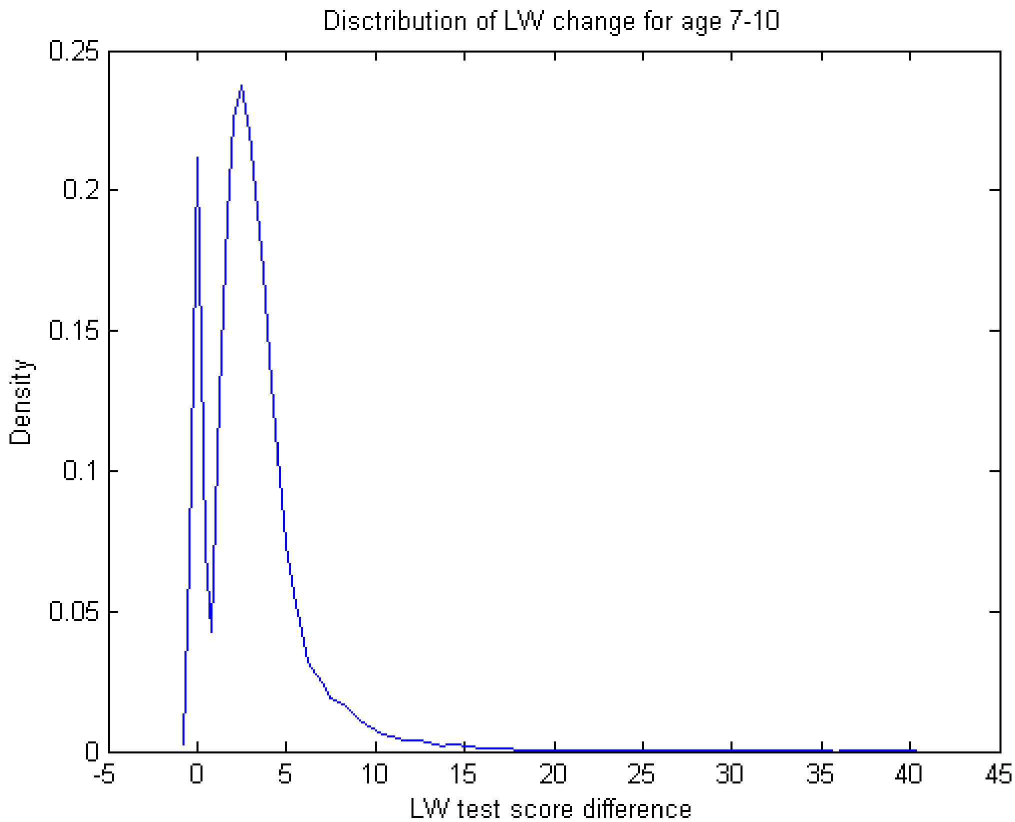


Figure 3.16: LW Test Score Change Due to Pollution Decline. Age 11-15

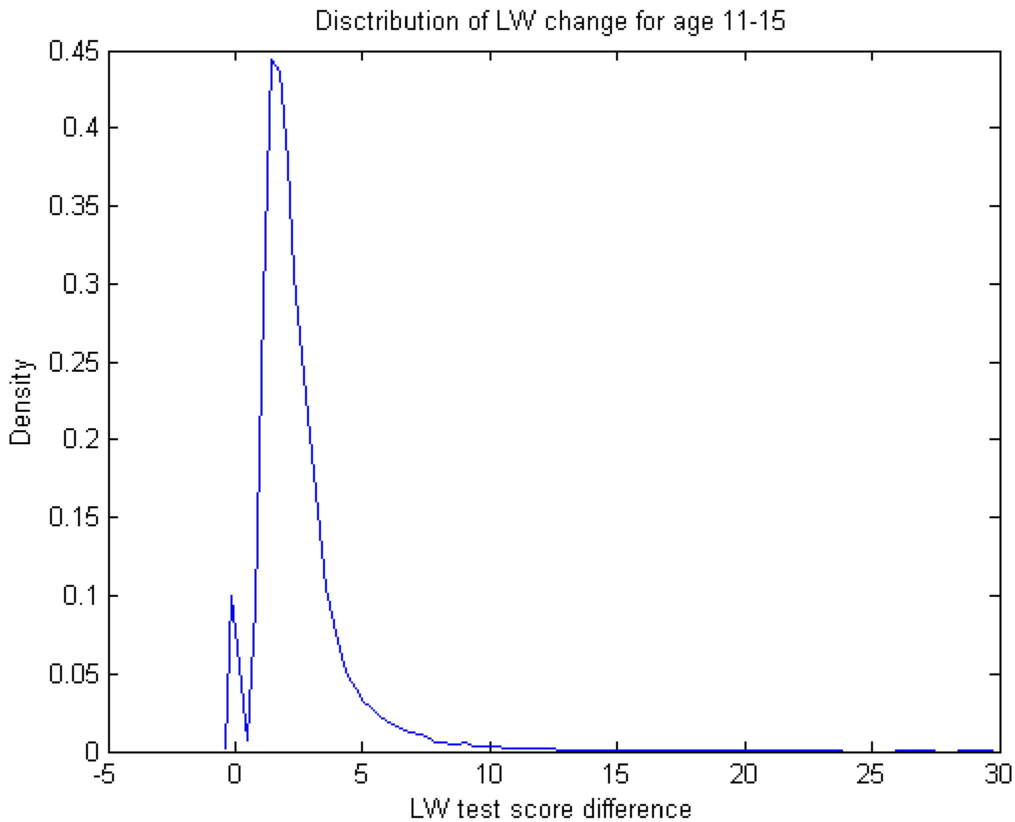
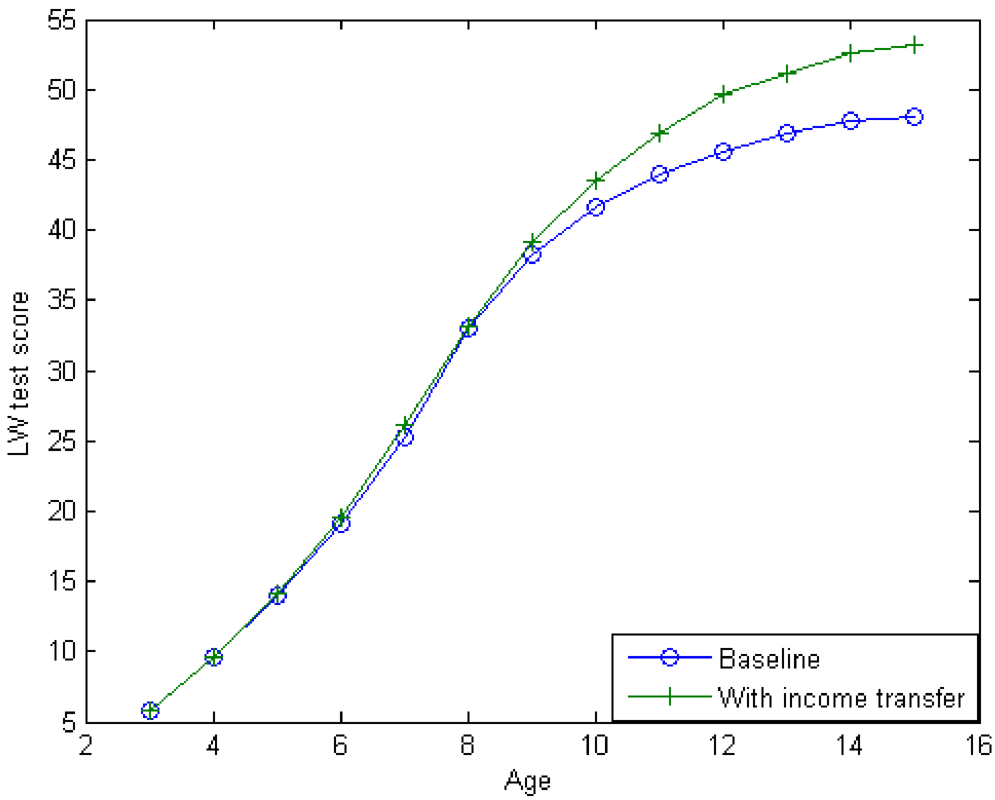
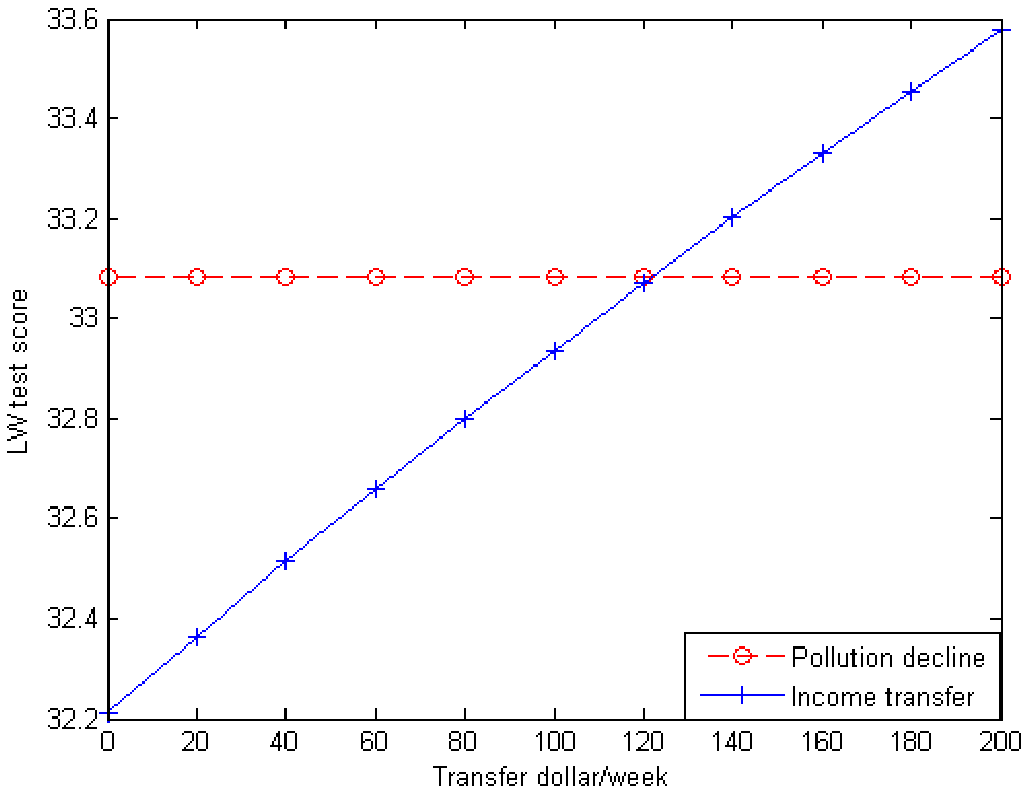


Figure 3.17: Average Child's Test Scores



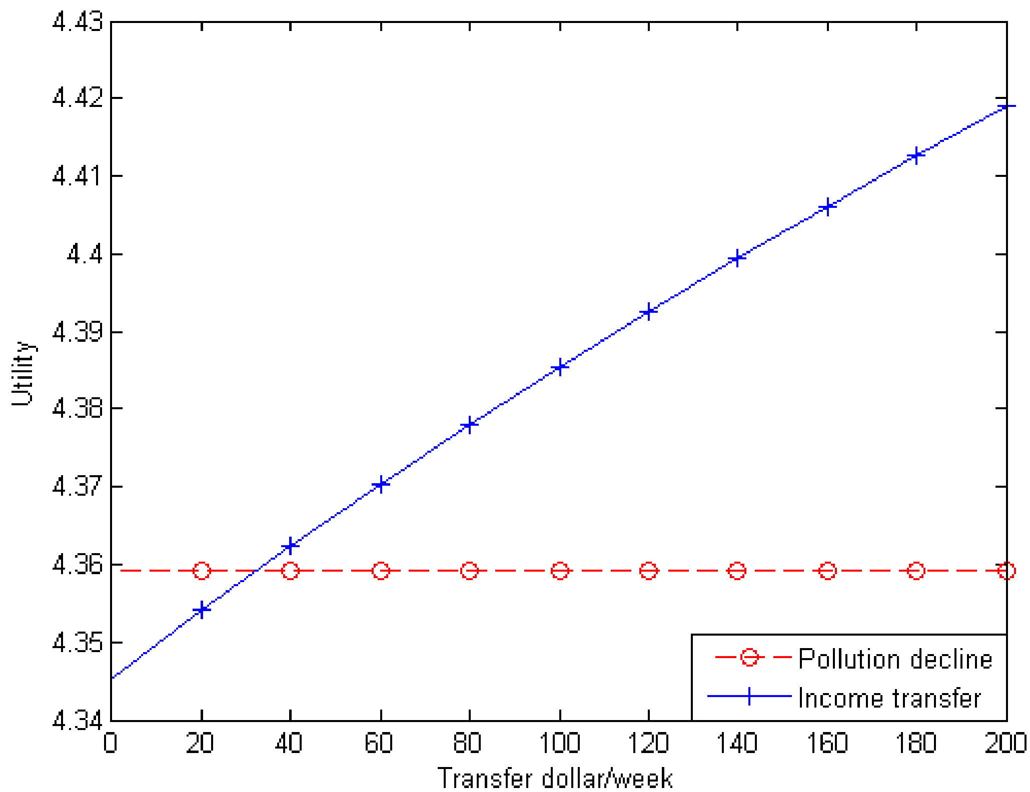
Note: In the baseline I draw the income value from the data. For the treatment group I give weekly transfer of one standard deviation income.

Figure 3.18: Average LW Test Score Under Two Separate Policies



Note: Average LW test score under two separate policies: pollution reduction, income transfer. Pollution decline curve presents the average LW test score from one standard deviation reduction of pollution level for all the ages without income transfer. Income transfer curve presents the average LW test score for different amount of transfer without pollution reduction. The graph shows that income transfer of around \$120 leads to the same average LW test score as one standard reduction in pollution levels for all the ages.

Figure 3.19: Average Household's Utility Levels under Two Separate Policies



Note: Average household's utility levels under two separate policies: pollution reduction, income transfer. Pollution decline curve presents the average household's utility levels from one standard deviation reduction of pollution level for all the ages without income transfer. Income transfer curve presents the household's utility levels for different amount of transfer without pollution reduction. The graph shows that income transfer of around \$30 leads to the same average household's utility levels as one standard reduction in pollution levels for all the ages.

3.9 Tables

Table 3.1: Data Sample

	Used variable from the data	Years	Source
I_t	Annual family income	1996,98,2000,02,04,06	PSID
h_t	Parents' labor hours	1996,98,2000,02,04,06	PSID
θ_t	Letter-Word score	1997,2002,07	CDS
τ_t^p	Active time parents spend with child	1997,2002,07	CDS
τ_t^c	Time that child spend at school study alone	1997,2002,07	CDS
x_t	Pollution	1997-2007	EPA

Table 3.2: Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Family income	971.98	600.29	100.06	2882.64	5302
LW Test Score	35.98	14.71	1	57	1940
Total time on education alone	30.97	15.9	0	88.5	1899
Parents' active time with child	23.74	14.9	0	143.5	1899
Ozone design value	5.43	0.89	1.85	9.45	4086

Notes: This table shows the sample's characteristics at the beginning of the study, year 1997.

Table 3.3: Summary Statistics for Sample at 1997

Variable	**Mean**	**Std. Dev.**	**N**
Mothers education (years)	14	1.9	854
Family size	2.2	0.9	958
Mothers age at first birth	24.28	5.74	855
Family income ($/week)	899.39	589.11	851

Notes: This table shows the sample's characteristics at the beginning of the study, year 1997.

Table 3.4: Estimated Parameters

	Parameter	Estimate	SE
Panel A: preference parameters			
Consumption impact	α_1	0.2569	0.0099
Parents leisure impact	α_2	0.1972	0.0240
Child leisure impact	α_3	0.1936	0.0069
Child's human capital impact	α_4	0.3523	0.0109
Child's human capital multiplier at final period	φ	9.0648	15.0449
Panel B: technology parameters			
Monetary investment parameter	p_1	0.0381	1.2955
Monetary investment intercept parameter	p_2	0.0712	0.0351
Parental time parameter	q_1	0.2522	0.0985
Parental time intercept parameter	q_2	2.5191	2.2786
Education time alone parameter	r_1	-0.0003	0.0002
Education time alone intercept parameter	r_2	5.2308	11.6220

Table 3.5: Sample Fit for the LW Test Scores

Age Category	Letter-Word Test Scores	
	Data	Simulated
3-6	9.51	10.53
7-10	34.09	33.62
11-15	45.61	45.07

Table 3.6: Sample Fit for Time Investment

Age Category	Parents time with child (hr/week)		Childs time on education alone (hr/week)	
	Data	Simulated	Data	Simulated
3-6	33.88	23.68	18.28	13.70
7-10	23.46	20.01	33.99	31.30
11-15	19.73	19.83	35.74	38.23

3.10 Model Solution Detail

Solving the household's optimization problem using the backward induction leads to the following first order conditions:

$$\frac{\beta\delta_{3,t}(1+p_1x_t)A_t}{e_t+p_1e_tx_t+p_2}-\frac{\alpha_1}{I_t-e_t}=0$$

$$\frac{\beta\delta_{4,t}(1+q_1x_t)A_t}{\tau_{pt}+q_1\tau_{pt}x_t+q_2}-\frac{\alpha_2}{T-L_t-\tau_{pt}}-\frac{\alpha_3}{T-\tau_{pt}+\tau_{ct}}=0$$

$$\frac{\beta\delta_{5,t}(1+r_1x_t)A_t}{\tau_{ct}+r_1\tau_{ct}x_t+r_2}-\frac{\alpha_3}{T-\tau_{pt}+\tau_{ct}}=0$$

For M periods of child development process A_t is calculated as

$$A_M=\varphi,$$

$$A_{M-1}=\alpha_4+\beta\delta_{1,M}\varphi,$$

$$\vdots$$

$$A_t=\alpha_4+\beta\delta_{1,t+1}A_{t+1},$$

$$\vdots$$

$$A_1=\alpha_4+\beta\delta_{1,2}A_2,$$

Where βA_t is the period t marginal utility of the period $t+1$ log child skill to the household, i.e. $\frac{\partial V_t(.)}{\partial \ln\theta_{t+1}}$. The future marginal utility depends on the flow utility of child skill which is measured by α_4, the productivity of child skill in forming the next period child skill which is measured by δ_{1t}, and the discount factor β. In the last period the household only values child skill at the end of period M with weight of φ. Hence, there is no flow utility of period $M+1$ and only the marginal utility is weighted by φ without adding α_4.

From the first order conditions the optimal value of monetary investment on child's skill, e_t, is

$$e_t^* = \frac{\beta\delta_{3,t}(1+p_1x_t)A_tI_t - \alpha_1p_2}{(1+p_1x_t)(\alpha_1+\beta\delta_{3,t}A_t)}$$

The optimal value of parental time, τ_{pt}, is a solution for the quadratic equation of $a\tau_{pt}^2 + b\tau_{pt} + c = 0$ where

$$a = (1+q_1x_t)[\alpha_2+\alpha_3+\beta A_t\delta_{3t} + \frac{\beta A_t\delta_{4t}(\alpha_2+\beta A_t\delta_{3t})}{\alpha_3+\beta A_t\delta_{4t}}]$$

$$\begin{aligned} b = \;& (1+q_1x_t)(\alpha_2T+\alpha_3T_1+\beta A_t\delta_{3t}(T_1+T)) - q_2(\alpha_2+\alpha_3) \\ & + {}^{(1+q_1x_t)(\alpha_2+\beta A_t\delta_{3t})}\!/_{(1+r_1x_t)(\alpha_3+\beta A_t\delta_{4t})}(\alpha_3r_2 - T\beta A_t\delta_{4t}(1+r_1x_t)) \\ & - {}^{\beta A_t\delta_{4t}}\!/_{(\alpha_3+\beta A_t\delta_{4t})}(T_1\beta A_t\delta_{3t}(1+q_1x_t) - \alpha_2q_2) \end{aligned}$$

$$\begin{aligned} c = \;& {}^{(T_1\beta A_t\delta_{3t}(1+q_1x_t)-\alpha_2q_2)}\!/_{(1+r_1x_t)(\alpha_3+\beta A_t\delta_{4t})}(T\beta A_t\delta_{4t}(1+r_1x_t) - \alpha_3r_2) \\ & + (q_2(\alpha_2T+\alpha_3T_1) - T_1T\beta A_t\delta_{3t}(1+q_1x_t)) \end{aligned}$$

The rest of the optimal values are as following

$$\tau_{ct}^* = \frac{\beta\delta_{4,t}A_t(1+r_1x_t)(T-\tau_{pt}^*) - \alpha_3r_2}{(1+r_1x_t)(\alpha_3+\beta\delta_{4,t}A_t)},$$

$$l_{pt}^* = T - L_t - \tau_{pt}^*,$$

$$l_{ct}^* = T - \tau_{pt}^* - \tau_{ct}^*,$$

3.11 Identifying Preference Parameters

From 3.10, I can simplify time investments as following

$$\tau_{pt}^* = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

$$\tau_{ct}^* = \frac{\beta\delta_{4,t}A_t(1+r_1x_t)(T-\tau_{pt}^*) - \alpha_3r_2}{(1+r_1x_t)(\alpha_3+\beta\delta_{4,t}A_t)}$$

Parameters a, b, and c are presented in section 3.10. Technically there are two equations in hand and four parameters to be identified. Since A_t is time varying, actual measures of time investments for a household in two different points in time provides four equations to back out the preference parameters, α_2, α_3, α_4, and φ. Similarly, time investments for two different households can do the job ,too. Empirically if there are enough variation in time investments and children's ages, I can recover preference parameters.

REFERENCES

"The ozone weekend ozone effect in california", Tech. rep., California EPA Air Resources Board (2014).

Almond, D., L. Edlund and M. Palme, "Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in sweden", The Quarterly Journal of Economics **124**, 4, 1729–1772 (2009).

Back, D., N. V. Kuminoff, E. V. Buren and S. V. Buren., "National evidence on air pollution avoidance behavior", Land Economics Working Paper (2013).

Becker, G. S., "A theory of the allocation of time", The Economics Journal **75**, 299, 493–517 (1965).

Bento, A. M., M. Freedman and C. Lang, "Redistribution, delegation, and regulators incentives: Evidence from the clean air act", Tech. rep., working Paper (2013).

Bharadwaj, P., M. Gibson, J. G. Zivin and C. A. Neilson, "Gray matters: Fetal pollution exposure and human capital formation", The National Bureau of Economic Research **Working Paper** (2014).

Blanchard, C. L. and S. J. Tanenbaum, "Differences between weekday and weekend air pollutant levels in southern california", Journal of the Air & Waste Management Association **53**, 7, 816–828 (2003).

Boca, D. D., C. Flinn and M. Wiswall, "Household choices and child development", Review of Economic Studies **81**, 3, 137–185 (2014).

Brauer, M., C. Lencar, L. Tamburic, M. Koehoorn, P. Demers and C. Karr, "A cohort study of traffic-related air pollution impacts on birth outcomes", Environmental Health Perspectives **116**, 5, 680–686 (2008).

Cameron, A. C. and P. K. Trivedi, *Microeconometrics: Methods and Applications* (Cambridge University Press, The Edinburgh Building, Cambridge CB2 2RU, UK, 2005).

Chay, K. Y. and M. Greenstone, "Does air quality matter? evidence from the housing market", Journal of Political Economy **113**, 2, 376–424 (2005).

Cunha, F. and J. Heckman, "The technology of skill formation", The American Economic Review **97**, 2, 31–47 (2007).

Cunha, F., J. J. Heckman, L. Lochner and D. V. Masterov, "Interpreting the evidence on life cycle skill formation", Handbook of the Economics of Education **Chapter 12**, 697–812 (2006).

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**, 3, 1003–1030 (2005).

Currie, J., M. Neidell and J. F. Schmieder, “Air pollution and infant health: Lessons from new jersey”, Journal of Health Economics **28**, 688–703 (2009).

Currie, J. and R. Walker, “Traffic congestion and infant health: Evidence from e-zpass”, The National Bureau of Economic Research **working paper**, 15413 (2009).

Dahl, G. B. and L. Lochner, “The impact of family income on child achievement: Evidence from the earned income tax credit”, American Economic Review **102**, 5, 1927–1956 (2012).

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

Evens, A., D. Hryhorczuk, B. P. Lanphear, K. M. Rankin, D. A. Lewis, L. Forst and D. Rosenberg, “The impact of low-level lead toxicity on school performance among children in the chicago public schools: A population-based retrospective cohort study”, Environmental Health **14**, 21, 1–9 (2015).

Fujita, E. M., W. R. Stockwell, D. E. Campbell, R. E. Keislar and D. R. Lawson, “Evolution of the magnitude and spatial extent of the weekend ozone effect in california’s south coast air basin, 1981-2000”, Journal of the Air & Waste Management Association **53**, 7, 802–815 (2003).

Goolsbee, A. and P. J. Klenow, “Valuing consumer products by the time spent using them: An application to the internet”, The American Economic Review **96**, 2, 108–113 (2006).

Heckman, J. J., “Schools, skills and synapses”, Economic Inquiry **46**, 3, 289–324 (2008).

Isen, A., M. Rossin-Slater and W. R. Walker, “Every breath you take - every dollar you’ll make: The long-term consequences of the clean air act of 1970”, The National Bureau of Economic Research **Working Paper** (2014).

Knittel, C. R., D. L. Miller and N. J. Sanders, “Caution, drivers! children present: Traffic, pollution, and infant health”, NBER Working Paper , 17222 (2011).

Krueger, A. B., “Experimental estimates of education production functions”, The Quarterly Journal of Economics **114**, 2, 497–532 (1999).

Larson, D. M. and S. L. Shaikh, “Recreation demand choices and revealed values of leisure time”, Economic Inquiry **42**, 2, 264–278 (2004).

Lefohn, A. S., M. J. Hazucha, D. Shadwick and W. C. Adams, “An alternative form and level of the human health ozone standard”, Inhalation Toxicology **22**, 12, 999–1011 (2010).

Moretti, E. and M. Neidell, “Pollution, health, and avoidance behavior: Evidence from the ports of los angeles”, NBER Working Paper , 14939 (2009).

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

Neidell, M. and P. L. Kinney, "Estimates of the association between ozone and asthma hospitalizations that account for behavioral responses to air quality information", Environmental Science and Policy **13**, 2, 97–103 (2010).

Neidell, M. J., "Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma", Journal of Health Economics **23**, 1209–1236 (2004).

Neilson, C. and S. Zimmerman, "The effect of school construction on test scores, school enrollment, and home prices", Journal of Public Economics **120**, 18–31 (2014).

Palmquist, R. B., D. J. Phaneuf and V. K. Smith, "Short run constraints and the increasing marginal value of time in recreation", Environmental and Resource Economics **46**, 1, 19–41 (2010).

Phaneuf, D. J., "Can consumption of convenience products reveal the opportunity cost of time?", Economics Letters **113**, 1, 92–95 (2011).

Qin, Y., G. Tonnesen and Z. Wang, "Weekend/weekday differences of ozone, nox, co, vocs, pm10 and the light scatter during ozone season in southern california", Atmospheric Environment **38**, 19, 3069–3087 (2004).

Ransom, M. R. and C. A. Pope, "External health costs of a steel mill", Contemporary Economic Policy **13**, 2, 86–97 (1995).

Reyes, J. W., "Environmental policy as social policy? the impact of childhood lead exposure on crime", The B.E. Journal of Economic Analysis & Policy **7**, 1, 1–43 (2007).

Sanders, N. J., "What doesnt kill you makes you weaker", The Journal of Human Resources **47**, 3, 826–850 (2011).

Schlenker, W. and W. R. Walker, "Airports, air pollution, and contemporaneous health", The Review of Economic Studies **83**, 2, 768–809 (2016).

Shaikh, S. L. and D. M. Larson, "A two-constraint almost ideal demand model of recreation and donations", The Review of Economics and Statistics **85**, 4, 953–961 (2003).

Vahedi, S., "Accounting for exposure time in estimating the impact of ozone pollution on respiratory disease and asthma", Working paper (2013).

Yarwooda, G., J. Granta, B. Kooa and A. M. Dunker, "Modeling weekday to weekend changes in emissions and ozone in the los angeles basin for 1997 and 2010", Atmospheric Environment **42**, 16, 3765–3779 (2008).

Zivin, J. G. and M. Neidell, "Days of haze: Environmental information disclosure and intertemporal avoidance behavior", Journal of Environmental Economics and Management **58**, 2, 119–128 (2009).

Zivina, J. G. and M. Neidell, "Environment, health, and human capital", Journal of Economic Literature **51**, 3, 689–730 (2013).

Zivina, J. G. and M. Neidell, "Temperature and the allocation of time: Implications for climate change", Journal of labor economics **32**, 1, 1–26 (2014).