Pollution Avoidance and Willingness-to-Pay: Evidence from Travel Mode Choice in Beijing*

Dingyi Li Shanjun Li C.-Y. Cynthia Lin Lawell

August 2022

Abstract

We estimate the short-term willingness-to-pay (WTP) to avoid air pollution by developing a model to capture the trade-offs between avoidance behavior and its costs. In particular, we use fine-scale travel survey data in Beijing to model the trade-offs between indoor and outdoor travel modes for compulsory work trips during high polluted hours. Our model indicates that the short-term WTP, which we estimate to be 0.00223 dollars per hour to avoid 1 $\mu g/m^3$ of ambient fine particles ($PM_{2.5}$), forms the lower bound for the long-term WTP, which is around 19.53 dollars per year to avoid 1 $\mu g/m^3$ $PM_{2.5}$. Our estimation strategy uses a machine learning IV method in a high dimensional econometrics setting. We find that a longer potential exposure to air pollution prevents people from walking and cycling. People older than 55 years old, who are more vulnerable to pollution and thus more likely to avoid pollution, have a 28% higher WTP than the young. Likewise, richer people, who value their health more, are willing to avoid a unit of pollution with 36% more cost. Finally, we find evidence that information affects the behavioral adjustment: people start to reduce their exposure to the toxic air only after extensive media coverage of air pollution.

*D. Li: Cornell University; dl922@cornell.edu S. Li: Cornell University; sl2448@cornell.edu. Lin Lawell: Cornell University; clinlawell@cornell.edu. We thank Panle Jia Barwick, Moshe Ben-Akiva, Luming Chen, Todd Gerarden, Dalia Ghanem, Yongmiao Hong, Cathy Kling, Francesca Molinari, Yang Ning, Avralt-Od Purevjav, Chen Qiu, Ivan Rudik, Alberto Salvo, Shuyang Si, Haokun Sun, Marten Wegkamp, Dennis Wesselbaum, Binglin Wang, and Tianli Xia for helpful comments and discussions. We benefited from comments from conference participants at the North American Summer Meetings of the Econometric Society and the World Conference of the Spatial Econometrics Association (SEA). We received funding for our research from a Cornell Graduate School Conference Travel Grant. S. Li and Lin Lawell are Faculty Fellows at the Cornell Atkinson Center for Sustainability. All errors are our own.

1 Introduction

Air pollution has adverse impacts on human health, particularly in developing countries. While a growing literature in economics quantifies the causal effect of air pollution on health (Neidell, 2004; Currie and Walker, 2011; Schlenker and Walker, 2016; Deryugina et al., 2019; Cole et al., 2020; Barwick et al., 2021), there has been little research to date focusing on how pollution distorts behavior, which might cause the health effects to be underestimated. Air pollution may influence behavior by affecting one's mood, via physical or mental health channels (Chen et al., 2018; He et al., 2019; Liu and Salvo, 2018; Salvo, 2020; Chu et al., 2021); or by causing those aware of the harmful effects of air pollution to take action to avoid it (Barwick et al., 2021). On the one hand, researchers can use the distortion in behavior to reveal people's cognition about the pollution. On the other hand, not taking these behaviors into account biases estimates related to the dirty air.

In this paper, we focus on air pollution avoidance behavior and develop a model to capture the trade-offs between avoidance behavior and its costs. In the environment and health economics literature, researchers mention the existence of avoidance behaviors, acknowledging that people may strategically avoid air pollution by staying at home (Neidell, 2009; Bäck et al., 2013; Deryugina et al., 2019; Deschenes et al., 2020; Barwick et al., 2020). The real air pollution exposure might therefore be lower than what is in the data record, and thus empirical results may underestimate the actual welfare loss due to air pollution. By focusing on avoidance behavior, we validate this concern and then evaluate people's cognition about the harmfulness of air pollution.

How people value air quality improvement, as measured by their willingness-to-pay (WTP), is an important determinant of the optimal level of environmental regulation (Michael and Kelsey, 2015). Nevertheless, well-identified estimates of this parameter are scarce for air quality in developing countries. Therefore, an accurate estimation of the WTP for air quality is crucial for individuals' well-being and for policy design.

The revealed preference approach, which explores the correlation between pollution levels and house or filter prices, is currently a common method for estimating the price of clean air (Chay and Greenstone, 2005; Bayer et al., 2009; Ito and Zhang, 2020). These estimates in the literature are based on a specific market over long-term air pollution reduction, assuming that consumers could access complete and perfect information on future air pollution levels, and do not take into account other avoidance strategies.

Meanwhile, a large literature has documented that congestion increases air pollution (Zhong et al., 2017; Green et al., 2020; Lu et al., 2021), but there has been limited research investigating the short-term reverse effect of air pollution on individual behavior. Our setting enables us to relax some of the assumptions commonly imposed in the literature in order to estimate the WTP for clean air. In our compulsory work trip scenario, most citizens have to choose between an indoor and outdoor travel mode to work when the air pollution can be directly observed. Moreover, accurate information is readily available on phone applications for those decision-makers, so we assume that they bear air pollution levels in mind when they make their decisions.

We analyze how air pollution affects travel mode decisions in China with a large and detailed hourly household-level data set on travel mode decisions collected in Beijing. We also collect data on hourly air pollution, weather, wind speed, and wind directions inside and around Beijing. To address the potential endogeneity of air pollution due to reverse causality and measurement error, we use atmospheric chemistry and a machine learning instrumental variable (IV) method in a high dimensional econometrics setting to select and construct strong instruments for air pollution from a set of over 2,115 variables. Building on the previous literature on the least absolute shrinkage and selection operator (LASSO) by Belloni et al. (2012), we show the estimator's properties under the weak instrument framework. Our proof not only works as the theoretical foundation for our empirical analysis but also contributes to the crosscutting area of machine learning and econometrics empirically. We use economic intuition to further reduce the number of selected instruments and deal with the potential bias of the estimator. Our final instrument is the southeast wind in Tangshan, which blows from upwind industrial areas towards Beijing.

According to our results, the short-term WTP is 0.00223 dollars per hour to avoid 1 $\mu g/m^3$ of ambient fine particles ($PM_{2.5}$). The long-term WTP is over 19.53 dollars per hour to avoid 1 $\mu g/m^3$ of $PM_{2.5}$. We find that a longer potential exposure to air pollution prevents people from walking and cycling. People older than 55 years old, who are more vulnerable to pollution and thus more likely to avoid pollution, have a 28% higher WTP than the young. Likewise, richer people, who value their health more, are willing to avoid a unit of pollution with 36% more cost.

We also find evidence that information affects the behavioral adjustment. We find that the WTP for air quality is significant and positive in the year 2014, after China launched its air quality monitoring and disclosure program. In contrast, we find no significant WTP in 2010, when people were still by and large unaware of the air pollution hazards in China. Our results suggest that people started to reduce their exposure to the toxic air only after extensive media coverage of air pollution.

We make three main contributions to the literature. First and foremost, we add to the growing literature on measuring the WTP for clean air by using a novel approach based on short-term avoidance behavior. The traditional revealed preference method, which is employed for example in the seminal air filter paper by Ito and Zhang (2020), is based on the relationship between long-term air pollution exposure reduction and its price. In contrast, our method, which exploits the trade-off between expensive indoor travel modes and less expensive outdoor travel modes, focuses on the pattern between air pollution fluctuation and immediate defensive behavioral reactions. We relax the three common assumptions made in the previous WTP literature: representative consumer, rational expectations, and exclusive avoidance. These assumptions are further discussed in Section 7.

Second, we extend the literature on the effects of pollution information (Barwick et al., 2020). During 2013-2014, China launched a nationwide, real-time air quality monitoring and disclosure program. The conventional understanding or perception of air pollution and its harmfulness can change significantly and quickly with information and media coverage. The dissemination of information is conducive to public welfare. In particular, air pollution in China had been perceived as harmless fog by the public in 2010, but, owing to the 2013-2014 air quality monitoring and disclosure program, was considered toxic smog in 2014. Once they were aware of air pollution and its harmfulness, people began to take protective measures on smoggy days (Barwick et al., 2020).

Finally, our econometric analysis contributes to the overlap between machine learning and econometrics theoretically and empirically. Belloni et al. (2012) develop results for the use of LASSO and post-LASSO methods to form first-stage predictions and estimate optimal instruments in linear instrumental variables (IV) models with many instruments when the first stage is approximately sparse – that is, when there exists a relatively small set of important instruments whose identities are unknown that well-approximate the conditional expectation of the endogenous variables given the instruments (Bunea et al., 2007). We extend the high dimensional setting of Belloni et al. (2012) by relaxing the sparse IV assumption to a weak IV assumption. From an empirical point of view, we show that economic intuition can effectively assist the theoretical framework with the potential biasedness.

The balance of this paper proceeds as follows. Section 2 sets up the theoretical foundation for our empirical analysis. Section 3 provides the background of the avoidance behavior under the travel scenario in China. Section 4 describes the data. Section 5 presents the empirical strategies and Section 6 presents the results. We carry out our welfare analyses and discuss the relaxed assumptions in Section 7. Finally, Section 8 concludes.

2 Background

2.1 Pollution Information and Avoidance Behavior

Air pollution has negative effects on health, productivity, and quality of life. Air pollution is a severe environmental and health issue in China, where the daily average concentration of fine particulate matter ($PM_{2.5}$) is over 60 $\mu g/m^3$, or about six times that in the World Health Organization guideline (Barwick et al., 2020). Fine particulate matter ($PM_{2.5}$)[1] is particularly deadly, with an 18% increase in lung cancer per 5 $\mu g/m^3$, as it can penetrate deeper into the lungs (Raaschou-Nielsen et al., 2013). As a result, people may take actions to avoid polluted air, including wearing face masks, purchasing air purifiers, reducing their outdoor activities, or staying at home; as a consequence, the real air pollution exposure is lower than that of the data record. Avoidance behaviors also widely exist in travel decisions. Recent studies show that when an air quality alert is issued, the amount of cycling could shrink remarkably (Tribby et al., 2013; Saberian et al., 2017;

[1] Fine particulate matter ($PM_{2.5}$) consist of tiny droplets in the air that are two and one half microns or less in width. Rather than having a single chemical composition, $PM_{2.5}$ is a mixture of various compounds including nitrates, sulfates, ammonium, and carbon (Kundu and Stone, 2014). In addition to natural sources, $PM_{2.5}$ is created from atmospheric conversion of power plant and auto emissions.

Zhao et al., 2018).

People's views towards air pollution in China were different before the year 2013, however. In 2013, China launched a nationwide, real-time air quality monitoring and disclosure program, the first of its kind in history. Until very shortly prior to the reform, there was a lack of awareness in air pollution exposure, and Chinese people believed that the air pollution was merely "fog" (Barwick et al., 2020). Based on this idea, we compare the effects of daily air pollution on outdoor travel mode shares in 2010, prior to the reform, to that in 2014, after the reform.

To investigate the existence of avoidance behavior and an informational effect, we examine the relationship between the daily outdoor mode share and air pollution in Figure 1.[2] The negative correlation between air pollution and outdoor travel in the year 2014 is evidence for the existence of avoidance behavior in 2014. In contrast, the positive correlation in the year 2010 suggests that in 2010 the WTP, which depends on information, was based on the expected cost of fog rather than the actual damage from smog. In 2010, people might not have been willing to pay anything for clean air, despite the adverse effects of air pollution on their health. With the dissemination of knowledge about air pollution during China's 2013-2014 air quality monitoring and disclosure program, the WTP in 2014 better reflects the true social welfare loss from air pollution.

2.2 Outdoor and Indoor Air Pollution

To support our identification strategy, we explore the variation between outdoor and indoor air pollution. Many environmental scientists have developed models to discuss this difference (Chen and Zhao, 2011).

The prevailing method to compare outdoor and indoor air pollution is the indoor/outdoor (I/O) ratio. This ratio directly represents the relationship between indoor and outdoor air pollution concentrations. The I/O ratios in the literature are often for developed countries where indoor smoking and cooking are among the main pollution sources. As a consequence, the literature tends to show that the average level of indoor air pollution is higher than that of outdoor air pollution, which may appear to contradict common sense and our hypothesis that people avoid

[2] Travel data are from the Beijing Household Travel Survey (BHTS). The full data description is in Section 4. We summarize 6 modes $\{Walking, Car, Subway, Bus, Taxi, Bicycle\}$, and we define the outdoor travel mode as $\{Walking, Bicycle\}$.

the air pollution by participating in indoor instead of outdoor activities. In contrast, in Beijing as well as in many developing countries, the main sources of air pollution are outdoors, including dirty firms and heavy traffic congestion. Moreover, smoking is forbidden on China's indoor public transportation, and cooking is almost impossible in transportation facilities.

The indoor/outdoor ratio is related to the infiltration factor, which represents the equilibrium fraction of ambient particles that penetrate indoors. While there are many sophisticated engineering models that estimate the infiltration factor, the intuition of these estimates stems from the regression of indoor concentration on outdoor concentration. The infiltration factors measured by different researchers vary in a relatively large range of $0.3 - 0.82$ for $PM_{2.5}$. In our empirical analysis, we use the mean value of the infiltration factors in the literature, 0.56, as our indoor/outdoor ratio to distinguish the indoor/outdoor travel mode pollution exposure, because it is in the ballpark of most of the literature. We assume that individuals are aware of the difference between outdoor and indoor air pollution. The outdoor air pollution can be visually observed, and accurate information are readily available on mobile phone applications.[3]

3 Theoretical Foundations

In this section, we discuss the theoretical foundations that motivate our empirical analysis.

3.1 Utility Maximization

We develop a model to capture the daily trade-offs between avoidance behavior and its costs. In particular, we model the trade-offs between indoor and outdoor travel modes for compulsory work trips during high polluted hours.

Formally, on any given work day, an individual chooses their travel mode m for compulsory work trips from the choice set M, as well as health expenditures e and the consumption level x, to maximize their utility subject to a budget constraint that captures the idea that an individual who spends less time traveling and saves time for work can get extra salary to pay for the either travel costs, health expenditures, or consumption. The individual's daily work travel mode choice

[3] In our future work, we will conduct a short survey to investigate people's expectations about the indoor and outdoor air pollution difference.

optimization problem is given by:

$$\begin{aligned}
\max_{m\in M, e>0, x>0, t_l, t_a} \quad & f(h(p_m; e), L) + x + \varepsilon_m \\
s.t. \quad & \gamma_0 \cdot (x + e + c_m) \leq \gamma_1 W, \\
& W = \widetilde{W} - (t_l + t_m - t_a)^+ \\
& L = 24 - W - t_m \\
& t_l >= t_l*
\end{aligned} \tag{1}$$

The optimal schedule decided yesterday is

$$\begin{aligned}
\max_{m\in M, e>0, x>0, t_l^*, t_a^*} \quad & f(h(0; e), L) + x + \varepsilon_m \\
s.t. \quad & \gamma_0 \cdot (x + e + c_m) \leq \gamma_1 W, \\
& W = \widetilde{W} - (t_l^* + t_m - t_a^*)^+ \\
& L = 24 - W - t_m
\end{aligned} \tag{2}$$

where $h(\cdot)$ is health or well-being as a function of pollution exposure p_m and health expenditures e; x is the consumption good; ε_m is the unobserved random variable of taste, which we assume follows the type-I extreme value distribution for the simplicity of the welfare analysis and estimation; c_m and t_m are the travel cost and travel time, respectively, of travel mode m; γ_1 is a constant value of working; γ_0 is a constant value of money which can be normalized as 1 if we adjust the standard error in the utility; t_a^* is the optimal arriving time; t_l^* is the optimal leaving time; W is the work hours and L is the Leisure hours. Pollution exposure p_m depends on the outdoor air pollution level P_o, the indoor pollution level P_i, and whether the travel mode m is indoors or outdoors, and is given by:

$$p_m = P_i T + (P_o - P_i)\mathbf{1}[m = outdoor]t_m. \tag{3}$$

We assume that, conditional on health expenditures e, the health function is a decreasing and concave function of pollution exposure p_m, which captures the possibly nonlinear and convex costs of pollution exposure: $h'(\cdot; e) \equiv \frac{\partial h(\cdot;e)}{\partial p_m} < 0$ and $h''(\cdot; e) \equiv \frac{\partial^2 h(\cdot;e)}{\partial p_m^2} \leq 0$.

We can approximate the optimization problem with:

$$\max_{m \in M} \quad h'(\bar{P}T;\bar{e})(P_o - P_i)\mathbf{1}[m = \mathit{outdoor}]t_m - \gamma_1 t_m - \gamma_2 c_m + \varepsilon_m, \tag{4}$$

where $\bar{P}$ is the average air pollution exposure over the year, and $\bar{e}$ is the average health expenditure over the year. The proof of the approximation is in the appendix.

Under the assumption that indoor pollution P_i is a fixed proportion $\alpha \in (0,1)$ of outdoor pollution P_o, i.e., $P_i = \alpha P_o$, the approximate optimization problem reduces to:

$$\max_{m \in M} \quad h'(\bar{P}T;\bar{e})(1-\alpha)P_o\mathbf{1}[m = \mathit{outdoor}]t_m - \gamma_1 t_m - \gamma_2 c_m + \varepsilon_m, \tag{5}$$

where α is the indoor/outdoor (I/O) ratio.

Equation (5) is the foundation of our empirical analysis. It reflects the trade-off between outdoor pollution exposure and indoor travel cost.

Given the utility function $u(\cdot)$, the marginal cost of pollution exposure, MC_p, which is the negative of the marginal value of air pollution exposure, and which captures the monetized health costs of pollution exposure (which can include, for example, health expenditures, costs of medical treatment, disutility of poor health, etc.), is given by:

$$MC_p = -\frac{\frac{\partial u}{\partial p_m}}{\frac{\partial u}{\partial c_m}} = -\frac{h'(\cdot);\bar{e}}{\gamma_2}. \tag{6}$$

The marginal value of time, VOT, is given by:

$$VOT = \frac{\frac{\partial u}{\partial t_m}}{\frac{\partial u}{\partial c_m}} = \frac{\gamma_1}{\gamma_2}. \tag{7}$$

Let β be the coefficient on $P_o\mathbf{1}[m = \mathit{outdoor}]$ in the utility function in Equation (5):

$$\beta = h'(\cdot;\bar{e})(1-\alpha)t_m. \tag{8}$$

Since $h'(\cdot;\bar{e}) < 0$, $\beta < 0$.

Then the marginal cost of pollution exposure, MC_p can be approximated by:

$$MC_p = -\frac{h'(\cdot);\bar{e}}{\gamma_2} = -\frac{\beta}{(1-\alpha)\gamma_2 \bar{t}_o}, \tag{9}$$

where $\bar{t}_o$ is the average outdoor trip time.

3.2 Welfare Analysis

After deriving the parameters in the utility function, there are two alternatives to measure the welfare change due to an air pollution event. One is the pure willingness-to-pay (WTP), the other is the Hicksian compensating variation (CV). They both measure the money value of the event but in slightly different ways.

The pure willingness-to-pay (WTP) is more common in the economics literature on air pollution because of its simplicity. It answers the following question: how much would an individual be willing to pay to experience a change in an exogenous pollution variation holding all else constant? In the literature, "holding all else constant" includes holding the choice of the individual constant. In other words, the pure WTP assumes that the individual remains at the old optimal alternative when the exogenous variable changes, and therefore does not account for any behavioral adjustment (Bockstael and McConnell, 2007). The pure WTP thus implicitly assumes that the individual does not change their travel mode when the air pollution changes. As a consequence, an individual who drove when air pollution is severe is assumed to still drive even when the air quality improves. Similarly, an individual who walked when the air is clean is assumed to still walk even when air pollution increases.

Suppose we want to measure the welfare change due to a reduction in pollution level from a dirty pollution level P^0 to an improved level P^1. The associated reduction in pollution exposure from travel mode choice m when the pollution level reduces from a dirty pollution level P^0 to an improved level P^1 is given by:

$$p_m^1 - p_m^0 = \alpha(P_o^1 - P_o^0)T + (1-\alpha)(P_o^1 - P_o^0)\mathbf{1}[m = outdoor]t_m. \tag{10}$$

In our model, the pure WTP per day for an air quality improvement from a dirty pollution level P^0 to an improved level P^1, which is equal to the reduction in monetized health costs from a reduction in pollution level from a dirty pollution level P^0 to an improved level P^1, holding all else constant (including the travel mode choice), is given by:

$$\begin{aligned}
WTP &= \frac{h(p_m^1;\bar{e}) - h(p_m^0;\bar{e})}{\frac{\partial u}{\partial c_m}} \\
&\approx \frac{h'(\cdot;\bar{e})(p_m^1 - p_m^0)}{\frac{\partial u}{\partial c_m}} \\
&= \frac{h'(\cdot;\bar{e})}{\gamma_2}(p_m^1 - p_m^0) \\
&= -MC_p \cdot (p_m^1 - p_m^0) \\
&= \frac{h'(\cdot;\bar{e})}{\gamma_2}(p_m^1 - p_m^0) \\
&= \frac{\beta}{(1-\alpha)\gamma_2 \bar{t_o}}(p_m^1 - p_m^0) \qquad (11)\\
&= \frac{\beta}{(1-\alpha)\gamma_2 \bar{t_o}}\left(\alpha(P_o^1 - P_o^0)T + (1-\alpha)(P_o^1 - P_o^0)\mathbf{1}[m = outdoor]t_m\right) \\
&= \frac{\beta}{(1-\alpha)\gamma_2 \bar{t_o}}\left(\alpha(P_o^1 - P_o^0)T + (1-\alpha)(P_o^1 - P_o^0)\bar{t_o}\right) \\
&= \frac{\beta}{(1-\alpha)\gamma_2 \bar{t_o}}\left((1-\alpha)\bar{t_o} + \alpha T\right)(P_o^1 - P_o^0) \\
&= \left(\frac{\beta}{(1-\alpha)\gamma_2 \bar{t_o}}(1-\alpha)\bar{t_o} + \frac{\beta}{(1-\alpha)\gamma_2 \bar{t_o}}\alpha T\right)(P_o^1 - P_o^0) \\
&= \left(\frac{\beta}{\gamma_2} + \frac{\beta\alpha T}{(1-\alpha)\gamma_2 \bar{t_o}}\right)(P_o^1 - P_o^0),
\end{aligned}$$

where $\bar{t_o}$ is the average outdoor trip time.

The pure WTP per hour for an air quality improvement from a dirty pollution level P^0 to an improved level P^1, which is equal to the reduction in monetized health costs from a reduction in pollution level from a dirty pollution level P^0 to an improved level P^1, holding all else constant (including the travel mode choice), is then given by:

$$WTP \text{ per hour} = \frac{1}{24} \cdot WTP = \frac{1}{24} \cdot \left(\left(\frac{\beta}{\gamma_2} + \frac{\beta\alpha T}{(1-\alpha)\gamma_2 \bar{t_o}}\right)(P_o^1 - P_o^0)\right), \qquad (12)$$

where $\bar{t_o}$ is the average outdoor trip time.

The concavity of the health function $h(\cdot;\bar{e})$ with respect to pollution exposure p_m, which captures the possibly nonlinear and convex costs of pollution exposure, means that by the common linear approximation $h'(P^0T;\bar{e})(P^1-P^0)T$ in the literature always yields a lower bound of the true pure willingness-to-pay.

The compensating variation (CV) is defined in money terms as the change in exogenous income necessary to restore an individual to the utility level that she experienced before the change in air pollution. Following the convention of Just et al. (1982, 2004) that the CV associated with a change that is improving should itself be positive, this means that in the case of an improvement in air quality, the CV is the maximum amount the individual would pay rather than forego the air quality improvement (Bockstael and McConnell, 2007). When individuals face a choice among discrete alternatives, the CV allows the individual to freely adjust her choice following the change in the exogenous variable (McFadden, 1999), which in our case is air quality. Thus, the CV implicitly assumes that the consumer re-optimizes and chooses the optimal travel mode when the air quality changes. Using the expression for CV derived in Bockstael and McConnell (2007), the compensating variation (CV) for a change from a dirty pollution level P^0 to a clean level P^1 is the difference in the value of the daily optimization program from the optimal travel mode choice under the clean level P^1, and the value of the daily optimization program from the optimal travel model choice under the dirty pollution level P^0, and is given by:

$$\begin{aligned} CV =& \frac{1}{\gamma_2}\cdot \\ & \Big(\max_{m\in M} \quad h'(P^1T)(1-\alpha)P_o^1\mathbf{1}[m=outdoor]t_m - \gamma_1 t_m - \gamma_2 c_m + \varepsilon_m \\ & - \max_{m\in M} \quad h'(P^0T)(1-\alpha)P_o^0\mathbf{1}[m=outdoor]t_m - \gamma_1 t_m - \gamma_2 c_m + \varepsilon_m\Big) \end{aligned} \tag{13}$$

In our random utility model, since owing to the random component ε_m, the ultimate choice of the individual is unknown to the researcher, so the expected value of CV must be computed. Using the expression for expected CV derived in Bockstael and McConnell (2007), the expected compensating variation (CV) for a change from a dirty pollution level P^0 to a clean level P^1 is given by:

$$
\begin{aligned}
E[CV] = &\frac{1}{\gamma_2} \cdot \\
&\left(\ln \left(\sum_{m \in M} \exp\left(h'(P^1 T)(1-\alpha) P_o^1 \mathbf{1}[m = outdoor] t_m - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right. \\
&\left. - \ln \left(\sum_{m \in M} \exp\left(h'(P^0 T)(1-\alpha) P_o^0 \mathbf{1}[m = outdoor] t_m - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right),
\end{aligned} \tag{14}
$$

which simplifies to:

$$
\begin{aligned}
E[CV] = &\frac{1}{\gamma_2} \cdot \\
&\left(\ln \left(\sum_{m \in M} \exp\left(\beta P_o^1 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right. \\
&\left. - \ln \left(\sum_{m \in M} \exp\left(\beta P_o^0 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right).
\end{aligned} \tag{15}
$$

The expected compensating variation (CV) per hour for a change from a dirty pollution level P^0 to a clean level P^1 is then given by:

$$
\begin{aligned}
E[CV] \text{ per hour} = &\frac{1}{24} \cdot E[CV] \\
= &\frac{1}{24} \cdot \frac{1}{\gamma_2} \cdot \\
&\left(\ln \left(\sum_{m \in M} \exp\left(\beta P_o^1 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right. \\
&\left. - \ln \left(\sum_{m \in M} \exp\left(\beta P_o^0 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right).
\end{aligned} \tag{16}
$$

A central precept of Hicksian welfare theory is the equivalence between the WTP and the CV (Zhao and Kling, 2004). It is imprecise to argue that WTP and CV are equivalent, however. If there is no avoidance behavior, then the CV is identical to the pure WTP. In the presence of possible avoidance behavior, however, the CV is no longer equivalent to the pure WTP, since it allows the individual to re-optimize and adjust her choice following the change in air quality, and therefore accounts for any avoidance behavior.[4]

If we define avoidance behavior as any change in travel mode choice when the individual re-optimizes and adjusts her choice following the change in air quality, then the cost of the avoidance behavior is the cost (in terms of monetized travel time and travel cost) of the change in travel mode choice as a result of the change in air quality. This avoidance cost can be positive, negative, or zero. The avoidance cost is positive if the change in air quality results in a change to a travel mode that has a higher cost in terms of the sum of monetized travel time and travel cost. The avoidance cost is negative if the change in air quality results in a change to a travel mode that has a lower cost in terms of the sum of monetized travel time and travel cost. In our model, the expected avoidance cost $E[CA]$ of an air quality improvement from a dirty pollution level P^0 to an improved level P^1

[4] As shown by Zhao and Kling (2004), the Hicksian equivalence between WTP and CV also breaks down in a dynamic setting.

is given by:

$$
\begin{aligned}
E[CA] =& \frac{1}{\gamma_2} \cdot \\
& \left(\left(\ln \left(\sum_{m \in M} \exp \left(\beta P_o^1 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) - E[\beta P_o^1 \mathbf{1}[m = outdoor] | P_o^1] \right) \right. \\
& \left. - \left(\ln \left(\sum_{m \in M} \exp \left(\beta P_o^0 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) - E[\beta P_o^0 \mathbf{1}[m = outdoor] | P_o^0] \right) \right) \\
=& \frac{1}{\gamma_2} \cdot \\
& \left(\left(\ln \left(\sum_{m \in M} \exp \left(\beta P_o^1 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) - \beta P_o^1 E[\mathbf{1}[m = outdoor] | P_o^1] \right) \right. \\
& \left. - \left(\ln \left(\sum_{m \in M} \exp \left(\beta P_o^0 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) - \beta P_o^0 E[\mathbf{1}[m = outdoor] | P_o^0] \right) \right) \\
=& \frac{1}{\gamma_2} \cdot \\
& \left(\ln \left(\sum_{m \in M} \exp \left(\beta P_o^1 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) \right. \\
& - \ln \left(\sum_{m \in M} \exp \left(\beta P_o^0 \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right) \right) \\
& \left. - \beta \left(P_o^1 \Pr(m = outdoor | P_o^1) - P_o^0 \Pr(m = outdoor | P_o^0) \right) \right) \\
=& E[CV] - \frac{1}{\gamma_2} \cdot \beta \left(P_o^1 \Pr(m = outdoor | P_o^1) - P_o^0 \Pr(m = outdoor | P_o^0) \right) \\
=& E[CV] - \frac{\beta}{\gamma_2} \cdot \left(P_o^1 \Pr(m = outdoor | P_o^1) - P_o^0 \Pr(m = outdoor | P_o^0) \right),
\end{aligned} \tag{17}
$$

where $E[\mathbf{1}[m = outdoor] | P_o] = \Pr(m = outdoor | P_o)$ is the outdoor mode choice probability when outdoor pollution level is P_o, and is given by:

$$
\Pr(m = outdoor | P_o) = \frac{\exp \left(\beta P_o \mathbf{1}[m = outdoor] - \gamma_1 t_m - \gamma_2 c_m \right)}{\sum_{\tilde{m} \in M} \exp \left(\beta P_o \mathbf{1}[\tilde{m} = outdoor] - \gamma_1 t_{\tilde{m}} - \gamma_2 c_{\tilde{m}} \right)} \tag{18}
$$

Thus, $E[CV]$ can be written as the following function of expected avoidance cost $E[CA]$ and

WTP:

$$
\begin{aligned}
E[CV] =& E[CA] + \frac{\beta}{\gamma_2} \cdot \left(P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0) \right) \\
=& E[CA] + WTP \cdot \frac{(1-\alpha)\bar{t}_o}{((1-\alpha)\bar{t}_o + \alpha T) \left(P_o^1 - P_o^0 \right)} \cdot \left(P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0) \right) \\
=& E[CA] + WTP \cdot \frac{(1-\alpha)\bar{t}_o}{((1-\alpha)\bar{t}_o + \alpha T)} \cdot \frac{P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0)}{(P_o^1 - P_o^0)} \\
=& E[CA] + WTP \cdot \underbrace{\frac{(1-\alpha)\bar{t}_o}{((1-\alpha)\bar{t}_o + \alpha T)}}_{\in[0,1]} \cdot \underbrace{\frac{\overbrace{P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0)}^{}_{\le 0}}{\underbrace{P_o^1 - P_o^0}_{\le 0}}}_{\in[0,1]}
\end{aligned} \tag{19}
$$

where we assume that the likelihood of an outdoor mode choice increases when the air quality improves from a dirty pollution level P^0 to a cleaner level P^1 (i.e., $\frac{\Pr(m=outdoor|P_o^0)}{\Pr(m=outdoor|P_o^1)} \le 1$) so that:

$$
\begin{aligned}
\frac{P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0)}{P_o^1 - P_o^0} =& \frac{\Pr(m = outdoor|P_o^1) \left(P_o^1 - P_o^0 \frac{\Pr(m=outdoor|P_o^0)}{\Pr(m=outdoor|P_o^1)} \right)}{P_o^1 - P_o^0} \\
\le& \frac{\Pr(m = outdoor|P_o^1) \left(P_o^1 - P_o^0 \right)}{P_o^1 - P_o^0} \\
=& \Pr(m = outdoor|P_o^1) \in [0,1].
\end{aligned} \tag{20}
$$

The difference between $E[CV]$ and WTP is given by:

$$
\begin{aligned}
E[CV] - WTP =& E[CA] + WTP \cdot \left(\frac{(1-\alpha)\bar{t}_o}{((1-\alpha)\bar{t}_o + \alpha T)} \cdot \frac{P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0)}{(P_o^1 - P_o^0)} - 1 \right) \\
=& E[CA] + WTP \cdot \underbrace{\left(\underbrace{\frac{(1-\alpha)\bar{t}_o}{((1-\alpha)\bar{t}_o + \alpha T)}}_{\in[0,1]} \cdot \underbrace{\frac{\underbrace{P_o^1 \Pr(m = outdoor|P_o^1) - P_o^0 \Pr(m = outdoor|P_o^0)}_{\le 0}}{\underbrace{P_o^1 - P_o^0}_{\le 0}}}_{\in[0,1]} - 1 \right)}_{\le 0}
\end{aligned} \tag{21}
$$

Thus, $E[CV]$ can be less than, equal to, or greater than WTP. Whether $E[CV]$ is less than, equal to, or greater than WTP is therefore an empirical question.

3.3 Econometrics

Our empirical analysis uses a high dimensional instrumental variable model. While Belloni et al. (2012) propose an estimator to identify the instruments under a high dimensional setting, their underlying data generation process (DGP) is different from ours. Fortunately, we can still use the least absolute shrinkage and selection operator (LASSO) to identify our DGP.

In particular, Belloni et al. (2012) develop results for the use of LASSO and post-LASSO methods to form first-stage predictions and estimate optimal instruments in linear instrumental variables (IV) models with many instruments when the first stage is approximately sparse – that is, when there exists a relatively small set of important instruments whose identities are unknown that well-approximate the conditional expectation of the endogenous variables given the instruments (Bunea et al., 2007). We extend the high dimensional setting of Belloni et al. (2012) by relaxing the sparse IV assumption to a weak IV assumption. From an empirical point of view, we show that the economic intuition can effectively assist the theoretical framework to obtain higher estimation efficiency.

The motivation for our DGP is from the weak IV scenario by Bound et al. (1995):

$$x_i = \beta_1 z_{i,1} + \cdots + \beta_s z_{i,s} + \frac{1}{\sqrt{np}}\gamma_1 t_{i,1} + \cdots + \frac{1}{\sqrt{np}}\gamma_p t_{i,p} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{22}$$

where z_i and t_i are two vectors of potential instruments. ε_i is the error term. s, p, and n are not fixed. $s < n$ and $p > n$ in this case. Following the usual definition, only z_i is the set of strong instruments and we cannot get consistent results if we include all IVs z_i and t_i in a 2SLS regression. Moreover, a classical OLS regression for the equation (22) is not well defined since the number of variables on the right hand side is larger than the number of observations. Therefore, we require novel estimation procedure to deal with the endogenous issue in our setup.

Denote our model as $\theta = (\beta_1, \ldots, \beta_s, \frac{1}{\sqrt{np}}\gamma_1, \ldots, \frac{1}{\sqrt{np}}\gamma_p)'$, $x = (x_1, \ldots, x_n)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$. $x = A\theta + \epsilon$, where $A_{ij} \overset{iid}{\sim} N(0,1)$ and $\frac{1}{n}||\varepsilon||_1 = \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| < v$.[5] The following theorem validates the LASSO when $\sqrt{p\log(s+p)} << n$.

[5] $||\cdot||_1$ represents the l_1 norm: $||\varepsilon||_1 = \sum_{i=1}^{n}|\varepsilon_i|$.

Theorem Let $\hat{\theta}$ solve $min\ ||\theta||_1$ s.t. $\frac{1}{n}||A\theta - x||_2^2 \leq v^2$.[6]

Then, $E||\hat{\theta} - \theta||_2 \leq \sqrt{8\pi}\left[\left(\sqrt{s}||\beta||_2 + \frac{\sqrt{p}|\bar{\gamma}|}{\sqrt{n}}\right)\frac{\sqrt{\log(s+p)}}{n} + v\right]$.

See the proof in the appendix.

Our model is under a common weak IV scenario. The literature assumes the IVs to be sparse, meaning that most of the potential IVs are independent from the endogenous variable. However, our setting relaxes the independence to weak correlations. The price of the relaxation is that the required number of observations should satisfy $\sqrt{p\log(s+p)} << n$ while under original sparsity $s\log(s+p) << n$. In other words, when $p > s^2\log(s+p)$, researchers should obtain more observations under a weak IV case than the literature suggests.

4 Data and Reduced-form Evidence

Our data set includes three parts: individual travel mode choice, hourly air pollution, and instrumental variables. The travel data are from two rounds of the Beijing Household Travel Survey in 2010 and 2014. The air pollution data are from the U.S Embassy in Beijing. The potential instruments are weather conditions in China from the National Oceanic and Atmospheric Administration (NOAA).

4.1 Beijing Household Travel Survey

Our travel mode data set is from the Beijing Household Travel Survey (BHTS), a confidential data set on travel surveys conducted in 2010 and 2014 by the Beijing Municipal Commission of Transport (BTRC). This detailed cross-sectional data set with a million observations includes the characteristics of individual household members, including their occupations, ages, and education; and the characteristics of each trip taken during a designated 24-hour period, including which of six travel modes, the distance, the time, and the districts.

BTRC randomly selected 642 out of 1,191 Traffic Analysis Zones (TAZs) in year 2010, and 667 out of 2,050 TAZs in year 2014, from the entire city. TAZs are geocoded areas defined by

[6] $||\cdot||_1$ represents the l_1 norm: $||\theta||_1 = \sum_{j=1}^{s+p}|\theta_j|$. $||\cdot||_2$ represents the l_2 norm (or Euclidean norm): $||x||_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$. The minimization problem is a special case of LASSO. The equivalence comes from the duality of the optimization problem.

the BTRC for traffic analysis. Each of the administrative districts in Beijing has 16 to 238 TAZs, based on the size of the area and the population of the district. TAZs are smaller in districts with higher population densities. The average TAZ is about 1.5 square kilometers. In the inner eight districts, on which the sampling focuses, TAZs range from 0.21 to 16 square kilometers. Each TAZ are randomly selected for in-person interviews to collect data on trips taken during a designated 24-hour period. The survey covered 116,142 individuals from 46,900 households in year 2010 and 101,827 individuals from 40,004 household in year 2014. We have 253,648 trips for year 2010 and 205,148 trips for year 2014 in total. We utilize the work-commute travel since for these compulsory trips, staying at home is not an option. When people stay at home, we do not acquire their travel information and the observed preference cannot reflect the benefit of exposure to the air pollution from reducing the cost.

Table 1 and Table 2 provide summary statistics for the travel and personal information for the two years of data. The tables report respondent and trip characteristics of all work commuting trips in these two years. In Table 1, the individual income increases from from 73,141.48 RMB in year 2010 to 101,939.57 RMB in year 2014, a 40% increase. China has the world's fastest-growing major economy, with growth rates averaging 10%. The difference in income between year 2010 and 2014 satisfies the growth rate number, implying the random selection of the household survey. The increase in the number of cars owned by each individual also reflects the wealth increment in Beijing. The gender, age, and schooling are quite similar in these two years.

Table 2 presents the trip information. First, the travel time is lower in year 2014. The expansion of the transportation system can explain these numbers. Beijing's rapid subway expansion from 2008 to 2014 led to an increase in aggregate welfare with modest congestion reduction. Additionally, more people used a car in 2014, compared to other modes. No number contradicts our intuition given that Beijing has become richer over the two years. In Figure 2, we check the variation of the mode shares for the two years. There are 13 travel modes in the data and we aggregate them into 6 modes: $\{Walking, Car, Subway, Bus, Taxi, Bicycle\}$.

4.2 Air Pollution and Weather Controls

For air pollution, we use hourly data on $PM_{2.5}$ from the U.S. Embassy in Beijing. There are 35 observation stations from Chinese government for year 2014, but for year 2010 we only have one U.S. Embassy station (indicated in green in Figure 3). To check the different travel patterns for year 2010 and year 2014, with respect to air pollution variation under the same scale, we use the U.S. Embassy station data. The U.S. Embassy is located near the East Third Ring Road in Beijing. Figure 4 presents a data correlation diagram for the 8 monitoring stations near or inside of the third ring roads in Beijing. The Pearson correlation coefficients for the $PM_{2.5}$ data detected by these stations are above 0.94 and significant at the level of 0.01. Therefore, we believe that the data from the U.S. Embassy monitoring station is representative of the air conditions in Beijing.

Our data for our weather controls, including precipitation, temperature, wind speed and wind direction in Beijing, are from the National Oceanic and Atmospheric Administration (NOAA). We include the weather conditions to reduce the potential bias in the estimation. In Figure 1, we check the correlation between the air pollution and the outdoor travel share, to provide descriptive evidence for our further analysis. The negative correlation between air pollution and outdoor travel in the year 2014 is evidence for the existence of avoidance behavior in 2014. In contrast, the positive correlation in the year 2010 suggests that in 2010 the WTP, which depends on information, was based on the expected cost of fog rather than the actual damage from smog. In 2010, people might not have been willing to pay anything for clean air, despite the adverse effects of air pollution on their health. This suggests that awareness and behavior changed after the media coverage about air pollution. To investigate the potential reason of the phenomenon in year 2010, in Figure 5, we divide the temperature over high and low using 20 Celsius degree as a threshold. In 2010, outdoor travel declines with air pollution when temperatures are high, but increases with air pollution when temperatures are low.

4.3 Instruments

Air pollution is endogenous to travel decisions for two main reasons. First, air pollution is endogenous because of omitted variable bias. For example, an unobserved variable subsumed by the error term is the number of days a person is staying in the home, which is influenced by air pollution. The second reason that air pollution is endogeneous is due to reverse causality. For example, travel modes affect traffic congestion, which in turn affects air pollution levels. We can partially resolve these issues of unobserved variables and simultaneity by including day-of-the-week and hour-of-the-day fixed effects. A third concern is measurement error, which biases the estimation downward.

We therefore instrument for air pollution to address its endogeneity. The instruments should be uncorrelated with the errors, including unobserved economic activities. Instruments should additionally be factors that contribute to (instead of merely be correlated with) $PM_{2.5}$ in Beijing.

A common source of instruments are wind speeds and wind directions, for winds that blow from locations other than Beijing with high pollution. We control for wind directions and wind speeds in Beijing,[7] since wind in Beijing may directly affect travel decisions in Beijing, but we assume that wind at and from other locations would not affect travel mode decisions except through their effect on air pollution. Figure 6 shows the pollution heat and wind direction in China for a random day in year 2021. The wind directions connect the pollution areas and reflect that air pollution transmission depends highly on the air pollution.

Researchers also commonly use factory production near Beijing as instruments for the air pollution in Beijing. As economic activity in Beijing and factory production near Beijing are likely correlated, this potential set of instruments does not satisfy the exclusion restriction. Therefore, factory production near Beijing is an invalid instrument in our case. Our set of potential instruments therefore are winds blowing from locations other than Beijing that may have high pollution that might blow into Beijing.

NOAA has 235 monitoring stations in China. We divide the wind directions into eight 45-degree intervals: North, Northeast, East, Southeast, South, Southwest, West, and Northwest. Since we only have around 1,000 observations of hourly pollution in year 2010 or year 2014 but 2,035 possible candidate instruments, we run the risk of overfitting in the first stage. It is also inefficient to select the polluted city manually. Hence, we adopt machine learning and LASSO regressions to select our

[7] The amount of transported pollution is large (Zhang et al., 2017).

instruments for air pollution.[8]

5 Empirical Strategies

Equation (5) elicits the underlying utility function of a travel mode choice,

$$u_{imt} = \beta_1 \mathbf{1}[m = indoor] X_t + \beta_2 \mathbf{1}[m = outdoor] X_t \\ + \gamma_1 time_{imt} - \gamma_2 cost_{imt} + \mathbf{z_{it}}' \eta_m + \alpha_m + \xi_{mt} + \varepsilon_{itm},$$

where i represents the individual, t is the hour and m denotes the travel mode. We also include an alternative-time preference parameter, ξ_{tm}, a constant specific to alternative m at time t. It shows economic activities, which reflects the endogenous error term. Weather conditions and individual characteristics are in the vector z_i. Our X_i is the $PM_{2.5}$, $time_{imt}$ is the commute time of mode m, and $cost_{imt}$ is the out-of-pocket cost of mode m. These three variables form the core of our trade-off analysis. Finally, ε satisfies the type-I extreme value distribution.

The utility function (23) is quasilinear as equation (5). However, we relax the assumption that people know the air pollution difference between indoor and outdoor activities, if we interpret β_1 and β_2 as a combination of the cognition about indoor or outdoor exposure and the marginal effect of the air pollution. Alternatively, the model measures the same preference if the cognition is the same as the true exposure. Hence, our utility model for estimation is more flexible than the theoretical foundation.

We also make proper assumptions to increase the efficiency of the estimation. In each trip, though the time of exposure from cycling is lower then walking for the same distance, we assume the disutility of air pollution for these two alternatives are the same since exercise accelerates breathing. In each trip, the disutility of air pollution for the bus, car, taxi are also identical. While some of the alternatives might be faster than the others, when respondents arrive the working location, they stay in the building for the extra time, and experience the same indoor air pollution exposure as the slower counterfactual world.

[8] Belloni et al. (2011) prove the efficiency of the LASSO method for a sparse Gaussian IV model. In Section 3 we show the LASSO is still consistent under the weak IV scenario.

We rely on equation (23) to do reduced form regression as well as logit regression. To deal with their endogeneity,[9] we use machine learning and LASSO regressions to select strong instruments for air pollution.

5.1 Reduced-Form Regression

For the empirical analysis, we rely on a mean transformation of the cost and time over individuals on hours, to utilize the linear model's efficiency and convenience. The share regression therefore is easily available,

$$\begin{aligned}\ln S_{mt} - \ln S_{1t} =& \beta \mathbf{1}[m = outdoor]X_t + \\ & \gamma_1(\overline{time}_{mt} - \overline{time}_{1t}) + \gamma_2(\overline{cost}_{mt} - \overline{cost}_{1t}) + \overline{\mathbf{z_t}}'(\eta_m - \eta_1) + \alpha_m + \xi_{mt}\end{aligned}$$

where $\beta = \beta_2 - \beta_1$.[10] We can use a classical IV approach to deal with this endogeneity from ξ_{tm}, satisfying the inclusion restriction and exclusion restriction. The IV should be uncorrelated with ξ_{mt} but correlated with the air pollution. Meanwhile, unlike the discrete choice model in a consumption market, the average cost of the the travel mode is invariant over different markets t. Beijing used a one-price mechanism in year 2010 and 2014 for the public transportation. The parking fees of vehicles and the costs of taxi are also equivalent at different hours on average. Moreover, walking and bicycle are zero cost. Overall, for all modes, the price is unlikely to change over time. This is problematic since we cannot include the mode specific fixed effect α_m due to collinearity. To overcome the difficulty, in the reduced form regression, we use the logit model and the control function method.

5.2 Logit Model and Control Function

The control function method extends a logit regression (Heckman and Robb Jr, 1985; Petrin and Train, 2010), and discusses the endogenous errors at the individual level. We assume there is only

[9] In Section 4, we discussed three potential endogenous problems. Economic activities can represent all.

[10] The proof is in the appendix.

one market where people make decisions on the travel mode,

$$\begin{aligned} u_{im} &= \beta \mathbf{1}[m = outdoor]X_i + \gamma_1 time_{im} + \gamma_2 cost_{im} + \mathbf{z_i}'\eta_m + \alpha_m + \varepsilon_{im} \\ X_i &= \gamma_0 + \gamma_1 f(weather\ condition\ far\ away) + \mu_i \\ \mu_{im} &= \mathbf{1}[m = outdoor]\mu_i \end{aligned} \tag{23}$$

The endogenous error is ε_{im} and it is correlated with X_{im} through μ_{im}. To handle the potential bias, we then decompose the ε_{im} into its mean conditional on μ_{im} and deviations around this mean: $\varepsilon_{im} = E(\varepsilon_{im}|\mu_{im}) + \tilde{\varepsilon}_{im}$. The residual $\tilde{\varepsilon}_{im}$ is therefore uncorrelated with the μ_{im}.

We regress the air pollution on the weather conditions far away to get μ_{im}. Then decompose the utility function as,

$$u_{im} = \beta \mathbf{1}[m = outdoor]X_i + \gamma_1 time_{im} + \gamma_2 cost_{im} + \mathbf{z_i}'\eta_m + \alpha_m + \lambda\mu_{im} + \tilde{\varepsilon}_{im} \tag{24}$$

Due to the exogeneity of the $\tilde{\varepsilon}_{im}$, we can rely on a logit model to get the result.

6 Empirical Results

6.1 Instrumental Variable Selection

We run two LASSO regressions for the first stage to select among 2,035 wind conditions as strong instruments using 2,069 observations,[11] controlling for 18 hourly, 2 monthly, and 7 weekday fixed effects, with other weather conditions in Beijing. The algorithms pick out more than 200 instruments in total for two separate LASSO regressions for the year 2010 and 2014.

The number of selected instruments is too large that it is impossible to filter out those that contradict facts manually. Therefore, we check the overlap locations – Tangshan, Shenyang and Zhangjiakou – in year 2010 and year 2014, by running OLS as a first-stage regression of hourly air pollution on their wind directions with the same set of controls. The coefficients and intuition support the southwest wind in Tangshan as the optimal instrument for our empirical analysis. In

[11] The survey is in October, September, November year 2010 or 2014. We drop hours in the midnight. The pollution at theses hours are not contributors to the estimation.

Table 3, the signs of the coefficients are similar for three potential cities in two different years. Therefore, there is no marked change in air pollution transition structure from the year 2010 to the year 2014.

In Figure 7, the geographic locations with the economic intuition of three cities complement the instrument selection. To begin with, the southwest wind in Shenyang, southwest wind in Zhangjiakou, southeast wind in Tangshan, Northeast wind in Tangshan are four potential instruments because they are strong and share similar coefficients in different years. The southwest wind in Shenyang has -75.97 and -63.67 respectively for the years 2010 and 2014. The negative signs are reasonable according to the map considering Shenyang is a highly polluted city. However, a northeast wind pointing the city to Beijing should be a contributor to the the air pollution in Beijing. The -33.00 and -20.89 reveal that the wind directions in Shenyang are less convincing as instruments. Secondly, the southwest wind in Zhangjiakou also encounter the same obstacle because its northwest wind has a -20.61 coefficient in 2014. Finally, Tangshan satisfies the matching of the predicted signs from relative position and the estimated coefficients of the wind directions. Though the coefficient of northwest wind in Tangshan change from 2010 to 2014, the results are still reasonable since the northwest wind could either be a contributor or a eliminator to the air pollution in Beijing, depending on whether a relative north or a west wind dominates in the period. Hence, we use the southeast wind in Tangshan as the instrument, with a first-stage F-statistic of 54.91.

6.2 Reduced-From Estimation Results

Baseline Analysis The reduced-form estimation results uncover the distinguished travel patterns for the years 2010 and 2014 when the air is bad at different hours. We use the one way commuting trip and investigate the compulsory trade-off of the indoor and outdoor travel mode when the air is bad. In Table 4 from column 1 to 2, as Figure 1, without the instrument and alternative specific fixed effect, 100 $\mu g/m^3$ $PM_{2.5}$ exposure increases the the utility by 0.0343 in year 2010 but decreases 0.0608 in year 2014. In column 3 and 4, with the fixed effect,[12] the signs sustain

[12] The average cost of the the travel mode is invariant over different markets t. Beijing used an one price mechanism in year 2010 and 2014 for the public transportation. The parking fees of vehicles and the costs of taxi were also equivalent at different hours on average. Moreover, walking and bicycle are zero cost. Overall, for all modes, the price is unlikely to change over time.

as 0.0233 and -0.0948 for year 2010 and year 2014 respectively. Temperature control derives the insignificance for coefficients in column 1 and 3.

We use the southeast wind in Tangshan as the instrument with a F statistics 54.91. Then, -0.382 for 100 units of air pollution exposure, significant at 1% level, in year 2014 and -0.1811 for 100 units of air pollution exposure, significant at 10% level emphasize the pollution reaction after the media coverage to the pollution in year 2013. The numbers for year 2010 and year 2014 in column 7 and 8 are not significant but keep the directions of the coefficients.

With the negative coefficients of the cost and the negative coefficients of the time for all columns, the value of time (VOT) delivers a convincing increase over the years. We use columns 7 and 8 to calculate the parameters by $\frac{distuility\ of\ time}{disutility\ of\ cost}$. With a wealthier society, people value time more, from 4.78 RMB per hour to 9.05 RMB per hour. On averge, over yearly increase rate 20% for the parameter in Beijing is a large amount.

Preference Heterogeneity Intuitively, different age groups and income groups have heterogeneous preference to air pollution. We run the same set of regression but for different subgroups. As the pattern in Table 4, in Table 5 and Table 6, 2014 has a higher disutility for the same level of air pollution exposure than year 2010. Additionally, instruments promote the significance. These results show the robustness of the estimation.

Table 5 indicates responses to air pollution for young and old, with 50 as the threshold. With instrument validating the analysis, when the cost coefficient is standarized, the disutility of 100 $\mu g/m^3$ air pollution exposure for the old 6.13 is higher than the disutility of 6.08 for the young, implying respondents vulnerable to air pollution are reluctant to expose to the bad air. As for the VOT, 9.18 RMB per hour from the young is lower than 10.32 RMB per hour from the old for year 2014, suggesting the old value time more than the young.

Table 6 describes avoidance behavior adjustments over time for different income groups, with annual income 100,000 RMB as the cutoff. Column 7 and 8 are the cores of our analysis as before. The disutiliy of per 100 units of air pollution exposure is -0.4787 for the high income group, which is a 36% higher magnitude than the poor group. Moreover, the rich group has the negative utility for the air pollution exposure even in year 2010, arguing that the rich people are among the first set of people obtaining the harmful information about the air pollution. The VOT again works as a robust test for the empirical procedure. The value of time is higher for the high income group compared to the values from the low income group.

6.3 Logit Regression and Control Function

To avoid the potential problem of the share model,[13] we also use a logit regression and control function method parallel to ordinary least squares (OLS) and the two-stage least squares (2SLS). In Table 7, the coefficients match the results in the share regression section's estimates. The air pollution exposure per $100\mu g/m^3$ has an insignificant coefficient in year 2010 while in year 2014 the air pollution exposure per $100\mu g/m^3$ has -0.0454 coefficient significant at 1% level. The value of time is 6.70 RMB per hour in year 2010 and 12.47 RMB per hour.

After dividing the sample into different bins for ages or incomes, the heterogeneity disutility of one unit exposure to $PM_{2.5}$ results for year 2014 with IV are in Figures 8a and 8b. The cost coefficients are normalized to negative 1. Figure 8a argues that extreme young, acquiring pollution information easily, and extreme old people, vulnerable to the pollution, are more likely to avoid the pollution. While the monetary values for air pollution in different income groups do not reflect a intuitive trend, the large standard errors cannot reject the null hypothesis that high income people have a higher disutility for the same amount of $PM_{2.5}$.[14]

Though we calculate the disutility of an average trip time air pollution exposure and then the average WTP for the clean air, a longer time exposure to the air pollution will be more harmful than a shorter one. We divide the exposure time for trips into different time bins and show in Figure 8c the disutility increases as the exposure time increases linearly.

[13]There is collinearity between hourly average cost and alternative specific fixed effect.

[14]Currently, we use the reduced-form regression to discuss the income WTP difference due to the significance, while logit regressions are more reasonable.

7 Discussion

7.1 Willingness-to-Pay

To interpret the coefficients in Section 6 for welfare analysis, we base our willingness to pay for the clean air on Table 7 column 4, because people started to realize the harmfulness of smog after 2013. Equation (5) motivates our empirical framework (23), where $h'(\bar{P}T) = \frac{\beta}{(1-I/O)\gamma_2 \bar{t}}$ and $\bar{t}$ is the average trip time. The disutilty of an hour exposure to the air pollution is therefore 0.014 RMB per $\mu g/m^3$ $PM_{2.5}$, which equals to 0.00223 dollars, according to the exchange rate in 2014, when we set the infiltration factor as 0.56. Since we have 8,760 hours in a year, the lower bound[15] of long-term individual WTP is therefore around 19.53 dollars per year to remove 1 $\mu g/m^3$ $PM_{2.5}$. The number is similar to previous literature, but still have a different interpretation, since the rationale for our optimization framework on air pollution exposure is distinctive.

7.2 Air Pollution Exposure

Representative Consumer The main advantage of our method is that our estimates represent all citizens. In our analysis, the sample is random over the whole city. Moreover, almost every people in the city make a trade-off between the pollution and the travel cost.

By contrast, people who purchase air cleaning equipment or can afford to live at better locations are those who pay attention to their health, so their WTP for clean air is likely higher than that of the general population. Furthermore, WTP estimates based on air cleaning equipment purchasing behavior or residential housing location choice measure the WTP at equilibrium prices in the equipment or housing market, thus we have no information about the WTP of other people whose WTP is lower than the equilibrium price on the demand curve. Therefore, our estimation is more unbiased.

Rational Expectations When estimating the long term WTP for the clean air, researchers assume people can realize the total air pollution reduction induced by their behaviors. However, it is hard for people to predict the air pollution in the future and capture sufficient information

[15] A first order approximation might be a lower bound for the air pollution if the well being function h is convex. See Section 3 for the details.

about the clean neighborhood. Additionally, when purchasing a house, people may care more about intangible assets, like the neighborhood education, or other facilities that are hard to measure and correlated with the air pollution exposure. Therefore, the estimates might be biased to a greater magnitude and highly depends on the data set collection.

In contrast, in our paper, we estimate the short term effect: people can directly observe the air pollution through through the transparency or smell of the air. After waking up, they can look outside and decide the travel mode they will use. It is similar to people notice it is raining outside so they take an umbrella. Moreover, people can also check the air quality index through their phones easily after 2013.

Exclusive Avoidance Even if people perceived the air pollution change correctly, the change might not be the same as the data record. All revealed preference methods rely on some exclusive avoidance behavior. In the previous literature, indoor behaviors are the only source to bring down air pollution. When purchasing a house or buying a filter, people are assumed to stay in their home as long as possible. This overestimates the reduction of the purchasing behavior.

We also make a similar assumption in our estimation. Though people might start to wear masks in 2014 to protect themselves, which might lead to lower air pollution exposure than our estimation for the outdoor activities, the masks are not fine enough to protect people from $PM_{2.5}$ in 2014. There are many media reports in 2014 covered the uselessness of the masks and mentioned only some special version could protect people. The production of those special masks was very low. Additionally, similar to the status quo during the Covid-19 pandemic, it takes time for people to get used to the masks. In 2014, people were reluctant to wear a mask, because either most masks in the market were not protective, or they were uncomfortable to wear.

8 Conclusion

In this paper, we develop a model to estimate the short-term WTP to be 0.00223 dollars per hour to avoid 1 $\mu g/m^3$ $PM_{2.5}$, which translates to a lower bound of the long-term WTP at 19.53 dollars per year to avoid 1 $\mu g/m^3$ $PM_{2.5}$. Our estimation relaxes a few assumptions in the previous literature and provides a new direction to quantify the WTP. In our compulsory work trip scenario, most citizens have to choose between an indoor and outdoor mode to work, even if the air pollution can be directly observed. Accurate information is readily available on phone applications for those decision-makers, so we can assume that they reckon with air pollution levels when making decisions.

The WTP is higher in the elder and richer groups, and is higher for longer travels. A longer potential exposure to the air pollution prevents people from walking and cycling. People older than 55 years old are vulnerable to pollution and thus more likely to avoid pollution. Richer people, who value their health more, are more likely to avoid the pollution as well.

We also show the behavioral change hinges on the knowledge about the dirty air. The avoidance behavior not only depends on whether people can observe the air pollution or not, but is also determined by the knowledge toward the harmfulness of it. The air pollution may have long-lasting health effect but people would not use any strategy to avoid it before building the connection between the pollution and their sickness. Therefore, in 2010, before the realization of the harmful air pollution, people did not show any avoidance evidence.

Finally, our empirical strategy supplements the growing literature on machine learning in economics. Our proof validates that a LASSO regression is still useful under the weak IV scenario, though the model require more numbers of observations. As an application of choosing wind directions for the endogenous air pollution, we select the southwest wind in Tangshan, the wind direction pointing a high polluted city to Beijing. The empirical procedure after the algorithm exhibits the importance of economic intuition besides a tedious theoretical framework for a high-dimensional estimation.

Since we focus on the compulsory work trips, our estimates of the pure willingness-to-pay (WTP), expected compensating variation (CV), and expected avoidance cost are daily measures based on commuting. As discussed in discussed in Section 7, focusing on compulsory work trips enables us relax the three common assumptions made in the previous WTP literature: representative consumer, rational expectations, and exclusive avoidance. The presence of avoidance behavior in compulsory work trips suggests that avoidance behavior in non-compulsory activities may take place and possibly be at least as large.

Bibliography

D. Bäck, N. V. Kuminoff, E. Van Buren, and S. Van Buren. National evidence on air pollution avoidance behavior. *Unpublished paper, January*, 2013.

P. J. Barwick, S. Li, L. Lin, and E. Zou. From fog to smog: The value of pollution information. NBER Working Paper No. 26541, 2020.

P. J. Barwick, S. Li, D. Rao, and N. Zahur. The morbidity cost of air pollution: Evidence from consumer spending in China. *Available at SSRN 2999068*, 2021.

P. Bayer, N. Keohane, and C. Timmins. Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management*, 58(1):1–14, 2009.

A. Belloni, V. Chernozhukov, and L. Wang. Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806, 2011.

A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

N. E. Bockstael and K. E. McConnell. *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models*, volume 7. Springer Science & Business Media, 2007.

J. Bound, D. A. Jaeger, and R. M. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995.

F. Bunea, A. Tsybakov, and M. Wegkamp. Sparsity oracle inequalities for the lasso. *Electronic Journal of Statistics*, 1:169–194, 2007.

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

C. Chen and B. Zhao. Review of relationship between indoor and outdoor particles: I/o ratio, infiltration factor and penetration factor. *Atmospheric Environment*, 45(2):275–288, 2011.

S. Chen, C. Guo, and X. Huang. Air pollution, student health, and school absences: Evidence from China. *Journal of Environmental Economics and Management*, 92:465–497, 2018.

J. Chu, H. Liu, and A. Salvo. Air pollution as a determinant of food delivery and related plastic waste. *Nature Human Behaviour*, 5(2):212–220, 2021.

M. A. Cole, R. J. Elliott, and B. Liu. The impact of the Wuhan Covid-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach. *Environmental and Resource Economics*, 76(4):553–580, 2020.

J. A. Cook and C.-Y. C. Lin Lawell. Wind turbine shutdowns and upgrades in denmark: Timing decisions and the impact of government policy. *Energy Journal*, 41(3):81–118, 2020.

J. Currie and R. Walker. Traffic congestion and infant health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1):65–90, 2011.

T. Deryugina, G. Heutel, N. H. Miller, D. Molitor, and J. Reif. The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12): 4178–4219, 2019.

O. Deschenes, H. Wang, S. Wang, and P. Zhang. The effect of air pollution on body weight and obesity: Evidence from China. *Journal of Development Economics*, 145:102461, 2020.

D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via $l1$ minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

C. P. Green, J. S. Heywood, and M. N. Paniagua. Did the London congestion charge reduce pollution? *Regional Science and Urban Economics*, 84:103573, 2020.

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

J. J. Heckman and R. Robb Jr. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1-2):239–267, 1985.

K. Ito and S. Zhang. Willingness to pay for clean air: Evidence from air purifier markets in China. *Journal of Political Economy*, 128(5):1627–1672, 2020.

R. Just, D. Hueth, and A. Schmitz. *Applied Welfare Economics and Public Policy*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

R. Just, D. Hueth, and A. Schmitz. *The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation*. Edward Elgar, Northampton, MA, 2004.

S. Kundu and E. A. Stone. Composition and sources of fine particulate matter across urban and rural sites in the Midwestern United States. *Environmental Science: Processes & Impacts*, 16 (6):1360–1370, 2014.

H. Liu and A. Salvo. Severe air pollution and child absences when schools and parents respond. *Journal of Environmental Economics and Management*, 92:300–330, 2018.

J. Lu, B. Li, H. Li, and A. Al-Barakani. Expansion of city scale, traffic modes, traffic congestion, and air pollution. *Cities*, 108:102974, 2021.

D. McFadden. Computing willingness-to-pay in random utility models. In J. Moore, R. Riesman, and J. Melvin, editors, *Trade, Theory and Econometrics: Essays in Honour of John S. Chipman*. Routledge, London, 1999.

G. Michael and J. B. Kelsey. Envirodevonomics: a research agenda for an emerging field. *J. Econ. Lit*, 53(1):5–42, 2015.

V. Milman. Random subspaces of proportional dimension of finite dimensional normed spaces: approach through the isoperimetric inequality. In *Banach Spaces*, pages 106–115. Springer, 1985.

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

M. J. Neidell. Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma. *Journal of Health Economics*, 23(6):1209–1236, 2004.

A. Pajor and N. Tomczak-Jaegermann. Subspaces of small codimension of finite-dimensional Banach spaces. *Proceedings of the American Mathematical Society*, 97(4):637–642, 1986.

A. Petrin and K. Train. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13, 2010.

O. Raaschou-Nielsen, Z. J. Andersen, R. Beelen, E. Samoli, M. Stafoggia, G. Weinmayr, B. Hoffmann, P. Fischer, M. J. Nieuwenhuijsen, B. Brunekreef, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European study of cohorts for air pollution effects (ESCAPE). *The Lancet Oncology*, 14(9):813–822, 2013.

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

A. Salvo. Local pollution as a determinant of residential electricity demand. *Journal of the Association of Environmental and Resource Economists*, 7(5):837–872, 2020.

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

C. P. Tribby, H. J. Miller, Y. Song, and K. R. Smith. Do air quality alerts reduce traffic? an analysis of traffic data from the Salt Lake City metropolitan area, Utah, USA. *Transport Policy*, 30:173–185, 2013.

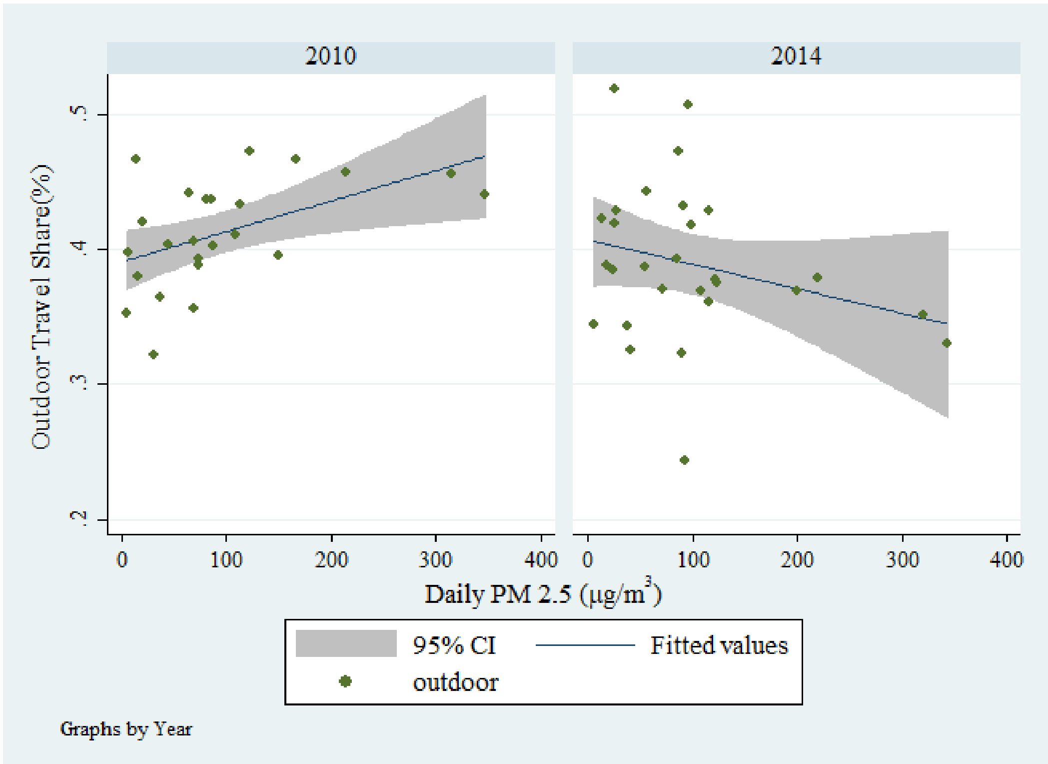
Q. Zhang, X. Jiang, D. Tong, S. J. Davis, H. Zhao, G. Geng, T. Feng, B. Zheng, Z. Lu, D. G. Streets, et al. Transboundary health impacts of transported global air pollution and international trade. *Nature*, 543(7647):705–709, 2017.

J. Zhao and C. L. Kling. Willingness-to-pay, compensating variation, and the cost of commitment. *Economic Inquiry*, 42(3):503–517, 2004.

P. Zhao, S. Li, P. Li, J. Liu, and K. Long. How does air pollution influence cycling behaviour?: Evidence from Beijing. *Transportation Research Part D: Transport and Environment*, 63:826–838, 2018.

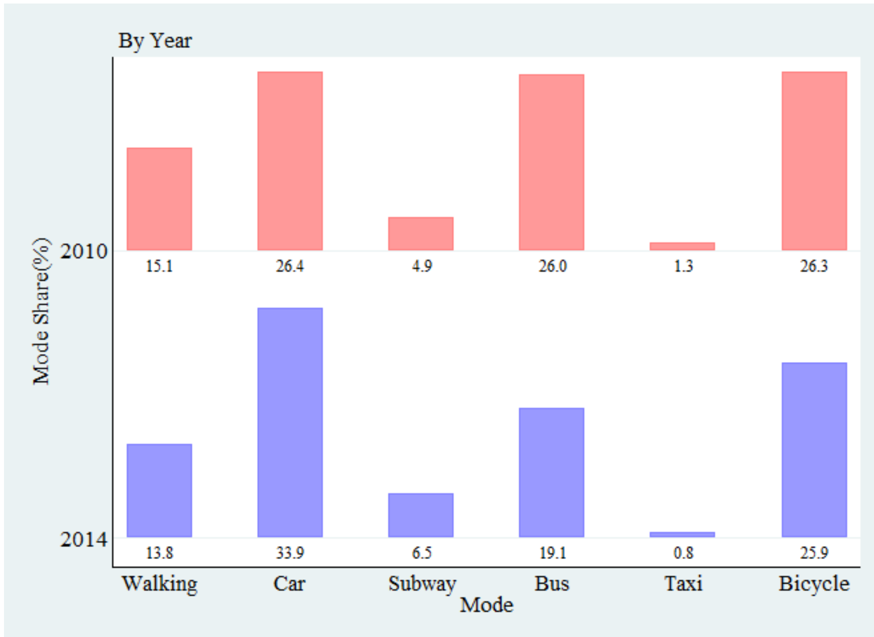
N. Zhong, J. Cao, and Y. Wang. Traffic congestion, ambient air pollution, and health: Evidence from driving restrictions in Beijing. *Journal of the Association of Environmental and Resource Economists*, 4(3):821–856, 2017.

Figure 1: Daily Average Air Pollution and Outdoor Mode Shares



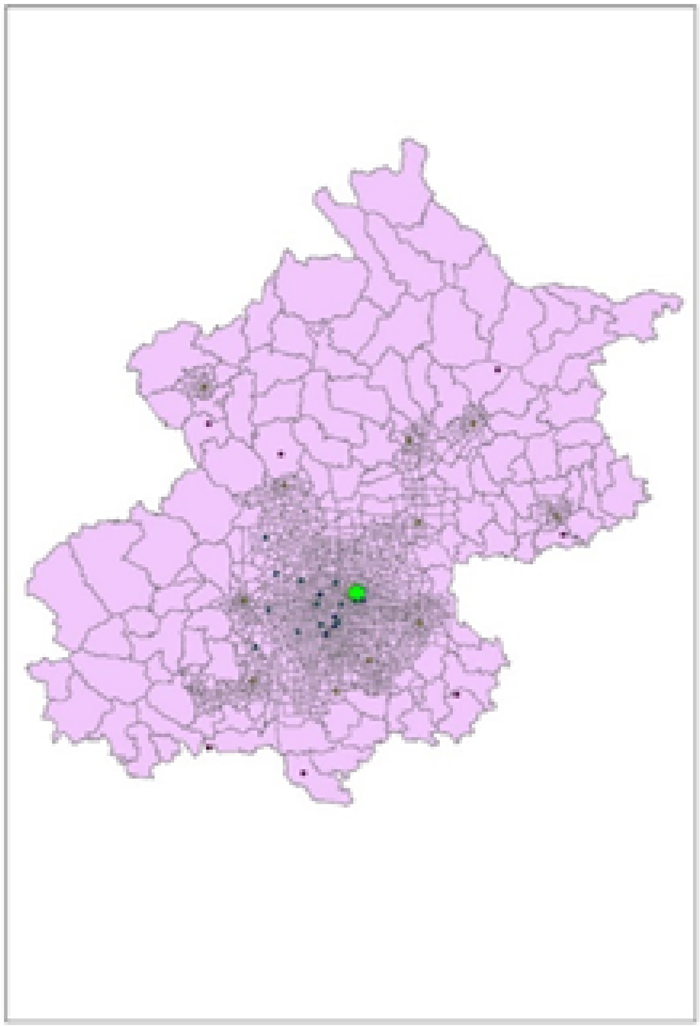
Note: Travel data are from the Beijing Household Travel Survey (BHTS) and air pollution average is from U.S. Embassy in Beijing. Outdoor travel modes are $\{Walking, Bicycle\}$. The full data description is in Section 4. The negative correlation between air pollution and outdoor travel in the year 2014 is evidence for the existence of avoidance behavior in 2014. In contrast, the positive correlation in the year 2010 suggests that in 2010 the WTP, which depends on information, was based on the expected cost of fog rather than the actual damage from smog. In 2010, people might not have been willing to pay anything for clean air, despite the adverse effects of air pollution on their health. With the dissemination of knowledge about air pollution during China's 2013-2014 air quality monitoring and disclosure program, the WTP in 2014 better reflects the true social welfare loss from air pollution.

Figure 2: Travel Mode Share



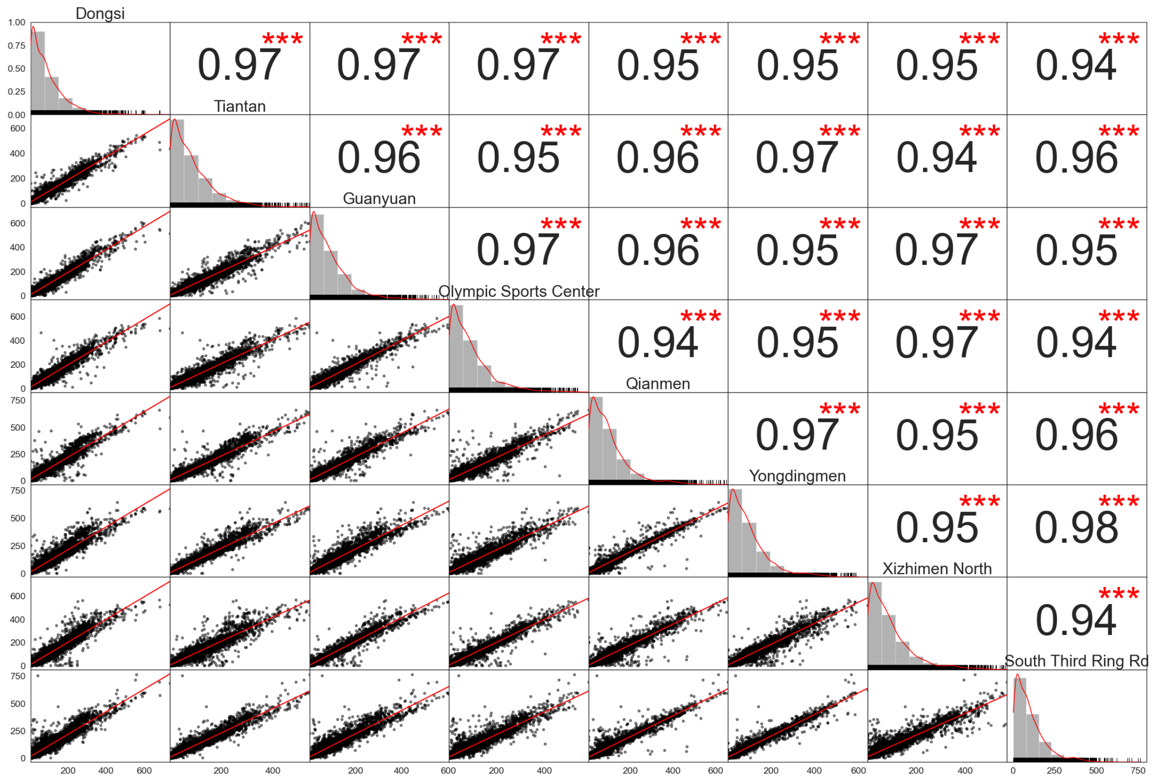
Note: Figure 2 presents the 6 modes shares of work commuting trips in Beijing. Beijing's rapid subway expansion from 2008 to 2014 led to an increase in aggregate welfare with modest congestion reduction. The increase in the number of cars owned by each individual also reflects the wealth increment in Beijing.

Figure 3: U.S. Embassy Station and Other Air Pollution Monitors



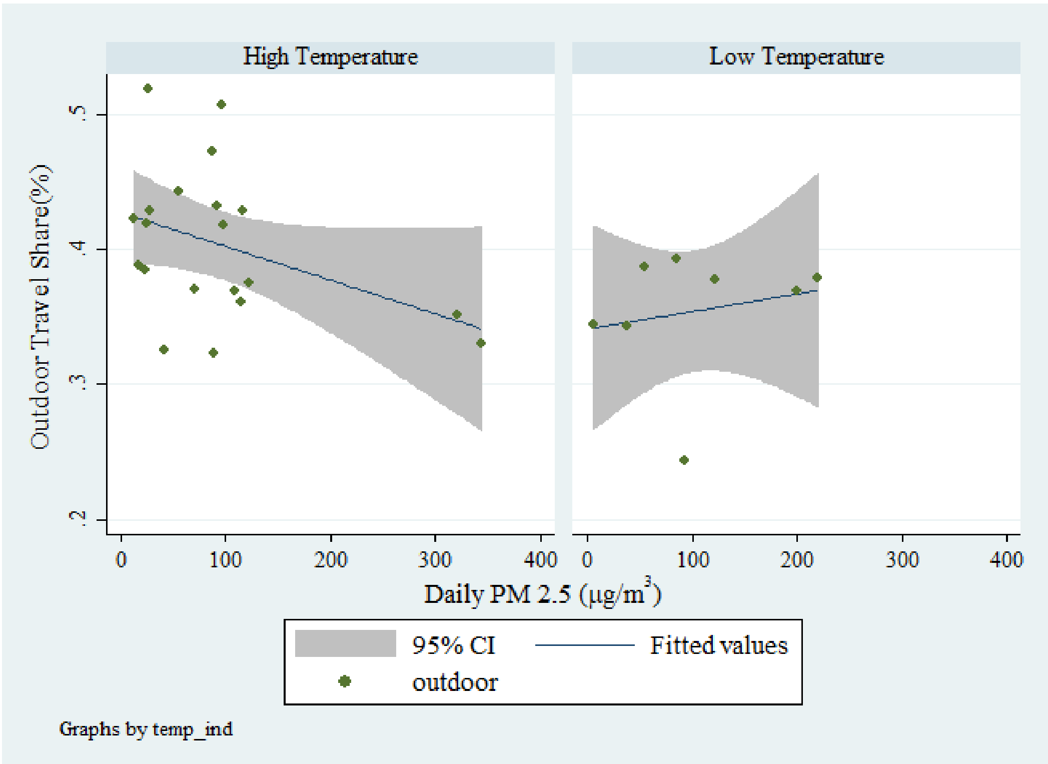
Note: The green dot represents the U.S. Embassy in Beijing. There are 35 observation stations from Chinese government as blue dots for year 2014, but for year 2010 we only have one U.S. Embassy station. To check the different travel patterns for year 2010 and year 2014, with respect to air pollution variation under the same scale, we use the U.S. Embassy station data.

Figure 4: Air Pollution Correlations from Different Monitors



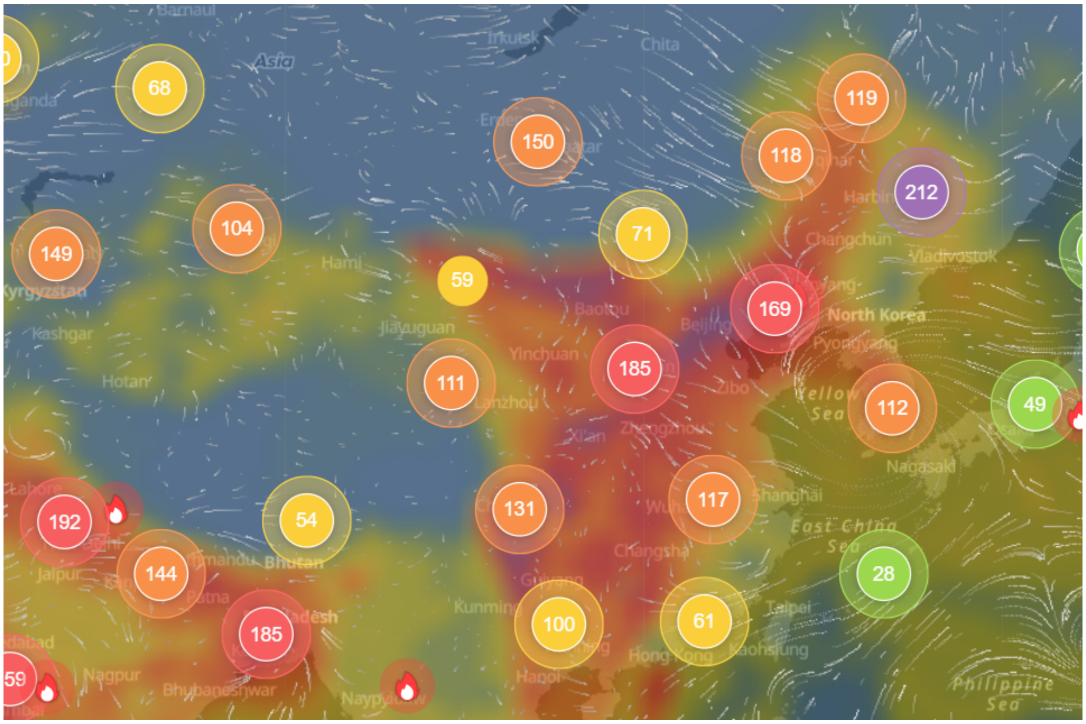
Note: The U.S. Embassy is located near the East Third Ring Road in Beijing. Figure 4 presents a data correlation diagram for the 8 monitoring stations near or inside of the third ring roads in Beijing. The Pearson correlation coefficients of the PM 2.5 data detected by these stations are above 0.94 and significant at the level of 0.01. Therefore, we believe that the data from the U.S. Embassy monitoring station is representative of the air conditions in Beijing.

Figure 5: Revisiting Figure 1 under Different Temperature Levels in Year 2010



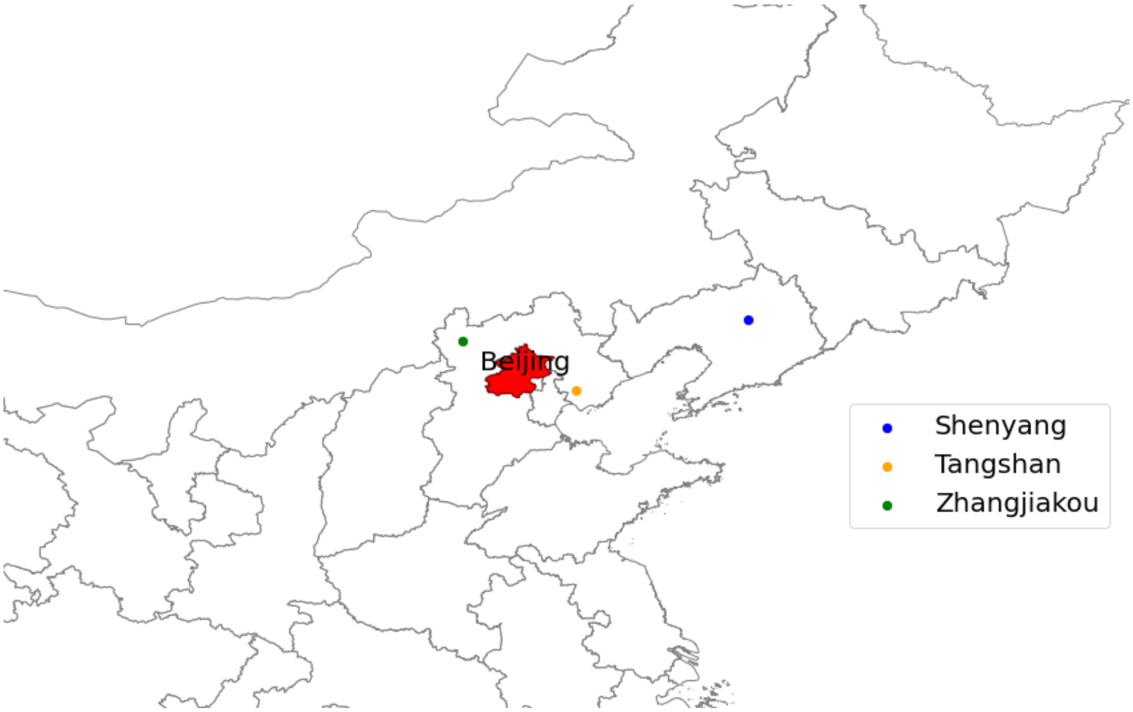
Note: In Figure 1, we examine the correlation between air pollution and the outdoor travel share. The positive correlation for the year 2010 suggests that in 2010, people might not have been willing to pay anything for clean air, despite the adverse effects of air pollution on their health. In Figure 5, we divide the temperature over high and low using 20 Celsius degree as a threshold to investigate the phenomenon in year 2010. In 2010, outdoor travel declines with air pollution when temperatures are high, but increases with air pollution when temperatures are low.

Figure 6: Wind and Air Pollution in China



Note: Figure 6 shows the pollution heat and wind direction in China for a random day in year 2021. It reflects the air pollution transmission highly depends on the air pollution. The source is https://www.iqair.cn/cn-en/air-quality-map.

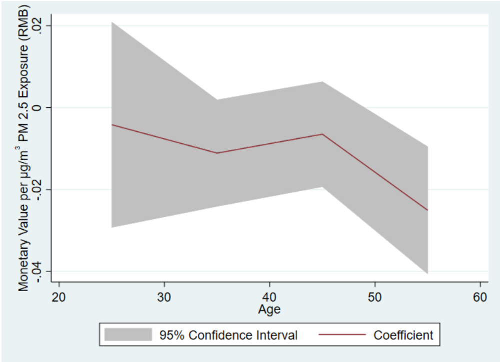
Figure 7: Selected Wind Locations



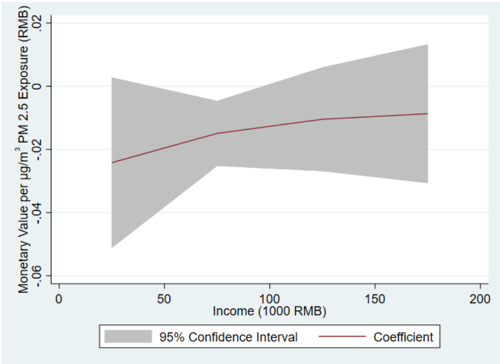
Note: The number of selected instruments is too large to attain optimal. Therefore, we check the overlap locations selected by LASSO – Tangshan, Shenyang and Zhangjiakou – in year 2010 and year 2014.

Figure 8: Logit Regressions for Different Age, Income, and Time Groups in Year 2014

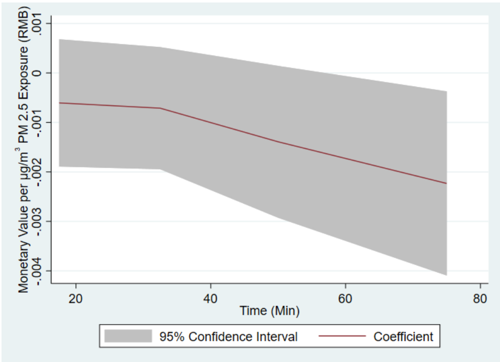
(a) Different Age Groups



(b) Different Income Groups



(c) Different Time Groups



Note: We divide the sample into different bins for ages [20, 30], [30, 40], [40, 50], and [50,∼], for income [0, 50,000], [50,000, 100,000], [100,000, 150,000], and [150,000,∼] RMB, and for potential walking time [10, 25], [25, 40], [40, 60], and [60,∼] minutes to run the logit regression. The figure reflects the heterogeneity disutility of one unit exposure to $PM_{2.5}$ for year 2014 with IV, standardizing the negative impact of cost.

Table 1: Individual Charateristics

	2010		2014	
	mean	sd	mean	sd
Annual inncome (RMB)	73,141.48	41,268.83	101,939.57	62,971.84
Number of cars owned	0.47	0.59	0.72	0.67
Female (=1)	0.45	0.50	0.44	0.50
Age (years)	37.78	10.43	38.69	10.11
Schooling (years)	13.74	2.79	13.87	2.86
Observations	21,657		27,055	

Notes: Table 1 reports respondent characteristics of all work commuting trips in these two years. The individual income increases from from 73,141.48 RMB in year 2010 to 101,939.57 RMB in year 2014, a 40% increase matching the growth rates averaging 10% in Beijing. The match implies the randomness of the household survey. The increase in the number of cars owned by each individual also reflects the wealth increment in Beijing. Other variables are similar in these two years.

Table 2: Travel Information

	2010		2014	
	mean	sd	mean	sd
Travel time (hours)	0.66	0.62	0.60	0.55
Travel cost (Yuan)	2.63	5.81	3.72	7.08
Walk	0.14	0.34	0.13	0.33
Bike	0.24	0.43	0.24	0.43
Bus	0.24	0.42	0.18	0.38
Subway	0.04	0.21	0.06	0.24
Car	0.24	0.43	0.31	0.46
Taxi	0.01	0.11	0.01	0.09
Observations	24,027		29,446	

Notes: Table 2 presents the work commuting trip information. The travel time is lower in year 2014. Beijing's rapid subway expansion from 2008 to 2014 led to an increase in aggregate welfare with modest congestion reduction. Additionally, more people uses car in 2014, compared to other modes. No number contradicts our intuition given that Beijing becomes richer over years.

Table 3: Hourly Air Pollution on Wind Directions First Stage

	(1) 2010 Tangshan	(2) 2014 Tangshan	(3) 2010 Shenyang	(4) 2014 Shenyang	(5) 2010 Zhangjiakou	(6) 2014 Zhangjiakou
North Wind	-77.45***	-20.75*	5.254	-21.45	-1.828	-11.06
	(18.718)	(11.787)	(27.424)	(13.714)	(17.190)	(13.390)
Northeast Wind	-34.77*	-35.21***	-33.00	-20.89	51.16***	-6.982
	(19.108)	(11.636)	(50.731)	(16.428)	(18.407)	(16.393)
East Wind	25.14	4.622	17.58	1.885	19.37	-28.79**
	(20.885)	(9.345)	(28.653)	(14.113)	(15.558)	(13.124)
Southeast Wind	64.53***	39.82***	45.76	19.63	121.3***	-20.04
	(20.684)	(12.863)	(28.595)	(14.906)	(21.556)	(13.576)
South Wind	-5.886	-20.11**	6.775	-5.456	102.6***	5.492
	(18.078)	(9.945)	(27.681)	(15.358)	(16.347)	(12.367)
Southwest Wind	-9.583	-51.96***	-75.97	-63.67**	73.44***	47.19***
	(24.983)	(14.080)	(48.067)	(29.265)	(20.601)	(14.437)
West Wind	-57.54***	-11.69	11.83	-30.67**	21.37	2.025
	(18.184)	(9.808)	(29.967)	(14.728)	(19.718)	(16.085)
Northwest Wind	-46.40**	76.32***	-13.81	-12.31	0.528	-20.61
	(22.301)	(18.658)	(29.571)	(15.069)	(27.412)	(25.947)
Wind Speed	0.345	-1.865***	-0.0773	0.244	0.814**	0.995***
	(0.312)	(0.333)	(0.407)	(0.210)	(0.356)	(0.275)
Observations	849	1026	849	1026	849	1026
Adjusted R^2	0.143	0.078	0.015	0.024	0.135	0.066

Notes: We check the overlap locations given by LASSO for year 2010 and year 2014, by running OLS for the first stage of hourly air pollution on their wind directions controlling for 18 hourly (6 AM to 11 PM), 2 monthly, and 7 weekday fixed effects, with other weather conditions in Beijing. Tangshan satisfies the matching of the predicted signs from relative position and the estimated coefficients of the wind directions. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Hourly Share Regression

	(1) 2010	(2) 2014	(3) 2010	(4) 2014	(5) 2010 IV	(6) 2014 IV	(7) 2010 IV	(8) 2014 IV
Pollution Exposure (100 $\mu g/m^3$)	0.0343	-0.0608*	0.0233	-0.0948***	-0.1811*	-0.3822***	0.2424	-0.4711
	(0.056)	(0.032)	(0.029)	(0.030)	(0.094)	(0.066)	(0.297)	(0.513)
Time (Hour)	-0.5681***	-0.7400***	-0.2515***	-0.2343***	-0.5797***	-0.7553***	-0.2558***	-0.2605***
	(0.082)	(0.067)	(0.047)	(0.052)	(0.086)	(0.068)	(0.049)	(0.053)
Cost (RMB)	-0.1155***	-0.0783***			-0.1214***	-0.0836***		
	(0.010)	(0.009)			(0.010)	(0.009)		
Alternative * 18 Hourly Fixed Effect	N	N	Y	Y	N	N	Y	Y
First Stage F test					54.91	54.91	54.91	54.91
Observations	900	1,118	900	1,118	860	1,116	860	1,116
Adjusted R^2	0.902	0.912	0.901	0.912	0.903	0.911	0.902	0.910

Notes: 6 alternatives are {Walking, Car, Subway, Bus, Taxi, Bicycle}. Standard errors are clustered at the daily level. Alternative specific weather conditions are controlled. The dependent variable is $lns_{mt} - lns_{1t}$ at hourly level. We exclude hours from 23 to 5. The instrument is the southeast wind in Tangshan. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Hourly Share Regression for Different Age Groups

	(1) 2010 Young	(2) 2010 Old	(3) 2014 Young	(4) 2014 Old	(5) 2010 Young IV	(6) 2010 Old IV	(7) 2014 Young IV	(8) 2014 Old IV
Pollution Exposure (100 $\mu g/m^3$)	-0.0072	-0.1972**	-0.0506*	-0.2120***	-0.2840***	0.0373	-0.4520***	-0.2819**
	(0.032)	(0.082)	(0.030)	(0.079)	(0.083)	(0.129)	(0.089)	(0.139)
Time (hour)	-0.6116***	-0.4197***	-0.8028***	-0.4740***	-0.5170***	-0.4272***	-0.8174***	-0.4778***
	(0.043)	(0.050)	(0.036)	(0.051)	(0.035)	(0.050)	(0.036)	(0.051)
Cost (RMB)	-0.1059***	-0.0117	-0.0844***	-0.0346**	-0.1227***	-0.0214	-0.0890***	-0.0463***
	(0.007)	(0.026)	(0.005)	(0.014)	(0.007)	(0.027)	(0.005)	(0.014)
Observations	850	354	1027	491	961	373	1030	493
Adjusted R^2	0.678	0.566	0.753	0.620	0.689	0.564	0.759	0.615

Notes: The age threshold is 50. Standard errors are clustered at the daily level. Alternative specific weather conditions are controlled. The dependent variable is $lns_{mt} - lns_{1t}$ at hourly level. The instrument is the southeast wind in Tangshan. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Hourly Share Regression for Different Income Groups

	(1) 2010 Low	(2) 2010 High	(3) 2014 Low	(4) 2014 High	(5) 2010 Low IV	(6) 2010 High IV	(7) 2014 Low IV	(8) 2014 High IV
Pollution Exposure (100 $\mu g/m^3$)	-0.0330	0.2078*	-0.0833***	0.0444	-0.2109***	-0.4539***	-0.3780***	-0.4787***
	(0.031)	(0.114)	(0.032)	(0.069)	(0.078)	(0.152)	(0.089)	(0.106)
Time (hour)	-0.5515***	-0.7464***	-0.6170***	-0.9816***	-0.4898***	-0.7794***	-0.6279***	-1.0076***
	(0.040)	(0.102)	(0.033)	(0.058)	(0.032)	(0.099)	(0.032)	(0.057)
Cost (RMB)	-0.1178***	-0.0527***	-0.0897***	-0.0758***	-0.1342***	-0.0618***	-0.0937***	-0.0808***
	(0.007)	(0.014)	(0.006)	(0.006)	(0.007)	(0.014)	(0.006)	(0.006)
Observations	871	290	930	627	980	305	934	628
Adjusted R^2	0.712	0.498	0.740	0.719	0.721	0.504	0.744	0.727

Notes: The income threshold is 100,000 RMB. Standard errors are clustered at the daily level. Alternative specific weather conditions are controlled. The dependent variable is $lns_{mt} - lns_{1t}$ at hourly level. The instrument is the southeast wind in Tangshan. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Logit Regression

VARIABLES	(1) 2010	(2) 2014	(3) 2010 IV	(4) 2014 IV
Pollution Exposure (100 $\mu g/m^3$)	0.0816**	-0.0429**	0.0511	-0.0454*
	(0.0338)	(0.0214)	(0.0396)	(0.0259)
Time (Hour)	-2.034***	-3.198***	-2.035***	-3.198***
	(0.0801)	(0.0853)	(0.0802)	(0.0853)
Cost (RMB)	-0.0350***	-0.0585***	-0.0350***	-0.0585***
	(0.00604)	(0.00513)	(0.00604)	(0.00513)
Observations	70,908	142,458	70,908	142,458

Notes: 6 alternatives are {Walking, Car, Subway, Bus, Taxi, Bicycle}. Alternative specific individual characteristics and weather condition are controlled, with instrument as the southeast wind in Tangshan. The standard errors are clustered at household level. The model also include 18 hourly fixed effects, excluding hours from 23 to 5. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix

A.1 Utility Maximization

The utility maximization problem is given by:

$$\begin{aligned} \max_{m \in M, e>0, x>0} \quad & h(p_m; e) + \gamma_2 x + \varepsilon_m \\ s.t. \quad & \gamma_2 \cdot (x + e + c_m) \leq \gamma_1 \cdot (T - t_m), \end{aligned} \tag{A.1}$$

where pollution exposure p_m depends on the outdoor air pollution level P_o, the indoor pollution level P_i, and whether the travel mode m is indoors or outdoors, and is given by:

$$p_m = P_i T + (P_o - P_i)\mathbf{1}[m = outdoor]t_m. \tag{A.2}$$

Proof. Since the utility function is monotone in x, the budget constraint is binding. Therefore, $\gamma_2 x = \gamma_1(T - t_m) - \gamma_2 e - \gamma_2 c_m$ and the maximization problem is equivalent to:

$$\max_{m \in M, e>0, x>0} \quad h(p_m; e) + \gamma_1(T - t_m) - \gamma_2 e - \gamma_2 c_m + \varepsilon_m. \tag{A.3}$$

A monotone transformation does not change the order of the preference. Thus, abstracting from the health expenditure choice problem and holding health expenditures fixed at $\bar{e}$, the average health expenditure over the year, we obtain:

$$\max_{m \in M} \quad h(p_m; \bar{e}) - \gamma_1 t_m - \gamma_2 c_m + \varepsilon_m. \tag{A.4}$$

We note that we can approximate the health or well-being function $h(\cdot; \bar{e})$ using a first-order

Taylor series approximation around $\bar{P}T$, where $\bar{P}$ is the average air pollution over the year, as follows:

$$\begin{aligned} h(p_m;\bar{e}) &= h(P_iT + (P_o - P_i)\mathbf{1}[m = outdoor]t_m;\bar{e}) \\ &\approx h(\bar{P}T;\bar{e}) + h'(\bar{P}T;\bar{e})((P_o - P_i)\mathbf{1}[m = outdoor]t_m + (P_i - \bar{P})T) \end{aligned} \tag{A.5}$$

Again the parts that do not contain m in the utility function do not influence the order of the preference. Thus, the utility maximization problem can be approximated by:

$$\max_{m\in M} \quad h'(\bar{P}T;\bar{e})(P_o - P_i)\mathbf{1}[m = outdoor]t_m - \gamma_1 t_m - \gamma_2 c_m + \varepsilon_m. \tag{A.6}$$

□

A.2 Appendix B. LASSO Regression

Denote our model as $\theta = (\beta_1, \ldots, \beta_s, \frac{1}{\sqrt{np}}\gamma_1, \ldots, \frac{1}{\sqrt{np}}\gamma_p)'$, $x = (x_1, \ldots, x_n)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$. $x = A\beta + \epsilon$, where $A_{ij} \overset{iid}{\sim} N(0,1)$ and $\frac{1}{n}||\varepsilon||_1 = \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| < v$. The following theorem validates the LASSO when $\sqrt{p\log(s+p)} << n$.

Theorem Let $\hat{\theta}$ solve $min\ ||\theta||_1$ s.t. $\frac{1}{n}||A\theta - x||_2^2 \leq v^2$.[16]

Then, $E||\hat{\theta} - \theta||_2 \leq \sqrt{8\pi}\left[\left(\sqrt{s}||\beta||_2 + \frac{\sqrt{p}|\bar{\gamma}|}{\sqrt{n}}\right)\frac{\sqrt{\log(s+p)}}{n} + v\right]$.

Proof. The intuition of this proof comes from Donoho and Elad (2003). Consider the convex hull $\bar{K} = conv\{\pm e_i\}_{i=1}^{s+p}$, where e_i are unit vectors in R^{s+p}, $x = \sum_{i=1}^{s}\beta_s e_i + \sum_{i=s}^{s+p}\frac{1}{\sqrt{np}}\gamma_i e_i$. Let $K = ||x||_1\bar{K}$.

[16] $||\cdot||_1$ represents the l_1 norm: $||\theta||_1 = \sum_{j=1}^{s+p}|\theta_j|$. $||\cdot||_2$ represents the l_2 norm (or Euclidean norm): $||x||_2 = \left(\sum_{i=1}^{n}x_i^2\right)^{\frac{1}{2}}$. The minimization problem is a special case of LASSO. The equivalence comes from the duality of the optimization problem.

Observe that $\hat{\theta}$ also minimizes the gauge $||\cdot||_K$ with respect to K:

$$min\ ||\theta||_{\bar{K}} \quad s.t.\ \frac{1}{n}||A\beta - x||_2^2 \leq v^2,$$

since

$$||\theta||_{\bar{K}} = min\ \{\lambda > 0|\ \theta = \lambda \sum_{i=1}^{s+p} \eta_i e_i,\ \eta_i > 0,\ \sum_{i=1}^{s+p} \eta_i = 1\}$$
$$= ||\theta||_1$$

We use the general M* bound (Milman, 1985; Pajor and Tomczak-Jaegermann, 1986), which ensures

$$E||\hat{\theta} - \theta||_2 \leq \sqrt{8\pi}\left(\frac{w(K)}{\sqrt{n}} + v\right),$$

where $w(K)$ is the gaussian mean width of K.

To conclude, according to the definition of K and θ,

$$\begin{aligned} w(K) &= ||\theta||_1 w(\bar{K}) \leq ||\theta||_1 \sqrt{\log(s+p)} \\ &= \left(\sum |\beta_i| + \frac{\sum |\gamma_i|}{\sqrt{np}}\right)\sqrt{\frac{\log(s+p)}{n}} \\ &\leq \left(\sqrt{s}||\beta||_2 + \frac{p\overline{|\gamma|}}{\sqrt{np}}\right)\sqrt{\frac{\log(s+p)}{n}}. \end{aligned}$$

The last inequality comes from Cauchy-Schwarz.

$\square$

A.3 Reduced-Form Regression

Proof. We rely on the utility function to get the reduced-form regression

$$
\begin{aligned}
u_{imt} = {} & \beta \mathbf{1}[m = outdoor] X_t \\
& + \gamma_1 time_{imt} - \gamma_2 cost_{imt} + \mathbf{z_{it}}' \eta_m + \alpha_m + \xi_{mt} + \varepsilon_{itm},
\end{aligned}
\tag{A.7}
$$

where $\beta = \beta_2 - \beta_1$ and it captures the utility of staying outside with an extra unit of $PM_{2.5}$. We assume the error term ϵ_{itm} is distributed type I extreme value. Therefore, we can get a closed-form probability function for individual i choosing mode m at time t with p_{imt},

$$
p_{imt} = \frac{exp(\beta \mathbf{1}[m = outdoor] X_t + \gamma_1 time_{imt} - \gamma_2 cost_{imt} + \mathbf{z_{it}}' \eta_m + \alpha_m + \xi_{mt})}{\sum_{k \in M} exp(\beta \mathbf{1}[k = outdoor] X_t + \gamma_1 time_{ikt} - \gamma_2 cost_{ikt} + \mathbf{z_{it}}' \eta_k + \alpha_k + \xi_{kt})}
$$

To obtain the the population share S_{mt} for mode m at time t, we take the expected value of p_{imt} over all individuals i (or, equivalently, integrate over the distribution of i):

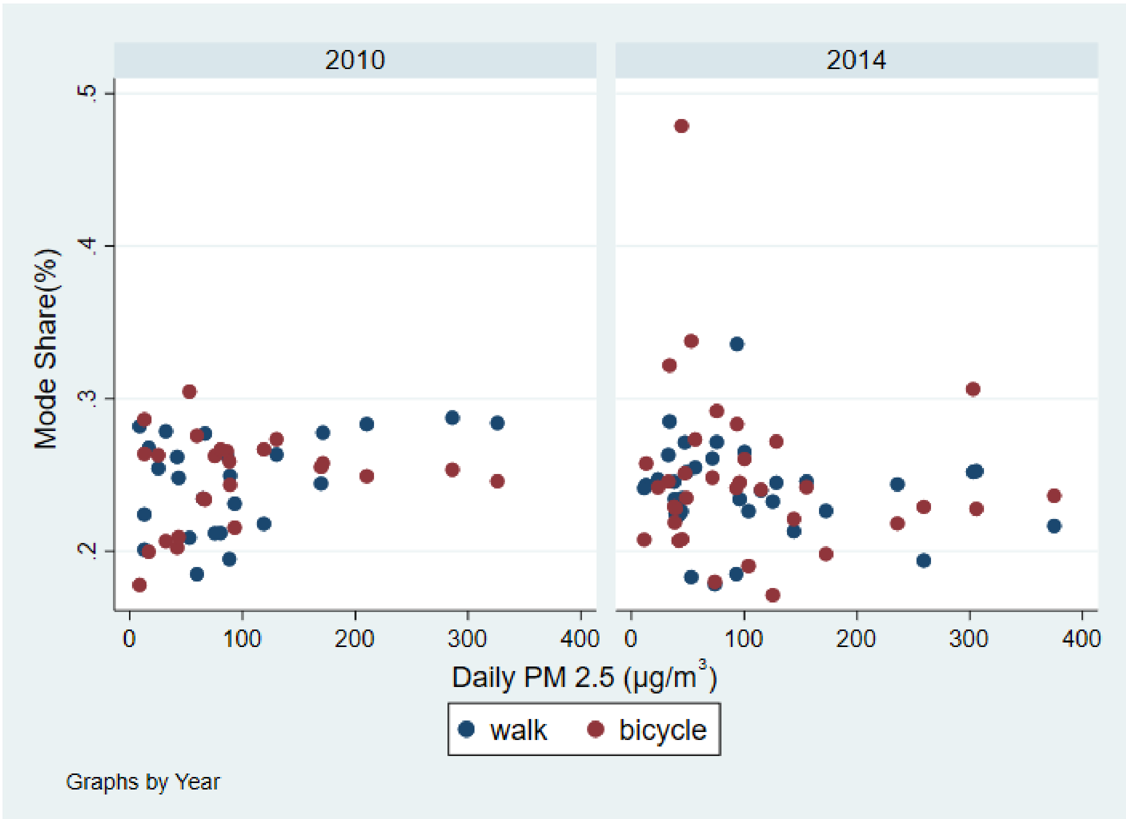
$$
S_{mt} = E\left[\frac{exp(\beta \mathbf{1}[m = outdoor] X_t + \gamma_1 time_{imt} - \gamma_2 cost_{imt} + \mathbf{z_{it}}' \eta_m + \alpha_m + \xi_{mt})}{\sum_{k \in M} exp(\beta \mathbf{1}[k = outdoor] X_t + \gamma_1 time_{ikt} - \gamma_2 cost_{ikt} + \mathbf{z_{it}}' \eta_k + \alpha_k + \xi_{kt})}\right]
$$

Normalizing the case-specific coefficients α_1 and ξ_{1t} for the base alternative $m = 1$ to 0, we get:

$$
\begin{aligned}
\ln S_{mt} - \ln S_{1t} = \beta \mathbf{1}[m = outdoor]X_t + \\
\gamma_1(E[time_{mt}] - E[time_{1t}]) + \gamma_2(E[cost_{mt}] - E[cost_{1t}]) + E[\mathbf{z_t}]'\eta_m + \alpha_m + \xi_{mt} \\
= \beta \mathbf{1}[m = outdoor]X_t + \\
\gamma_1(\overline{time}_{mt} - \overline{time}_{1t}) + \gamma_2(\overline{cost}_{mt} - \overline{cost}_{1t}) + \overline{\mathbf{z_t}}'\eta_m + \alpha_m + \xi_{mt},
\end{aligned}
$$

which is the closed form equation we are interested in. $\square$

Figure A.1: Daily Average Air Pollution and Outdoor Mode Shares



Note: Travel data are from the Beijing Household Travel Survey (BHTS) and air pollution average is from U.S. Embassy in Beijing. Outdoor travel modes are $\{Walking, Bicycle\}$. The full data description is in Section 4. The positive slope for year 2014 in the figure shows the existence of the avoidance behaviors in 2014, while the negative slopes for year 2010 implies that the WTP, depends on information and is the expected air pollution cost rather than the true loss. In 2010, people might not be willing to pay anything for the clean air, despite the adverse effects on their health. With the popularization of knowledge about air pollution, the WTP gets closer to the actual social welfare loss.