

NBER WORKING PAPER SERIES

THE HEALTHCARE COST OF AIR POLLUTION:
EVIDENCE FROM THE WORLD'S LARGEST PAYMENT NETWORK

Panle Jia Barwick
Shanjun Li
Deyu Rao
Nahim Bin Zahur

Working Paper 24688
<http://www.nber.org/papers/w24688>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2018, Revised May 2021

Previously circulated as “The Morbidity Cost of Air Pollution: Evidence from Consumer Spending in China.” We thank Yongmiao Hong, Solomon Hsiang, Matthew E. Kahn, Koichiro Ito, Doug Miller, Lucija Muehlenbachs, Matthew Turner, Shuang Zhang and seminar participants at Carnegie Mellon University, Chinese University of Hong Kong, Clemson University, MIT, National University of Singapore, Peking University, Penn State, University of Arizona, University of Chicago, and 2018 NBER Environmental and Energy Economics Workshop for helpful comments. We thank Nan Zhong and Youting Zheng for generous data sharing. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Panle Jia Barwick, Shanjun Li, Deyu Rao, and Nahim Bin Zahur. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Healthcare Cost of Air Pollution: Evidence from the World's Largest Payment Network
Panle Jia Barwick, Shanjun Li, Deyu Rao, and Nahim Bin Zahur
NBER Working Paper No. 24688
June 2018, Revised May 2021
JEL No. I15,Q51,Q53

ABSTRACT

Using the universe of credit- and debit-card transactions in China during 2013-2015, this paper provides the first nationwide analysis of the healthcare cost of PM2.5. We leverage spatial spillovers of PM2.5 from long-range transport for exogenous variation in local pollution and employ a flexible distributed lag model to capture semiparametrically the dynamic response of pollution exposure. Our analysis shows significant impacts of PM2.5 on healthcare spending in both the short and medium terms. A 10 mg/m³ decrease in PM2.5 would reduce annual healthcare spending by more than \$9.2 billion, about 1.5% of China's annual healthcare expenditure.

Panle Jia Barwick
Department of Economics
Cornell University
462 Uris Hall
Ithaca, NY 14853
and NBER
panle.barwick@cornell.edu

Shanjun Li
Cornell University
405 Warren Hall
Ithaca, NY 14853
and NBER
SL2448@cornell.edu

Deyu Rao
Warren Hall
Dyson School of Applied Economics
and Management
Cornell University
dr526@cornell.edu

Nahim Bin Zahur
Cornell University
Economics Department
Ithaca, NY 14853
nahimzahur@gmail.com

1 Introduction

The mortality and morbidity impact of air pollution is an essential component of the overall benefit of environmental regulations. The existing literature has primarily focused on the impact of air pollution on mortality.¹ Among the studies on the morbidity impact of pollution, most of them focus on specific health outcomes (such as asthma attacks) and the associated physiological channels of the impact.² There is a limited understanding of the aggregate morbidity cost of air pollution from all health outcomes and a lack of a commonly agreed method to measure it (WHO, 2015). Different from mortality, morbidity outcomes have diverse endpoints ranging from respiratory problems to cardiovascular diseases and lung cancer, as well as multiple complications that could arise for those with pre-existing conditions. Therefore, the morbidity outcomes are much harder to collect and measure on a large scale than mortality (Landrigan et al., 2018), especially in developing countries.

As a result of the increased pressure from economic development and lax environmental regulations, developing countries and especially emerging economies, such as China and India, are currently experiencing the worst air pollution in the world. This is especially concerning given the size of the population and the lack of access to adequate health care in these countries. While policymakers in these countries are increasingly aware of the negative impacts of air pollution on human health and quality of life, data on health outcomes are limited and rigorous empirical evidence on the health impact of air pollution is only emerging recently. Consequently, the dose-response relationships (between pollution exposure and health outcomes) estimated using data from developed countries have often been used as critical inputs for evaluating environmental regulations in developing countries, raising the question of external validity of this benefit-transfer approach (Arceo et al., 2015; OECD, 2016).

This study fills these two gaps in the literature by offering, to our knowledge, the first comprehensive, nationwide analysis of how air pollution affects health expenditures from all medical conditions for a developing country.³ We combine hourly air pollution readings from all monitoring stations from January 2013 to December 2015 with the universe of credit and debit card (or ‘bank card’) transactions in China during the same period. The transaction data come from the UnionPay Network, the largest payment network in the world, and the only inter-bank payment network in

¹For papers on mortality, see for example Chay and Greenstone (2003); Currie and Neidell (2005); Currie and Walker (2011); Chen et al. (2013); Knittel et al. (2015); Clay et al. (2016); Ebenstein et al. (2017); Anderson (2020).

²For example Pope (1989); Dockery (2009); Pope and Dockery (2012); Neidell (2004); Schlenker and Walker (2016).

³A growing literature uses health insurance claims data to examine the impact of air pollution on healthcare spending in the U.S. (Deschênes et al., 2017; Williams and Phaneuf, 2016; Deryugina et al., 2019). In developing countries, health insurance tends to be inadequately provided and detailed insurance data at the national level are hard to find. The current system of health-care delivery in China is fragmented, hospital-centered and treatment-dominated, with little effective collaboration among institutions in different tiers of the system (Wang et al., 2018b). This is a key challenge in obtaining consistent micro-level data of health outcomes for the whole country.

China. The data contain transactions for 2.7 billion bank cards that contribute to over \$5 trillion of economic transactions annually. In addition to covering 51% of private healthcare spending in China in 2015, this dataset also includes spending in over 300 non-healthcare categories. Our approach of using healthcare spending data (which include both the frequency and value of transactions) allows us to quantify the aggregate healthcare cost without explicitly examining every health outcome that is negatively affected by pollution. Although our data on bank card transactions in healthcare facilities do not contain information on the specific diagnoses or treatment associated with these transactions, we provide evidence on the strong correlation between our spending data and health outcomes at both the macro- and micro- levels.

There are two key empirical challenges in identifying the causal effect of air pollution on healthcare spending. The first challenge is the potential endogeneity in contemporaneous and lagged $PM_{2.5}$ that we use to capture pollution exposure. The endogeneity can arise from unobservables that affect both the pollution level and consumer spending (e.g., economic conditions). In addition, there could be measurement errors in constructing pollution exposure using air quality monitoring data. Ideally, residents' pollution exposure should be measured by the population-weighted local pollution, because the pollution level could vary greatly across locations within a city. However, monitoring stations are located sparsely across the country, preventing us from constructing population-weighted averages at a fine geographic scale.

To deal with this challenge, we construct instrumental variables by modeling the spatial spillovers of $PM_{2.5}$ due to fine particles' long-range transport property. Our IV approach is similar to the identification strategy used in [Bayer et al. \(2009\)](#), [Williams and Phaneuf \(2016\)](#), and [Deryugina et al. \(2019\)](#). The first two studies construct IVs based on air quality predictions from the EPA's source-receptor matrix that uses distant polluting facilities as inputs, while the latter study exploits changes in daily wind directions in a county as exogenous shocks to local air pollution. Based on a parsimonious model of $PM_{2.5}$ concentration in the spirit of EPA's air quality modeling, we disentangle the contribution of local and non-local sources and use $PM_{2.5}$ concentration from non-local sources as an exogenous variation. This allows us to leverage factors that directly affect pollution transport – wind patterns and other meteorological conditions in both the source and receptor cities, as well as geographic information such as distance – in constructing IVs.

Our instruments are essentially weighted averages of lagged $PM_{2.5}$ levels in distant cities where the weights are a function of the distance between the source and receptor cities, wind direction and speed, and other meteorological conditions. To examine the role of different identification variations, we experiment with alternative IVs, including the historical average and hence time-invariant level of air pollution in source cities, IVs that do not depend on local conditions, wind direction in the destination city interacted with regional dummies as in [Deryugina et al. \(2019\)](#), as well as placebo tests that randomize wind direction and speed. Our results indicate that both wind

direction and other meteorological conditions (wind speed, precipitation, and temperature) provide important exogenous identifying variation.

The second challenge in estimating the causal effect of pollution on healthcare spending arises from the nature of the high-frequency data. On the one hand, the rich data variation provides an opportunity to examine the dynamic impacts of past pollution exposure. On the other hand, daily pollution measures exhibit high autocorrelation. A direct OLS or IV estimation that includes many lagged terms leads to oscillating and imprecise estimates. We propose a flexible distributed lag model that extends the Almon technique (Almon, 1965) and uses finite-order B-splines (Corradi, 1977) to flexibly capture the effects of long lags. We combine this framework with the IV method to address endogeneity in contemporaneous and lagged air pollution measures. Our empirical framework is semiparametric in nature and can flexibly accommodate various data patterns.⁴

Our analysis based on daily healthcare spending by city shows that a short-run (i.e., contemporaneous) increase of $10 \mu\text{g}/\text{m}^3$ in $\text{PM}_{2.5}$ would lead to 0.65% more healthcare transactions. A medium-run (i.e., three-months) increase of $\text{PM}_{2.5}$ would lead to 2.65% more healthcare transactions. The impact of $\text{PM}_{2.5}$ differs across health facilities: spending in Children’s hospitals is more than twice as responsive as spending in other types of health facilities. For non-healthcare spending, we find a negative impact of $\text{PM}_{2.5}$ in the short-term but no significant impact beyond a few weeks. In addition, predicted worsening of air quality the next day increases the current day’s spending in both health and non-healthcare categories. These results provide evidence of avoidance behavior whereby consumers reduce outdoor activities (such as shopping) to mitigate pollution exposure.

We have examined a host of robustness checks, including various parametric specifications of the medium-term impact, different buffer zones, alternative B-spline segments, more flexible controls of meteorological conditions, the inclusion of other pollutants such as CO, SO₂ and average $\text{PM}_{2.5}$ in nearby cities, and different sample cuts. Our results are robust to these alternative specifications. The estimates are also similar if we conduct the analysis using the number of healthcare transactions per capita, or if we include controls for card penetration over time.

In monetary terms, a medium-run reduction of $10 \mu\text{g}/\text{m}^3$ in daily $\text{PM}_{2.5}$ would lead to annual savings in healthcare spending that exceed 59.6 billion *yuan*, or \$9.2 billion, with a 95% confidence interval ranging from 4.0-115.2 billion *yuan*.⁵ This is equivalent to \$22.4 per household per year. Bringing down China’s $\text{PM}_{2.5}$ to the World Health Organization’s (WHO) annual standard of $10 \mu\text{g}/\text{m}^3$ could lead to savings exceeding \$42 billion per year, nearly 7% of China’s national healthcare spending or 0.4% of China’s GDP in 2015.

How does the estimated healthcare cost from this study compare to the mortality cost estimates

⁴This framework is less restrictive than a more intuitive framework that regresses the current-day spending on the average pollution during a time window (e.g., the past week or month), where the effect of pollution is assumed constant over the time window.

⁵We use an exchange rate of \$1= 6.5 *yuan* throughout this analysis.

in the literature? [Ebenstein et al. \(2017\)](#) examine the mortality impact of PM₁₀ in China for different age groups. Their results imply that the monetized mortality cost based on the Value of a Statistical Life (VSL) is \$13.4 billion from a 10 unit increase in PM₁₀. Our estimated healthcare cost of \$9.2 billion is therefore about two-thirds of the mortality cost estimates in the literature. The ratio between pollution's healthcare cost and mortality cost in China is similar to the estimate derived by [Deschênes et al. \(2017\)](#) who analyze reductions in NO_x emissions in the U.S. These findings contribute to a better understanding of the significance of air pollution's morbidity cost and are in contrast to the common perception that morbidity is a minor component of the overall health impact of air pollution.⁶

Our study makes several contributions to the literature. First, to our knowledge, this is the first comprehensive study that analyzes the effect of pollution on healthcare spending at the national level for a developing country. Our paper adds to the growing literature that examines air pollution in developing countries ([Arceo et al., 2015](#); [Chen et al., 2013](#); [Greenstone and Hanna, 2014](#); [He et al., 2016](#); [Ebenstein et al., 2017](#)). Different from these studies which all focus on mortality, our analysis studies the impact of air pollution on spending in healthcare facilities. Among its recommendations to contain pollution's economic costs, the Lancet Commission on pollution and health ([Landrigan et al., 2018](#)) calls for further research to improve the morbidity cost estimates of pollution, recognizing that it is more difficult to measure the morbidity impact than mortality. Our analysis directly contributes to this research endeavor and highlights the economic magnitude of the morbidity impact.

Second, our analysis provides an alternative to the benefit-transfer approach commonly used in the literature to evaluate the health impact of air pollution in developing countries (due to a lack of rigorous empirical evidence from these countries). The benefit-transfer approach takes the dose-response function estimated in developed countries and interpolates the mortality or morbidity benefit from reduced air pollution to developing countries ([Lelieveld et al., 2015](#); [World Bank, 2007](#)). This approach may lead to significant inaccuracies due to differences in air pollution levels, baseline health conditions, and access to health care between these two groups of countries. In addition, to monetize the health impact, the dose-response function is then combined with sometimes ad hoc assumptions on the monetary costs for different illnesses (e.g., the cost of one asthma attack). Our analysis is not subject to these concerns. Our estimates suggest that China's elevated PM_{2.5} level relative to the WHO's annual standards entails \$42 billion additional healthcare expenditure in 2015. This estimate is an order of magnitude larger than the estimate in [OECD \(2016\)](#) based on the benefit-transfer approach.

Third, the rich spatial and temporal variation in our data allows us to examine both the short-

⁶[EPA \(2011\)](#) estimates that the morbidity benefit from the Clean Air Act from 1990 to 2020 is about 8% of the mortality benefit. [WHO \(2015\)](#) applies an additional 10% of the overall mortality cost as an estimate for the morbidity cost.

and medium-term impacts of air pollution on healthcare spending. Most studies focus on the contemporaneous impact by using daily or quarterly data and abstract away from the dynamic impact of air pollution. This is partly because it is difficult to disentangle the short-term and medium-term health impacts when current and lagged air pollution variables are both endogenous and at the same time exhibit high autocorrelations. We address this challenge by developing a novel approach that adapts a flexible distributed lag model to the IV setting. Our method is semiparametric, computationally light and has several advantages over existing methods such as VARs or local projection methods. It delivers a smooth impulse response function of both the short- and medium-term effects, easily incorporates instrumental variables, and can accommodate theoretical restrictions reflecting researchers' prior about the data generating process. To our knowledge, our study is the first analysis in the economics literature that exploits this technique to study the short- and medium-term health impacts with high-frequency data.

The rest of the paper is organized as follows. Section 2 describes the data and air pollution challenges facing China. Section 3 discusses our empirical framework and the identification strategy. Section 4 presents estimation results and Section 5 calculates the morbidity cost based on parameter estimates. Section 6 concludes.

2 Data

Our analysis is based on three comprehensive, nation-wide, micro-level datasets of air pollution, consumer spending by category, and meteorology conditions from January 2013 to December 2015, aggregated to daily and city-level. These datasets enable us to evaluate the impact of air pollution on spending in both the short- and medium-terms, as well as heterogeneous impacts across regions and pollution levels.

2.1 Air Pollution

For nearly four decades, China has maintained its GDP growth at an annual rate of nearly 10% and has transformed from an agricultural economy to a manufacturing-dominated economy. China became the world's largest exporter in 2009 and the largest trading nation in 2013. This unprecedented economic growth is largely propelled by fossil fuels, with coal accounting for about two-thirds of aggregate energy consumption and oil nearly twenty percent. China is by far the world's largest energy consumer, accounting for roughly a quarter of the world's total energy consumption and half of the world's coal consumption.

Fast economic growth and rising energy consumption have put enormous pressure on the environment, with air, water, and soil pollution becoming serious challenges that adversely affect

human health, ecosystems, and the quality of life.⁷ Improving air quality has become an important policy goal for the central government, which extensively revised the Environmental Protection Law in 2014 and defined goals of pollution abatement in both the 12th (2011 - 2015) and 13th (2016 - 2020) five-year plans.

Fine-scale air quality data at monitoring stations in China only became publicly available in 2013. The Ministry of Environmental Protection (MEP) publishes hourly measures of PM_{2.5}, CO, SO₂, NO₂, and O₃. The number of monitoring stations and cities covered increased steadily from 1003 stations in 159 cities in 2013 to 1582 stations in 367 cities in 2015. We calculate the daily concentration of PM_{2.5} and other pollutants at the city level by averaging data across monitoring stations within a city.

Air pollution affects human health mainly through its impact on respiratory and cardiovascular systems. Several decades of study in epidemiology and more recently in economics has associated exposure to air pollution with increases in mortality and morbidity risks (Brunekreef and Holgate, 2002; Pope and Dockery, 2012). Fine particles (PM_{2.5}), the focus of our analysis, are shown to be especially detrimental to health as they can penetrate deep into lungs and carry toxins to other organs. High levels of PM_{2.5} irritate respiratory and cardiovascular systems and can lead to aggravated asthma, lung disease, heart attacks, and stroke.

Figure 1 plots the three-year average of PM_{2.5} from 2013 to 2015 across cities. China's nationwide average during this period is 56 $\mu\text{g}/\text{m}^3$ (with a standard error of 46 $\mu\text{g}/\text{m}^3$), which is much higher than the annual standard of 12 $\mu\text{g}/\text{m}^3$ that is set by the U.S. Environmental Protection Agency and also higher than the standard of 35 $\mu\text{g}/\text{m}^3$ by China's MEP.⁸ Notably, there is considerable regional disparity. Cities in northern and central China with a high concentration of manufacturing industries suffer from the most severe pollution, with many of them experiencing a three-year average PM_{2.5} concentration of 90 $\mu\text{g}/\text{m}^3$ or higher. The less-developed regions in the west and wealthy regions in the south have better air quality. The latter, especially regions along the coast, has seen noticeable improvement in air quality as a result of shutting down or relocating polluting industries and reorienting the industry structure toward high tech and service industries.

One advantage of our empirical analysis is the rich variation in pollution measures both across cities and over time. To illustrate the time-series variation, we present in Appendix Figure A1 the daily PM_{2.5} concentration for the nation (the top panel) and separately for four regions (the bottom panel). The daily PM_{2.5} concentration is higher than 35 $\mu\text{g}/\text{m}^3$, the official MEP standard, in most days for all parts of the country. The northern regions have more pronounced peaks in winter than

⁷Lelieveld et al. (2015) estimate that air pollution led to 1.3 million premature deaths in China in 2010, accounting for 40% of the world's total premature deaths in the same year. World Bank (2007) puts the health cost of air pollution at 1.2-3.8% of China's GDP in 2003.

⁸U.S. EPA's daily standard is 35 $\mu\text{g}/\text{m}^3$ and annual standard is 12 $\mu\text{g}/\text{m}^3$. China's MEP sets limits on PM_{2.5} for the first time in 2012 to take effect in 2016: the daily standard is 75 $\mu\text{g}/\text{m}^3$ and annual standard is 35 $\mu\text{g}/\text{m}^3$.

the southern region, largely because of the coal-fired central heating systems north of the Huai River (Chen et al., 2013). The pollution level is trending downwards in all regions, driven by tightened government regulations, private and public investment in waste treatment, and changes in China’s overall industry structure.

2.2 Credit and Debit Transactions

The second main database for our analysis is the universe of credit and debit card (or ‘bank card’) transactions in China that are settled through the UnionPay network. The UnionPay network is the only inter-bank payment network in China and is state-owned. It is the largest network in the world in terms of both the number and value of transactions, ahead of Visa and Mastercard. There were 2.7 billion cards in use from 2013 to 2015, covering over 300 merchant categories and contributing to over \$5 trillion of economic transactions annually.^{9,10} We observe the location, time, merchant name, and amount for all transactions and aggregate the data to daily spending by category and city. To our knowledge, these are the most comprehensive and fine-scale data on consumer spending in China in temporal and spatial dimensions, and we are the first to utilize them for academic research. It is worth noting that during our sample (2013-15), the use of mobile payment (such as WeChat Pay and AliPay) was limited. In 2015, the share of mobile payments in China’s total retail consumption was only 8%, compared to 44% for debit and credit cards (Kapron and Meertens, 2017).

Healthcare spending includes transactions at hospitals, pharmacies, and other healthcare facilities (e.g. small health clinics). We exclude transactions exceeding 200,000 *yuan* (\$30,770).¹¹ In 2015, hospitals account for 83.5% of healthcare spending and 56.8% of healthcare transactions. Different from pharmacies in the U.S., such as CVS or Walgreens, most pharmacies in China only carry medicine and do not sell daily necessities. Pharmacies account for 6.0% of healthcare spending and 31.0% of healthcare transactions in 2015. The remaining transactions are accounted for by other healthcare facilities. Within hospitals, we identify People’s hospitals and Children’s hospitals based on merchant name. People’s hospitals are state-owned general hospitals and tend to be the largest health care facilities in a city. Each city has at least one People’s hospital but not all cities have Children’s hospitals, which accept mostly child patients. People’s and Children’s hospitals account for 24.1% and 4.2% of total healthcare spending respectively, and 26.2% and 9.0% of

⁹There were 800 million bank-card holders in China in 2015. The latest available census (2010-2011) shows that China’s population was 1.33 billion, out of which 221 million were below the age of 15 (and ineligible to hold bank cards). Of the 1.1 billion eligible individuals, 72% hold at least one bank card.

¹⁰There are seven major categories and 300 subcategories. The major categories are: retail; wholesale; direct sales; real estate and finance; residential and commercial service; hotel, restaurant, and entertainment; and education, health, and government service. Merchants are classified by these categories.

¹¹200,000 *yuan* (\$30,770) is the 99th percentile of transaction values across all categories. Transactions exceeding 200,000 *yuan* are excluded from the database due to UnionPay’s data quality assurance protocol that aims to remove potentially fraudulent transactions (a practice called “cash out”).

transactions in 2015.

Our data account for 31% of total private healthcare spending in 2013. As card penetration grew, the coverage rose to 51% in 2015, similar to the share of bank card transactions in other sectors. According to official statistics from the [Central Bank of China \(2015\)](#), bank card transactions accounted for 48% of overall spending in retail sales of consumer goods in the third quarter of 2015. In comparison, spending from credit and debit cards accounts for 55% of all consumer spending in 2012 in the U.S. ([Bagnall et al., 2014](#)).

Appendix Figure [A2](#) illustrates the spatial pattern of card adoption by plotting the number of active cards per resident by city in 2015. Card adoption is higher in coastal or high-income cities. Table [A1](#) in Appendix [A](#) correlates the cross-sectional card adoption rate with city demographics. Cities with a higher household income and education and a younger population are associated with higher adoption.

In addition to healthcare spending, we also analyze spending in non-healthcare categories, such as daily necessities. We closely follow the United Nations' Classification of Individual Consumption According to Purpose (COICOP) in defining necessity goods.¹² Relative to healthcare spending, spending on daily necessities is three times as large and transactions three times as frequent. A unique feature of Chinese consumers' shopping behavior is their frequent trips to supermarkets for groceries (often on a daily basis). We therefore use supermarket spending as another proxy for daily consumption, in addition to spending on necessities.¹³ Spending in supermarkets is over four times as large as healthcare spending in value and five times as frequent in 2015.

To illustrate inter-temporal spending patterns, Appendix Figure [A3](#) plots weekly healthcare spending and the number of transactions at the national level from 2013 to 2015. There is a significant drop in both the spending amount and the transaction frequency during holidays. In addition, both variables have more than tripled during our sample period due to the diffusion of bank cards. We control for these two salient features in our regression analysis through holiday fixed effects and city-specific time trends.

To graphically illustrate the relationship between pollution and spending, we plot the log number of transactions against contemporaneous $PM_{2.5}$ in Figure [2](#). All other controls (weather, city trend, etc.) are partialled out, so the figure displays the net correlation of pollution with spending. For ease of presentation, we group $PM_{2.5}$ by percentiles and plot the in-group average of log number of transactions against each percentile of $PM_{2.5}$. In addition to the aggregate number of healthcare transactions (top left), we also plot the relationship separately for pharmacies, People's hospitals, Children's hospitals, and two non-healthcare categories (necessities and supermarkets). $PM_{2.5}$ has a positive relationship with spending in all health categories and a negative relationship

¹²United Nations' COICOP defines necessity goods as 1) food and non-alcoholic beverages, 2) alcoholic beverages, tobacco and narcotics, 3) clothing and footwear, 4) recreation and culture, and 5) restaurants and hotels.

¹³We exclude supermarkets from necessity spending because they sell a large variety of goods other than necessities.

with non-health spending across nearly all quantiles of $PM_{2.5}$. The contrast in the relationship of $PM_{2.5}$ with health versus non-health spending is interesting and suggests potential causal impacts: elevated air pollution negatively affects health and leads to avoidance behavior among consumers. We quantify the causal impact in our regression analysis below.

2.3 Health Insurance and Health Outcomes

Health care in China is financed by government programs, individuals' out-of-pocket spending, and commercial health insurance. There are three major public health insurance programs, covering urban employees, urban non-employee residents, and rural residents, respectively. Through massive government subsidies and successful public campaigns, China achieved nearly universal health care coverage in 2011, when 95% of the population was covered through these three government supported insurance programs, up from 65% in 2009 (Yu, 2015). Commercial health insurance is rare and accounts for a negligible fraction of national health spending (Choi et al., 2018).

Despite the nearly universal health insurance in China, the coverage is low with high coinsurance rates (the fraction of health care expenses paid by individuals out of pocket) and low coverage ceilings that vary across insurance programs, healthcare facilities, and cities (or counties) which are the local risk pooling units in China (Meng and Yang, 2015). For example, the 2016 coinsurance rate (after deductible) in Shanghai varied from 25%-50% across tiers of healthcare facilities for employees and 10%-30% for retirees for out-patient visits and 8-20% for in-patient visits.¹⁴ In terms of the drug coverage, the Ministry of Human Resources and Social Security maintains the National Reimbursement Drug List. Only drugs on this list are covered by China's public health insurance programs, some in full (type A drugs) and others partially (type B). In most cases, individuals can purchase drugs without a doctor prescription.

Nearly all covered medical expenses (e.g., hospital visits and drug purchases) require some individual contributions through either bank card payments (which are included in our database) or cash.¹⁵ In most cases, out-patient care requires payment up front before receiving treatment, while in-patient care is billed several times a week (Jha, 2014; Browne, 2005). In light of this, the number of health-related transactions recorded in our database should capture well the number of visits to healthcare facilities and serves as a key outcome variable in our empirical analysis.

The bank card transactions do not identify specific disease diagnoses or treatment associated with the spending. This may raise concerns of whether the healthcare spending data correspond well to health outcomes. We provide several pieces of evidence that validate the data quality. We first obtain data on the aggregate number of hospital visits by in-patients, out-patients, and ERs in

¹⁴<http://sh.bendibao.com/zffw/2016325/158153.shtm>. Last accessed in Nov 2018.

¹⁵The fraction of medical expenses that is covered by the government health insurance programs is directly billed on health insurance cards and goes through a different clearing system from UnionPay.

each province from the annual China Statistical Yearbook that is published by the National Bureau of Statistics (NBS). This allows us to examine the correlation between our healthcare spending data and the national-level healthcare statistics. Appendix Figure A4 plots the number of card transactions in hospitals against the number of hospital visits in logarithms at the province-year level for our sample period. There is a close relationship between these two series with a high correlation coefficient: 0.86 in logs and 0.75 in levels, indicating that the number of card transactions is a good proxy for hospital visits.

Appendix Table A2, Table A3, Figure A5, and Figure A6 provide further evidence based on two confidential micro-level data sets, including the universe of medical emergency calls in Beijing and healthcare insurance claims in Ganzhou city, Jiangxi Province. As the capital city, Beijing has a highly educated population and a high penetration of bank cards. Ganzhou, on the other hand, is a medium-sized city that is primarily rural. In both cases, there is a strong correlation between our spending data and micro-level health outcomes, which is reassuring. Patterns from these very different examples suggest that the spending data provide reliable measures of health outcomes. Given the lack of micro-level data on health outcomes at the national level, our data provide to our knowledge the only alternative health-related measures that are both granular and have national coverage in China.

2.4 Meteorology Data and Summary Statistics

We obtain meteorology data from the Integrated Surface Database (ISD) that is hosted by National Oceanic and Atmospheric Administration (NOAA). The ISD dataset includes hourly measures of temperature, precipitation, wind speed, and wind direction for 407 monitoring stations in China, covering most major Chinese cities. We match cities with the nearest weather station according to their geographic coordinates and compute daily temperature and wind speed from a simple average of the hourly data.

ISD’s hourly measure of precipitation suffers from noticeable measurement errors, so we use daily precipitation from NOAA’s *Global Surface Summary of the Day* database (GSOD) instead.¹⁶ Daily wind direction is calculated by adding up twenty-four hourly vectors of wind direction, where the length of each vector is the hourly wind speed.

Table 1 reports the summary statistics for all variables used in our study at the city-day level. The daily PM_{2.5} concentration is on average 56 $\mu\text{g}/\text{m}^3$, with the inter-quartile range from 27 to 69 $\mu\text{g}/\text{m}^3$. The maximum recorded daily PM_{2.5} is 985 $\mu\text{g}/\text{m}^3$. Sixty-seven percent of these city-day observations record a concentration level that is above the U.S. daily standard of 35 $\mu\text{g}/\text{m}^3$. For healthcare spending, the average daily number of transactions is 7,229 per city, and the average

¹⁶GSOD reports daily precipitation using Greenwich Mean Time, which is the cumulative rainfall from 8 a.m. Beijing time to 8 a.m. the next day. We use this measure as our daily precipitation.

daily spending is 6.7 million *yuan*.

3 Empirical Framework

In this section, we first present a flexible econometric model that allows us to estimate the short- and medium-term impacts of air pollution on healthcare spending. Then we discuss our estimation strategy and the construction of instrumental variables.

3.1 Flexible Distributed Lag Model

Air pollution has both short- and long-term consequences on healthcare spending. Different from quarterly or annual data commonly used in the literature, our daily data allow us to characterize the path of health impacts from both contemporaneous and past air pollution exposure. We use the following distributed lag model (DL) to capture this relationship:

$$y_{it} = \sum_{\tau=0}^k \beta_{\tau} p_{i,t-\tau} + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \quad (1)$$

where y_{it} is daily healthcare spending in city i on day t , $p_{i,t-\tau}$ is either contemporaneous ($\tau = 0$) or lagged pollution exposure ($\tau \geq 1$), and k is the number of lagged pollution variables. \mathbf{x}_{it} includes a rich set of controls such as weather conditions, holiday fixed effects, day-of-week fixed effects, and seasonality. $\theta_i \cdot t$ is a city-specific linear time trend, ξ_i is city fixed effect, and η_w is week-of-the-sample fixed effect. The key parameters of interest are β_{τ} 's, which capture the short- and longer-term causal impacts of pollution exposure on healthcare spending.

The short-term impact of pollution is characterised by β_0 , which captures responses in healthcare spending to a contemporaneous increase in pollution concentration, everything else fixed. The long-term impact, or the cumulative impact of pollution, is characterised by $\sum_{\tau=0}^k \beta_{\tau}$, which reflects changes in healthcare spending as a result of persistent elevation in past pollution exposure. We are interested in quantifying both the short-term and the long-term impact of pollution.

Suppose, for a moment, that there is no measurement error in pollution exposure $p_{i,t-\tau}$ and that there are no omitted variables, two important issues we return to in the next section. Then the DL model can be estimated using OLS. But the linear estimation with a large number of lags is undesirable due to the high autocorrelation among the lag variables $p_{i,t-\tau}$. The parameter estimates tend to be imprecise with artificial oscillations as shown in Appendix Table B4. Alternatively, one can use the average pollution during a time window (such as the past week or past month) as in the following framework:

$$y_{it} = \beta \bar{p}_{it} + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it}, \quad (2)$$

where $\bar{p}_{it} = \frac{1}{k+1} \sum_{\tau=0}^k p_{i,t-\tau}$. Although this specification is easy to implement and addresses the issue of high autocorrelation, it imposes a strong restriction that all lagged pollution variables within the time window have a constant impact on current-day spending and does not allow researchers to examine the dynamic time-path of pollution's impact. We present results from this specification as a robustness check in Section 4.2.

To allow for flexible and smooth longer-term impacts and at the same time dealing with the issue of high autocorrelation, we extend Almon (1965) and specify β_τ 's in equation (1) as cubic B-spline functions of time with z segments (where z is a constant chosen by econometricians), following Corradi (1977).¹⁷ The intuition is that any smooth function (here β_τ can be treated as a function of time) defined on a closed interval $[a, b]$ can be approximated uniformly closely by basis splines.

To illustrate our approach, consider the example of cubic B-splines with one segment which amounts to a simple 3rd order polynomial:

$$\beta_\tau = \gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3, \quad (3)$$

where the contemporaneous effect of pollution on spending is captured by $\beta_0 = \gamma_0$, the effect of yesterday's pollution is $\beta_1 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3$, and the effect of pollution from τ days' in the past is $\beta_\tau = \gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3$. Appendix B.1 describes how to extend this to the more general case where there are multiple segments and the coefficients β_τ are piecewise polynomials in τ . Plug (3) into (1) and rearrange terms, we have:

$$\begin{aligned} y_{it} &= \sum_{\tau=0}^k \beta_\tau p_{i,t-\tau} + \mathbf{x}_{it} \alpha + \theta_{it} + \xi_i + \eta_w + \varepsilon_{it} \\ &= \gamma_0 p_{it} + (\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3) p_{i,t-1} + \dots + (\gamma_0 + \gamma_1 k + \gamma_2 k^2 + \gamma_3 k^3) p_{i,t-k} \\ &\quad + \mathbf{x}_{it} \alpha + \theta_{it} + \xi_i + \eta_w + \varepsilon_{it} \\ &= \gamma_0 v_{1,it} + \gamma_1 v_{2,it} + \gamma_2 v_{3,it} + \gamma_3 v_{4,it} + \mathbf{x}_{it} \alpha + \theta_{it} + \xi_i + \eta_w + \varepsilon_{it} \end{aligned} \quad (4)$$

where $v_{1,it} = p_{it} + p_{i,t-1} + p_{i,t-2} + \dots + p_{i,t-k}$, $v_{2,it} = p_{i,t-1} + 2p_{i,t-2} + \dots + kp_{i,t-k}$, $v_{3,it} = p_{i,t-1} + 2^2 p_{i,t-2} + \dots + k^2 p_{i,t-k}$, and $v_{4,it} = p_{i,t-1} + 2^3 p_{i,t-2} + \dots + k^3 p_{i,t-k}$, respectively. These four terms in equation (4) now constitute our key regressors. The first term, $v_{1,it}$, is the sum of past pollution exposure. The second to the fourth terms, $v_{2,it}, \dots, v_{4,it}$, are weighted sums of past exposure with the weights being polynomial terms of time. With this reformulation, we only need to estimate four coefficients $\{\gamma_i\}_{i=0}^3$ rather than $k+1$ coefficients (the number of lags plus current day). Once we

¹⁷Almon (1965) first proposed approximating the lag coefficients with polynomial functions. Poirier (1975), Corradi and Gambetta (1976) and Corradi (1977) suggested using spline functions, which impose weaker restrictions on the lag coefficients than polynomials while keeping the number of parameters small.

obtain the OLS or IV estimates and standard errors of γ 's, we can recover β_τ 's using equation (3).

In summary, the flexible distributed lag model transforms a series of many lagged pollution variables $\{p_{i,t-\tau}\}_\tau$ into a small number of $\{v_{\cdot,it}\}$'s, which are weighted sums of past pollution exposure with the B-spline functions of time as weights. This approach has several advantages over competing distributed lag models, the most popular one being the geometric decay model. First, these new regressors $\{v_{\cdot,it}\}$ exhibit much less multicollinearity than lagged pollutions $\{p_{i,t-\tau}\}_{\tau=0}^k$. Second, this model allows for much more flexible time-series patterns of the marginal impact β_τ than those in geometric decay models. Third, it is straightforward to impose additional restrictions that are generated by economic theories or reflect prior knowledge of the data generating process. For example, if tomorrow's pollution exposure (forward one period) should not affect current healthcare spending, then $\beta_{-1} = 0$. If pollution exposure prior to k lags has no effect, then $\beta_{k+\tau} = 0, \forall \tau \in \mathbb{N}$ and $\tau > 0$. These assumptions can be imposed individually or jointly as estimation constraints and tested as linear restrictions.¹⁸ Finally, we allow for an arbitrary correlation between the contemporaneous error term ε_{it} and past error terms, which is difficult in geometric decay models.

Our benchmark specification incorporates 90 daily lags ($k = 90$) and allows the time series patterns of the marginal impact β_τ in each month to be characterized by a separate cubic polynomial. This corresponds to a cubic B-spline with three segments, which leads to six regressors $\{v_{1,it}, \dots, v_{6,it}\}$ and six γ parameters to be estimated. We examine robustness to different numbers of lagged pollution variables and spline segments in Section 4.2.

3.2 Identification

3.2.1 Sources of Endogeneity

There are multiple factors that would render the OLS estimates as discussed above inconsistent. As recognized in the recent literature on estimating the causal impact of air pollution on health (Currie and Neidell, 2005; Arceo et al., 2015; Knittel et al., 2015; Schlenker and Walker, 2016; Deryugina et al., 2019), the pollution exposure variable likely suffers from measurement errors. This is because pollution levels vary across locations within a city and pollution readings from different monitoring stations are averaged to the city level. For example, among the 9 monitoring stations in the urban core of Beijing, the average difference between the maximum and minimum pollution level in a day is $35 \mu\text{g}/\text{m}^3$ in 2014, a sizable gap given the daily average of $87 \mu\text{g}/\text{m}^3$ at the city level. Since population is unevenly distributed within a city and the spatial distribution of monitoring stations does not align with residential areas, the arithmetic mean across all stations

¹⁸Another benefit over other distributed lag models is that this specification does not require instruments for the lagged dependent variable (lagged consumption in our setting), which is often challenging.

within a city may not accurately reflect the city population’s pollution exposure. An ideal measure would be the population-weighted average of local air quality, but this is impractical due to the lack of air pollution data at the finer spatial level (e.g., city block or zip code) and the fact that many monitoring stations are located outside of population centers. In addition, our daily pollution measure is a simple average of hourly measurements and abstracts away the temporal variation. To the extent that these measurement errors are classical, our OLS estimates would suffer from attenuation bias.¹⁹ City fixed effects are unlikely to adequately address these measurement errors, which vary over time. For example, on days when there is more local pollution in densely populated areas, the difference between the population’s pollution exposure and the simple average pollution will be larger.

Another factor that makes OLS estimates inconsistent is unobservables that are correlated with pollution. Despite our rich set of controls for weather and local conditions (e.g., city specific time trend and seasonality), there are various sources of temporal variations that cannot be adequately controlled for. For example, permanent local shocks to healthcare spending, such as income shocks, could be correlated with economic activities and thus with air quality. Temporary local shocks, such as major sport and political events and traffic congestion, could affect both the air pollution level and healthcare spending (and consumer activities in general).²⁰ These unobservables that are not absorbed by our location fixed effects and trend/seasonality interactions would render the air quality variable endogenous.

3.2.2 IV Construction

To address these concerns, we construct instruments by exploiting the spatial spillovers of PM_{2.5} due to its long-range transportability. PM_{2.5} particles are light, can travel at a speed of 10 mph, and often reside in the atmosphere for 3-4 days (Yassin et al., 2012; Díaz and Dominguez, 2009; Riva et al., 2011; Joksić et al., 2009). Their region of influence is determined by wind speed and direction. Based on atmospheric modeling, Zhang et al. (2015) document significant regional pollutant transport in China. For example, nearly half of the pollution in Beijing originates from sources outside of the municipality. These results suggest that PM_{2.5} from other cities could serve as exogenous shocks to the pollution level for a given city.

We use a parsimonious model to apportion observed pollution levels into components from local and non-local sources (see Appendix B.2 for more details on modeling and IV construction). Denote the pollution level of city i in time t as p_{it} , which is a function of past pollution and pollution

¹⁹Satellite data on Aerosol Optical Depth (AOD) offer an alternative measure of the ground level pollution with finer spatial resolutions (e.g., 3 km by 3 km from Terra satellite and 10 km by 10 km from Aqua) (Zou, 2018). However, there are a lot of missing values at the daily level due to cloud coverage.

²⁰An unexpected increase in congestion on a given day (e.g., due to accidents or weather conditions) raises air pollution and at the same time reduces healthcare spending (residents might prefer to stay at home on more congested days).

from other cities:

$$p_{it} = \theta_1 p_{i,t-1} + \underbrace{\sum_{j \neq i, d_{ij} \leq r} p_{j \rightarrow i, t}^+}_{\text{PM}_{2.5} \text{ imported from nearby cities}} + \underbrace{\sum_{j \neq i, d_{ij} > r} p_{j \rightarrow i, t}^+}_{\text{PM}_{2.5} \text{ imported from distant cities}} + \mu_{it}$$

where θ_1 captures the amount of pollution that is carried over from the previous day (which is affected by local meteorological conditions), $p_{j \rightarrow i, t}^+$ denotes the amount of PM_{2.5} pollutants in city i at time t that is originated from city j , d_{ij} represents the distance between cities i and j , r is the radius of a buffer zone, and μ_{it} is the error term. The total amount of PM_{2.5} imported by city i is the sum of $\sum_{j \neq i, d_{ij} \leq r} p_{j \rightarrow i, t}^+$ (pollution imported from cities within the buffer zone) and $\sum_{j \neq i, d_{ij} > r} p_{j \rightarrow i, t}^+$ (pollution imported from cities outside the buffer zone).

The contribution of non-local sources to the pollution level of a given city could be affected by a host of meteorological conditions and is the subject of sophisticated air quality modeling.²¹ We use the following parsimonious model to capture the key feature that PM_{2.5} pollutants dissipate over time and across space as they move:

$$p_{j \rightarrow i, t}^+ = \max[\cos \Phi_{ji}, 0] \cdot p_{j, t-s_{ijt}} \cdot f(d_{ij}, w_{j, t-s_{ijt}}, w_{i, t}), \quad (5)$$

where $p_{j \rightarrow i, t}^+$ is the amount of pollution that enters city i on day t , having originated from city j on day $t - s_{ijt}$. Pollution decays over time as it travels and only part of the pollution from city j enters the atmosphere of city i . This is represented by $f(d_{ij}, w_{j, t-s_{ijt}}, w_{i, t}) \in [0, 1]$, which is a function of the distance between the two cities (d_{ij}), weather conditions in the source city when the pollution is generated ($w_{j, t-s_{ijt}}$), and weather conditions in the destination city when the pollution enters its atmosphere ($w_{i, t}$). To take into consideration the effect of wind direction and speed, we invoke a vector decomposition. Let Φ_{ji} denote the angle between the wind direction and the direction from city j to city i , and $v_{j, t-s_{ijt}}$ denote the wind speed in city j . The amount of pollutants carried toward city i from city j is assumed to be $\cos(\Phi_{ji}) p_{j, t-s_{ijt}}$ at speed $\cos(\Phi_{ji}) v_{j, t-s_{ijt}}$. Note that $p_{j \rightarrow i, t}^+$ is zero if $\cos(\Phi_{ji})$ is negative: when wind blows away from city i , pollution from the source city j should not contribute to that of the receptor city. The number of days it takes pollutants to travel from city j to city i , s_{ijt} , is rounded to the next smallest integer: $s_{ijt} = \left\lfloor \frac{d_{ij}}{\cos(\Phi_{ji}) v_{j, t-s_{ijt}}} \right\rfloor$. As an example, Appendix Figure B7 illustrates graphically all subvectors of pollutants that are blown towards Beijing on Dec. 5, 2013.

We now describe how to construct instruments using the above model. The decay function $f(d_{ij}, w_{j, t-s_{ijt}}, w_{i, t})$ in equation (5) is unknown. We approximate it by a set of polynomial functions

²¹Meteorological conditions play a key role in the diffusion of PM_{2.5}. See, for example, Seibert and Frank (2003); Tai et al. (2012); Wang and Ogawa (2015); Wang et al. (2018a) and Wang et al. (2019).

$\{u_l(d_{ij}, w_{j,t-s_{ijt}}, w_{i,t})\}_{l=1}^L$. The total amount of pollution imported from cities outside of the buffer zone, \hat{p}_{it}^{far} , is the following:

$$\begin{aligned}
\hat{p}_{it}^{far} &= \sum_{j:d_{ij}>r} p_{j \rightarrow i, t}^+ = \sum_{j:d_{ij}>r} \max[\cos \Phi_{ji}, 0] \cdot p_{j,t-s_{ijt}} \cdot \sum_l^L \gamma_l u_l(d_{ij}, w_{j,t-s_{ijt}}, w_{i,t}) \\
&= \sum_l^L \gamma_l \sum_{j:d_{ij}>r} \max[\cos \Phi_{ji}, 0] \cdot p_{j,t-s_{ijt}} \cdot u_l(d_{ij}, w_{j,t-s_{ijt}}, w_{i,t}) \\
&= \sum_l^L \gamma_l Z_{it}^l
\end{aligned}$$

where $Z_{it}^l = \sum_{j:d_{ij}>r} \max[\cos \Phi_{ji}, 0] \cdot p_{j,t-s_{ijt}} \cdot u_l(d_{ij}, w_{j,t-s_{ijt}}, w_{i,t})$, $l = 1, \dots, L$ (6)

Our instruments for current day pollution p_{it} is the set of $\{Z_{it}^l\}_{l=1}^L$. These are valid instruments since they only depend on weather in city i at time t , which we control for in our regressions, and on pollution and weather variables in cities outside of the buffer zone at time $t - s_{ijt}$, which are uncorrelated with city i 's spending shocks by our identification assumption. Equation (6) makes it explicit that this strategy exploits a number of restrictions to construct powerful IVs. For example, if the prevailing wind conditions are such that it takes two days for pollution generated in city j to reach city i , we would expect $p_{j,t-2}$ instead of $p_{j,t}$ or $p_{j,t-1}$ to affect $p_{i,t}$. Our instrument Z_{it} is therefore a function of $p_{j,t-s_{ijt}}$, where s_{ijt} is the number of days it takes for pollution generated in city j to arrive in city i . As such, the calculation of Z_{it} properly dates the relevant pollution source in origin city $p_{j,t-s_{ijt}}$ and aggregates over all origin cities.

Set of IVs In the baseline specification, we use 15 second-order polynomial terms $\{u_l(\cdot)\}_{l=1}^{L=15}$ to flexibly approximate the decay function: 1) constant, the inverse distance, and origin city's weather (wind speed, precipitation, temperature) (5 terms); 2) the quadratic terms of the inverse distance and origin city's weather (4 terms); 3) the product of the inverse distance and the origin city's weather (3 terms); 4) the destination city's weather (wind speed, precipitation, temperature) (3 terms). Hence, we have 15 instruments $\{Z_{it}^l\}_{l=1}^{L=15}$ for current day pollution p_{it} .

As shown in Section 3.1, the flexible distributed lag model transforms many lagged pollution variables $\{p_{i,t-\tau}\}_\tau$ into a few $\{v_{\cdot,it}\}$'s, which are weighted sums of past pollution exposure with B-splines as weights (see equation (4) for a simple example). In our main specification that uses a cubic B-spline with three segments, there are six B-spline terms and hence six endogenous variables $\{v_{1,it}, \dots, v_{6,it}\}$. The instruments for these endogenous variables are constructed analogously, except that the lagged endogenous pollution variables are replaced with the corresponding lagged vector of exogenous IVs $\{Z_{i,t-\tau}^l\}_{l=1}^{L=15}$. There are fifteen IVs for each $v_{\cdot,it}$ and a total of 90 instruments.

Appendix B.2 provides more details.

Identification Assumptions Our approach that exploits $PM_{2.5}$'s region of influence is analogous to the source-receptor matrix constructed by the US EPA for air pollution prediction. The instruments we construct leverage variation in $PM_{2.5}$ in non-local sources, wind patterns and other meteorological conditions such as temperature and precipitation in both the source cities and the destination city, which have been shown to affect the long-range transport of $PM_{2.5}$. These instruments provide ample variation that allows us to simultaneously identify the short-term and medium-term impacts of pollution and quantify the time-path of these impacts. An alternative strategy involves using variation in local wind direction to estimate the health impacts of particulate matter pollution (Deryugina et al., 2019). Although changes in local wind direction are more plausibly exogenous and well suited towards identifying the short-run impact of pollution, they may lack enough variation to explain changes in both current and lagged pollution variables. As we illustrate in Section 4.2, IVs that only use variation in wind direction (interacted with region dummies) fail to pass the weak IV tests and lead to insignificant estimates, though the estimated impact of $PM_{2.5}$ on aggregate health spending is broadly similar to our baseline estimates.

Our identification assumption is that pollution shocks (e.g., economic activities) in regions outside of the buffer zone are uncorrelated with local shocks to spending.²² This assumption would be violated if spending shocks (e.g., high temperature that leads to more hospital visits as well as increased demand for electricity) in city i affect production activities in other cities (e.g., electricity generation) outside the buffer zone, which in turn affect the pollution level in city i . To the extent that economic shocks in city i affect production and hence pollution in other cities, this should induce correlation between the error term ε_{it} and *future* pollution levels rather than lagged pollution levels in other cities. In contrast, our instruments are weighted sums of *lagged* pollution levels in distant cities, where the weights are the inverse distance and meteorological conditions in both the source and receptor cities. In addition, averaging over the exogenous variation in wind speed and direction across a large number of source cities should reduce such correlations, if any.

To further address potential concerns on the validity of our IVs, we proceed in three ways. First, we show in section 4.2 that results are robust to different radii of the buffer-zone. Second, we construct an alternative set of IVs using the historical average (time-invariant) level of air pollution in source cities, rather than the observed lagged pollution that could be subject to regional economic spillovers. The within-city variation of these IVs is solely driven by wind patterns and other weather conditions rather than time-varying pollution levels in source cities, hence should not be correlated

²²Pollution imported from regions outside the buffer zone is assumed to be uncorrelated with the measurement errors in local pollution exposure. This is plausible because changes in measurement errors arise from within-city variation, while imported pollution is affected by economic activities at the source cities and weather conditions determining the diffusion of pollutants.

with unobserved economic shocks in the destination city. The results from this specification are similar to the benchmark estimates. Third, we include the average $PM_{2.5}$ in other cities outside of the buffer zone but within the same region as an additional regressor to control for regional spillovers in economic activities. This has little impact on the parameter estimates.

Finally, our identification strategy is different from the regression discontinuity (RD) approach based on the Huai River heating policy used in [Chen et al. \(2013\)](#) and [Ito and Zhang \(2018\)](#). The RD design exploits the long-term cross-sectional variation in pollution and is better suited to study long-term impacts, such as that on mortality. This study focuses on the short- and medium-term impacts and our IV approach is designed to leverage the data’s rich spatial and temporal variations.

4 Empirical Results

4.1 Impact of Pollution on Health Spending

We now describe our empirical analysis of air pollution’s effect on health spending. We use the logarithm number of transactions as the dependent variable rather than the value of transactions, following the literature that uses similar transaction-level purchase data ([Einav et al., 2014](#)). As explained in [Sec 2.3](#), the number of transactions is a good proxy for visits to healthcare facilities. In [Appendix C](#), we report results using the value of transactions as the dependent variable. They are similar in magnitude to those based on the number of transactions but less precise. This is partly because the distribution of healthcare spending is right-skewed with many large transactions (e.g., surgeries) that are unlikely caused by air pollution in the short run. While our baseline specification utilizes the total number of transactions as the dependent variable, the estimates are very similar if we instead use the number of transactions per capita, as discussed in [Section 4.2](#).

All regressions include city fixed effects to control for time-invariant unobservables, week-of-the-sample fixed effects to control for nationwide shocks, and city-specific time trend and city-specific seasonality (i.e., interactions of city fixed effects and quarterly dummies) to control for trends in bank card adoption and seasonal diseases. We also add fixed effects for state holidays, working weekend,²³ day of the week, as well as weather variables to control for their direct effects on spending. For example, people may reduce non-urgent hospital visits during holidays or on raining days. All standard errors are clustered at the city level.

First-Stage Results To address the issue of measurement errors and endogeneity, we instrument $PM_{2.5}$ using pollution imported from distant cities outside the buffer zone as discussed in

²³Weekends near multi-day holidays are usually swapped with weekdays next to the actual holidays to create a longer holiday. As a result, businesses and schools treat those weekends as *working weekends*.

Section 3.2. To assess the strength of instruments in first stage regressions, we follow the best practice as suggested in the weak IV literature. Since the Cragg-Donald F-statistic that assumes homoskedasticity tends to overstate the strength of instruments in non-homoskedastic settings, we report F-statistics that are robust to heteroskedasticity throughout this analysis. When there is one endogenous regressor, Andrews et al. (2019) recommend the effective F-statistic of Olea and Pflueger (2013), which is the appropriate statistic for formally testing weak identification in non-homoskedastic settings. We therefore report the effective F-statistic in all regressions with one endogenous variable.

The benchmark specification of our distributed lag model has six endogenous variables and a total of 90 instruments. To our knowledge, the literature on weak instruments has not yet developed formal methods for detecting weak identification in the presence of multiple endogenous regressors and non-homoskedastic errors. As such, we report the Kleibergen-Paap Wald rk F-statistic that is clustered at the city level in all regressions with multiple endogenous regressors.

Appendix Table C5 reports the first-stage result where we regress p_{it} on different sets of IVs. The signs for included IVs are expected: pollution is lower in holidays and decreases with local precipitation and wind speed. In Column (1), the only excluded instrument is a simple sum of $PM_{2.5}$ from distant cities traveling toward the destination city. Column (2) takes into account that $PM_{2.5}$ decays as it travels and uses the sum of $PM_{2.5}$ from distant cities weighted by the inverse distance and weather variables of the origin cities as excluded IVs. This corresponds to a linear decay function. The coefficient estimates suggest that both higher temperatures and greater precipitation in origin cities lead to a faster decay of $PM_{2.5}$. In addition, the further $PM_{2.5}$ has to travel, the more it decays. As a result, the distance weighted sum of $PM_{2.5}$ has a much higher predictive power of local pollution than a simple sum of pollution from origin cities.²⁴ Column (3) allows for a second-order polynomial decay function in the inverse distance and weather conditions in the origin cities, as well as weather conditions in the destination city, leading to a total of fifteen instruments as discussed in Section 3.2.2. The effective F-statistic is 161 and 112 in Columns (2) and (3), respectively. They exceed the critical value by a large margin and indicate a strong first stage. Our preferred specification is Column (3) which allows for a more flexible decay function of $PM_{2.5}$ than Column (2), though the estimated health impacts are similar with either four or fifteen IVs.

For our baseline specification, the robust F-statistic varies from 38 to 48 (Table 3), which suggests that weak identification is unlikely to be a concern in our setting. In comparison, if we were to assume homoskedastic errors, the Cragg-Donald F-statistic varies from 213 to 255 and far exceeds the critical value (which converges to 21 with a large number of IVs).²⁵

²⁴The raw correlation between local pollution and distance weighted pollution from origin cities is 0.21, while the raw correlation between local pollution and a simple sum of pollution from origin cities is close to 0.

²⁵We extend the analysis in Stock and Yogo (2005) by using the programming code in Skeels and Windmeijer (2018) to

Short-term Impacts Our empirical analysis begins with the current-day $PM_{2.5}$ as the only key variable of interest, i.e., $k = 0$ in Equation (1). The coefficient estimate on current-day $PM_{2.5}$, β_0 , captures the effect of both current-day and past pollution exposure, the latter of which are correlated with current-day pollution but omitted from the regression. As a result, β_0 is not the marginal impact of the current-day exposure on spending. Nevertheless, we can view the estimate as a short-term impact. Appendix Tables C6 and C7 report the OLS and IV estimates of the short-term impacts, respectively. A $10 \mu g/m^3$ increase in current-day $PM_{2.5}$ is associated with a 0.65% contemporaneous increase in transactions in the aggregate health care sector. The effect of air pollution on spending at Children’s hospitals is the largest among different health care categories and is nearly twice as large as that for the overall healthcare spending.

The IV estimates of the health impact of air pollution are several times as large as their OLS counterparts.²⁶ The large differences between OLS and 2SLS results are common in this literature. For example, Schlenker and Walker (2016) use runway congestion at airports on the US East Coast as exogenous variation to measure the contemporaneous health impact of air pollution exposure for communities near large airports in California. Their 2SLS estimates on the respiratory-related emergency room admissions due to CO and NO₂ exposure are 7-10 times as large as the OLS estimates. Ebenstein et al. (2017) use a regression discontinuity design based on the Huai River policy to examine the long-term impact of PM₁₀ on mortality. Their RD estimates are 2-3 times as large as the OLS estimates. Similarly, in Deryugina et al. (2019)’s study of the mortality and medical costs of PM_{2.5} in the US, the IV estimates are between 6 and 17 times larger than the corresponding OLS estimates. The smaller OLS estimates are consistent with the attenuation bias due to (classical) measurement errors in PM_{2.5}. The downward bias could also be driven by temporary local shocks, such as economic activities or major local events, that are positively correlated with air pollution but negatively correlated with healthcare spending (more outdoor activities and fewer hospital visits).

Longer-Term Impacts Exposure to $PM_{2.5}$ could have dynamic longer-term health impacts that are nonlinear. Directly including a large number of lagged $PM_{2.5}$ suffers from high autocorrelation. For example, Appendix Table B4 reports coefficient estimates from including up to 5-day lags of $PM_{2.5}$ in equation (1). Although these results indicate that the effect of $PM_{2.5}$ persists beyond one day, the high autocorrelation makes it difficult to tell apart the effect of $PM_{2.5}$ on consecutive days. As such, many coefficients are imprecise with oscillating signs. To address this issue, we employ the flexible distributed lag model discussed in Section 3.1 and allow pollution impacts to follow a

compute the critical value for the homoskedastic F-statistic with a large number of IVs. The critical value converges to 21 at the 5% significance level when there are 30 or more IVs with one endogenous variable.

²⁶The magnitude of the differences between OLS and IV estimates is theoretically possible, especially with a large number of fixed effects (Cameron and Trivedi, 2005).

smooth path.

Table 2 reports the cumulative effects of elevated PM_{2.5} concentration over different time periods, $\sum_{\tau=0}^k \beta_{\tau}$, from OLS regressions. Our benchmark specification incorporates daily pollution exposure for the past three months (90 lags). Effects beyond 90 days are modest and imprecisely estimated, thus excluded from the cumulative effects.²⁷ A one-day surge of 10 $\mu\text{g}/\text{m}^3$ in PM_{2.5} concentration increases today's transactions in all healthcare facilities by 0.03%. A medium-run (three-month) elevation of 10 $\mu\text{g}/\text{m}^3$ raises the number of transactions by 0.86%, eight times as large as the effect reported in Appendix Table C6 when only the contemporary PM_{2.5} concentration is included in the regression. There is a statistically significant negative impact on necessities and supermarket spending within two weeks, but not in the long run.

To deal with measurement errors and the endogeneity in current and lagged PM_{2.5}, we use IVs discussed in Section 3.2 and present results in Table 3. Several important findings emerge. First, the estimated 2SLS longer-term impacts of PM_{2.5} across all healthcare categories are positive and much larger than the short-term impact, consistent with the comparisons from the OLS estimates. Specifically, a 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} concentration over the past 90 days raises the number of transactions in the aggregate healthcare sector by 2.65%. Second, the impact on Children's hospitals is the largest and more than twice as large as the impact on aggregate healthcare spending, consistent with the fact that children are among the most vulnerable groups. Pharmacy is the second most responsive category among the four healthcare categories. When elevated air pollution aggravates symptoms for people with respiratory problems, they may go to pharmacies without visiting hospitals. Third, the effects on daily necessities and supermarket spending are all negative and appear to be short-lived.

To examine how the impact on spending changes over time, Figure 3 plots the path of the cumulative effects of past pollution exposure across different categories. Solid lines (and solid segments) indicate significance at the 5% level. The optimal number of lags should in theory differ across categories. For example, the effect of pollution on non-healthcare categories appears to be short-lived, while for children's hospitals it could last for more than three months. To keep the results comparable, we impose the same lag structure on all categories. Panel (a) depicts the cumulative effect for the aggregate health spending and spending in Children's hospitals. Consistent with the results in Table 3, the cumulative effect increases over the 90-day window and is stronger (in percentage terms) for spending in Children's hospitals. For spending in the aggregate health category, the cumulative effect appears to stabilize at three months, which is confirmed in the robustness analysis below.

In contrast, air pollution reduces spending on necessities and in supermarkets in the short term. The cumulative effect appears to peak at around two weeks, reduces in magnitude afterward and

²⁷We examine the robustness of our results to the choice of lags and B-spline segments in section 4.2.

becomes imprecise past one month. One explanation for the short-term reduction in non-health spending is the budget constraint: if consumers have to spend more on health care to mitigate the negative health impact of air pollution, they may have less to spend on non-health-related categories. However, the temporary reduction we find is inconsistent with the budget constraint hypothesis, since a sustained increase in healthcare spending would lead to a sustained reduction in necessities with a fixed budget. Instead, our results lend support to the hypothesis of avoidance behavior, whereby consumers postpone or reduce shopping trips to reduce pollution exposure in response to poor air quality. This is consistent with recent evidence in the literature (Mu and Zhang, 2016; Ito and Zhang, 2018; Sun et al., 2017).

Chinese consumers are increasingly aware of the air quality and its impact on health (Barwick et al., 2020) and PM_{2.5} readings are becoming readily accessible through cell phone apps and from government websites in recent years.²⁸ Given the importance of avoidance behavior, we provide additional evidence that individuals engage in avoidance in Appendix F by examining how expectations about future air pollution affect current healthcare and non-health spending.

Table 4 examines how the estimates of the short- and medium-run effects of PM_{2.5} on total health transactions depend on various controls included in regressions. City fixed effects control for baseline differences across cities where more polluted cities tend to have higher health spending. Week-of-the-sample fixed effects allow for nation-wide temporal variation in spending and air pollution. City-specific time trends capture the heterogeneous card adoption rates across cities, which are important given the rising card penetration in our sample periods. The coefficient estimates for short- and medium-run effects become intuitive after controlling for city fixed effects and week-of-the-sample fixed effects and remain stable once we include city-specific time trends. The holiday fixed effects, day-of-the-week fixed effects, and weather variables play a less important role in determining the magnitude of pollution's impact on health spending.

Our results suggest that a 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} would raise health-related transactions by 2.65% in the medium term. In terms of the value of transactions, the effect is 1.5% over the out-of-pocket expenses (Appendix Table C8). The estimates are somewhat less precise than those based on the number of transactions, driven by the larger noise inherent in the value of healthcare spending. The smaller impact on the transaction value makes intuitive sense in that the illnesses due to air pollution likely cost less to treat than other diseases on average.²⁹ In our analysis in Section 5.1, we use these estimates to bound the healthcare cost.

²⁸Hourly air pollution data in major Chinese cities have been published on the websites of the Ministry of Environment Protection and other non-government entities since 2013.

²⁹The average medical spending for in-patients in China is 6140 *yuan* (\$944) in 2013, while the average is 4109 *yuan* (\$632) for in-patients treated for respiratory diseases (National Health Commission, 2013). Relatedly, in the US, average expenditure per in-patient for respiratory diseases is \$1491, ranked 22 among 30 disease types according to Agency for Healthcare Research and Quality (2017).

4.2 Robustness Checks

We conduct an extensive set of robustness checks to examine whether the results discussed above depend on the empirical specification and choices of IVs.

Robustness to B-splines and buffer zone radii Table 5 reports the cumulative impact for overall healthcare spending under three different numbers of B-spline segments (1, 2, and 3) and five lags (30, 60, 90, 120 and 150). The estimates across different segments are similar. Our base specification uses three segments which provides a good balance between flexibility and precision. In terms of the number of lags, the cumulative impact is considerably smaller using 30-day lags but stabilizes after 90 lags.³⁰ We prefer 90-day lags because the estimated effects for lagged pollution are significant until around 90 days and start to lose significance for further distant lags.

Next, we carry out robustness checks with regard to the buffer zone radius in constructing IVs. We fix the radius at 150 km in the benchmark specification and assume that unobservables outside of the buffer zone of a city would not affect healthcare spending in that city. There is an inherent trade-off in the choice of the radius. On the one hand, the larger the buffer zone, the easier it is for the exclusion restriction to hold. On the other hand, the bigger the radius, the weaker the correlation between the predicted PM_{2.5} using non-local pollution and the observed PM_{2.5} in a given city. Table 6 presents several choices of the buffer zone from 100 km to 300 km with an increment of 50 km.³¹ The top panel reports the first-stage results. Generally, both the R² and the F-statistics decrease with the radius of the buffer zone, suggesting a weaker correlation between the IV and the endogenous variable as the buffer zone gets larger. The bottom panel shows the cumulative medium-term impact on healthcare spending, which varies from 2.42% to 2.88% across different radii when PM_{2.5} increases by 10 $\mu\text{g}/\text{m}^3$ over a 3-month period. Our preferred specification with a 150 km radius delivers an estimate that is in the middle of this range, though results are relatively stable across buffer zones within 300km.³² This is consistent with: (1) the long-range transportability of PM_{2.5}, and (2) non-local sources accounting for a significant share of local air pollution level (e.g., on average 35% for Beijing during 2005-2010 (Wang et al., 2015)).

Additional Controls The next set of robustness analysis includes various additional controls. Panel A in Table 7 reports estimates controlling for other pollutants including O₃, SO₂, NO₂ and CO. Emission sources such as electricity generation and transportation produce both particulate matters and other pollutants which also have harmful health impacts, though our IV strategy should

³⁰Cross-validation results indicate that models with long lags are preferred to the model with 30 days of lags.

³¹Williams and Phaneuf (2016) construct IVs for air pollution using pollutants 60 km (or 120 km) away without exploiting wind patterns.

³²300km is well within the travel distance of PM_{2.5}. For example, at a moderate speed of 15 miles per hour, it takes only one day for PM_{2.5} from 360 miles away to be transported to a destination.

address this to some extent since it leverages the long-range transport property of $PM_{2.5}$, which is different for other pollutants, especially O_3 and CO . Results with these four additional pollutants are similar to those in Table 3 for both healthcare and non-healthcare spending categories.³³

To address potential spillovers in regional economic activities, Panel B in Table 7 includes as a regressor the average level of $PM_{2.5}$ of nearby cities in the same region outside of the buffer zone.³⁴ If regional economic activities have systematic spillover beyond the buffer zone, one might be concerned with the exogeneity of our IVs: local unobservables could be correlated with economic activities in other cities, which are in turn correlated with pollution levels in other cities. Including $PM_{2.5}$ of nearby cities directly controls for economic activities in other cities and delivers similar results as those in the benchmark specifications.

Card penetration is growing rapidly over time during our sample period, which raises a concern that our results might be driven by uneven rates of card adoption across cities. The city-specific time trends in our baseline specification should capture this. In Panel C of Table 7, we introduce as additional controls the annual number of active cards and the annual number of point-of-service terminals in each city. Including these variables has little effect on the estimated impacts of $PM_{2.5}$.

In our baseline regressions, we control for local temperature, precipitation and wind speed. Table 8 explores the effect of including more flexible weather controls. Panel A is identical to the baseline specification. Panel B includes 2nd-order polynomial terms in weather variables. In Panel C, we create ten temperature bins similar to Deschênes and Greenstone (2011) and six bins for each of precipitation and average wind and include interaction terms of the bins. Panel D accommodates medium-term effects of local weather on health spending by additionally controlling for lagged weather variables up to 90 days.³⁵ Finally, in Panel E, we include lagged weather at both the origin and the destination cities, given that weather patterns are spatially correlated and weather conditions in a destination city may be correlated with lagged weather conditions in origin cities.³⁶ Overall the results appear reasonably robust though somewhat smaller with the inclusion of more flexible weather variables. The long-run effect of a $10\mu g/m^3$ increase in $PM_{2.5}$ on the number of all health transactions is 2.16% in Panel C with the most flexible weather controls, 1.98% in Panel D with lagged local weather variables, and 1.80% in Panel E with lagged weather in both the source

³³The correlation coefficient between daily levels of $PM_{2.5}$ and O_3 , SO_2 , NO_2 and CO is -0.13, 0.55, 0.66, 0.03, respectively. While we directly control for these pollutants in addition to $PM_{2.5}$ in our robustness checks, we do not address the potential endogeneity in these pollutants. Therefore, our estimated impact of $PM_{2.5}$ may reflect the impact of other pollutants. Disentangling the impacts of different pollutants is an important gap in the literature.

³⁴We follow the National Bureau of Statistics' classification that groups provinces into seven regions: East, North, Mid, South, Southwest, Northwest and Northeast.

³⁵We include the following lags for each weather variable at the destination city: the weather of day (t-1), (t-2), up to day (t-7); the average weather from day (t-8) to (t-14), the average weather from (t-15) to (t-28), the average weather from day (t-29) to (t-56) and the average weather from (t-57) to (t-90).

³⁶For each destination city, we construct an unweighted average of the lagged weather in each of the source cities, under the assumption that the weather travels to the destination city at the prevailing wind speed. We then include the same set of lags for source weather as we do for local weather. We thank an anonymous referee for this suggestion.

and destination cities, as opposed to 2.65% in the baseline specification.

Finally, Appendix Table C9 illustrates that adjusting for population size by using the number of transaction per capita as the dependent variable leads to similar results to those in the baseline specification.

Specifications Using Average Pollution Our flexible distributed lag model delivers a smooth marginal impact estimates of past pollution on current-day spending. The robustness checks presented in Table 5 illustrate that our results are not driven by the B-spline choices. To further address concerns over the functional form assumption, we estimate the more conventional specification in Equation (2) that uses the average pollution during a certain time window (e.g., current day + the past week) as the key variable of interest. While the specification may appear to be less restrictive than the flexible distributed lag model, it actually imposes a strong restriction that the marginal impact of lagged daily pollution on current-day spending is constant within the specified time window. The epidemiology literature has documented hump-shape (nonlinear) responses to air pollutants due to either normal physiological considerations or behavioral factors such as harvesting (Zanobetti et al., 2000, 2002; Schwartz, 2000). On the one hand, the impact of air pollution on respiratory system could take time to manifest. On the other hand, patients may postpone hospital visits until the symptoms are fully developed or cannot be treated by home remedies. Figure 3 corroborates these findings in the epidemiology literature and indicates that the marginal impact of lagged daily pollution on healthcare spending is unlikely to be constant.

Nonetheless, Appendix Table C10 presents the IV results for Equation (2) across several windows: current day, a week, a month, two-months, and 90 days. While the overall patterns are broadly consistent with those from the more flexible model in Table 3, the estimates from the restrictive model over current-day are substantially larger than those in our baseline specification across all health categories, while the estimates over the 90-day period are smaller. These differences are driven by two considerations. First, when pollution exhibits serial correlation, the estimated impact for the average pollution over a given window also captures the impact of pollution exposure in earlier periods. This is especially true for the current-day estimate. Second, restricting lagged pollution to have a constant impact on health spending could either overestimate or underestimate the true effect.

Choices of Instruments and Placebo Tests We have carried out a series of robustness analyses to examine the role of the instruments. To address the concern on the exogeneity of the instruments, we create an alternative set of IVs in Panel A of Table 9 using the historical average (time-invariant) level of air pollution in source cities, rather than the observed lagged pollution that could be subject to regional economic spillovers. The within-city variation in these IVs comes purely from changes

in wind and weather patterns, and thus the IVs should be uncorrelated with local unobserved economic shocks after controlling for city fixed effects. Though the IVs are not as strong as those in the main specification as indicated by a reduction in the F-statistic, the estimated effect of pollution on healthcare spending is similar to the benchmark specification.

A subset of the instruments depend on the destination city's weather. Panel B in Table 9 drops instruments that are functions of the destination city's weather so that none of the IVs uses information related to local conditions. The estimated aggregated health impact is 2.91%, slightly larger than our baseline result of 2.65%. In Panel C of Table 9, we drop the following large cities: Beijing, Shanghai, Guangzhou, Shenzhen, Wuhan, Chongqing, Chengdu, and Nanjing. Due to superior medical facilities in these large transportation hubs, these cities receive a large number of patients from other areas. If some out-of-town patients come from areas that export pollution to these major cities, this could lead to a correlation between the instruments and unobserved healthcare spending shocks. The estimated aggregated health impact is 2.25%, somewhat lower than our baseline result though the difference is insignificant statistically.

Appendix Table C11 conducts a placebo test that constructs IVs based on randomly generated wind direction and wind speed. To offer a useful benchmark, we first drop IVs that depend on temperature and precipitation and limit to a parsimonious set of IVs that are only interactions between wind speed and direction and source cities' pollution. As shown in Column (2), the estimated impact of $PM_{2.5}$ on health spending is comparable to our baseline specification, though the F-stats is lower indicating weaker IVs. In Column (3), we randomize wind direction and speed using random draws from their empirical distribution. The first-stage F-stat is merely 7.24. Moreover, the impact of air pollution is noisily estimated with wrong signs and statistically insignificant, as expected.

Alternative Identification Strategies An alternative identification strategy is to drop source cities' pollution and other meteorological conditions altogether and only uses wind direction in destination cities to instrument for changes in air pollution. Since the effect of wind direction on air pollution depend on geography, we interact wind direction with region dummies that are created by a K-means clustering algorithm to spatially classify cities.³⁷ This is essentially the empirical strategy used in [Deryugina et al. \(2019\)](#).

Appendix Table C12 compares our baseline results (Column 1) with results from these alternative IVs that are purely based on destination wind direction (Column 2). One challenge we encounter with wind-IVs is that they fail to pass the weak IV test. While wind directions are well-suited to identify short-term impacts as in [Deryugina et al. \(2019\)](#), they lack enough variation to explain changes in both current and lagged daily pollution in our context. Using the identification-robust confidence intervals proposed by [Andrews \(2018\)](#) that are valid under weak IVs, the impacts

³⁷Results are similar if use wind direction interacted with province fixed effects as IVs.

of pollution on healthcare spending are estimated with much less precision. The long-run effect of a $10\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ on the aggregate healthcare transactions is estimated to be an insignificant 1.48%, as opposed to 2.65% in the baseline.

Lastly, we also implement the RD design based on the differential heating policy across the Huai River. Following [Chen et al. \(2013\)](#); [Ebenstein et al. \(2017\)](#); [Ito and Zhang \(2018\)](#), we collapse the sample to a single cross-section which removes the high-resolution temporal variation. These papers examined (heavier) TSP or PM_{10} in earlier years. In our context, the Huai River policy turns out a weak IV for $\text{PM}_{2.5}$, likely a result of $\text{PM}_{2.5}$'s long-range transport property, the recent reform on the heating policy (e.g., pay for heating in the north), and other environmental and energy regulations (e.g., switching coal to natural gas for winter heating) in recent years.

Pollution Monitoring and Sample Cities The number of cities where air pollution was monitored has grown considerably over time as the Chinese government rolled out the nationwide pollution monitoring and public disclosure program from 2013 ([Barwick et al., 2020](#)). To check if the effects of air pollution are different for cities where monitoring began early and cities where monitoring began later, we carry out a robustness check that limits to the 159 cities where pollution has been monitored since 2013 (thus dropping cities where monitoring began in 2014 or afterwards). The results are shown in Appendix Table C13. The long-run effect of a $10\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ on the number of health transactions is 2.08% for cities with pollution monitoring since 2013, as opposed to 2.65% for the full sample. This reflects the fact that pollution monitoring began in larger, richer cities and was later extended to smaller cities with lower per capita income. As we discuss in the next section, the marginal effect of $\text{PM}_{2.5}$ is smaller for wealthier cities.

4.3 Nonlinearity and Heterogeneity

One concern regarding the external validity of the benefit-transfer approach is the potential nonlinearity of the dose-response function. The pollution level observed in developing countries such as China and India is far greater than the prevailing level studied in the literature. Linear projections in the benefit-transfer approach could either under- or over-estimate the health costs of air pollution in developing countries if the underlying effect is nonlinear ([Lelieveld et al., 2015](#); [World Bank, 2007](#)). Despite of its important implications, there is a lack of empirical evidence on the nonlinearity of the dose-response function ([Lelieveld and Poschl, 2017](#)). The rich spatial and temporal variation in our data allows us to examine the health impacts of $\text{PM}_{2.5}$ for a wide range of pollution levels.

To capture nonlinearity, we include the quadratic term of $\text{PM}_{2.5}$ in addition to its linear form.³⁸

³⁸To conserve the number of parameters, we use one-segment instead of three-segment B-splines, since cumulative effects are similar across different segments (section 4.2). Appendix Table D14 reports the coefficient estimates.

The top panel of Figure 4 plots the estimated surface of the marginal response for varying levels of $PM_{2.5}$ and along the time path for up to 100 days. For each value of $PM_{2.5}$, the slice of the surface along the p-axis is the estimated dynamic response. The surface is tilted upwards with a higher marginal response for a higher pollution level, indicating an increasing marginal impact of $PM_{2.5}$ on healthcare spending. In the bottom panel of Figure 4, we plot the cumulated marginal effect over three months ($\sum_{\tau} \beta_{\tau}$) against pollution level. The cumulative impact on healthcare spending increases in $PM_{2.5}$, but the overall nonlinearity of the health impact does not appear very pronounced. Based on this finding, we extrapolate our estimates across a wide range of pollution levels in evaluating the pollution's healthcare cost in China (Section 5).

Appendix Figure D8 examines the impact of air pollution across cities with different per capita income. In 2015, China's average annual disposable income per capita varied from 12,000 *yuan* to 53,000 *yuan* across cities, with an average of 25,530 *yuan*. Pollution's impact on healthcare spending appears to diminish monotonically within our range of disposable income: it is largest in poor cities and decreases with income. This may be driven by the limited avoidance behavior (e.g., use of air purifiers) among low income households and a lack of preventive healthcare in poor cities. While the differences across income levels could be meaningful economically, the evidence is suggestive given the statistical insignificance of income coefficients (Appendix Table D14).

We have also estimated heterogeneity across seasons and years (Appendix Table D15). Most of the heterogeneity coefficients are statistically insignificant, except for the coefficient on winter, suggesting that health spending is more responsive to pollution in winter than in other seasons. This is consistent with results on the nonlinearity analysis, as pollution peaks in winter and higher pollution invokes a larger marginal response.

5 Healthcare Cost of Air Pollution

In this section, we estimate the healthcare cost of $PM_{2.5}$ in China and compare it with the mortality cost estimated from the literature. It is important to note that the impact of particulate matter pollution on health spending will generally understate the welfare impact of morbidity. This is because in addition to increased healthcare costs, individuals who fall sick due to air pollution also suffer from reduced productivity (e.g. sick-days) and reduced quality of life. Moreover, individuals may engage in costly avoidance behavior in order to reduce exposure to air pollution as shown in Appendix F. Since avoidance behavior is a response to pollution (i.e., an outcome), rather than an unobserved confounding factor, the presence of avoidance behavior does not bias our estimates *per se*. It does, however, change the interpretation of the results. Our estimates provide the healthcare cost of pollution *conditional* on defensive behaviors undertaken by individuals, which is different from (and in general lower than) the morbidity cost of pollution in the absence of any avoidance

behavior.

5.1 Healthcare Cost

To better understand the magnitude of our estimates, we first benchmark our results with the findings in the related literature in Appendix Table E16. Our preferred specifications show that a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ would lead to a 2.65% increase in the number of health-related transactions (Table 3) and a 1.5% increase in the value of transactions (Table C8) in the long term. In a study on preventive expenditure, Mu and Zhang (2016) estimate that face mask purchases in China increase by 5.45% for a 10-point increase in Air Quality Index (AQI) and 7.06% for anti- $\text{PM}_{2.5}$ masks. Using the piecewise linear relationship between $\text{PM}_{2.5}$ and AQI, this means that exposure to $10 \mu\text{g}/\text{m}^3$ more $\text{PM}_{2.5}$ leads to a 3.6% to 7.3% increase in preventive spending.

Williams and Phaneuf (2016) use data in the U.S. and find that a one-standard-deviation ($3.78 \mu\text{g}/\text{m}^3$) change in $\text{PM}_{2.5}$ leads to 8.3% more spending on asthma and COPD, which is equivalent to a 22% increase for $10 \mu\text{g}/\text{m}^3$ more $\text{PM}_{2.5}$. According to China's National Health Commission (2013), spending on respiratory diseases accounts for 8% of total health expenditure in 2012. Assuming all additional spending induced by air pollution is for respiratory diseases, our estimates translate to a 33% increase in respiratory-related spending, about 50% larger than the estimate from Williams and Phaneuf (2016).

We now calculate the healthcare cost from elevated $\text{PM}_{2.5}$. Assuming that the health impact is the same for both bank-card and non-bank-card spending (see discussions in Appendix E), the 1.5% impact from a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ translates to 59.6 billion yuan (\$9.2 billion) based on the national healthcare spending in 2015.³⁹ Taking 2.65% as the upper bound, the healthcare cost from a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ would be 105.3 billion Yuan (\$16.2 billion). To be conservative, we use the lower bound estimate for our analysis below. These estimates can directly inform the overall economic cost of $\text{PM}_{2.5}$ and related policy discussions. For example, OECD (2016) estimates that $\text{PM}_{2.5}$ and ground level ozone are associated with a \$20 billion direct cost on health expenditures (due to morbidity) worldwide based on the benefit-transfer approach, with half of these costs coming from non-OECD countries. A simple linear interpolation based on our estimates implies that the elevated $\text{PM}_{2.5}$ ($56 \mu\text{g}/\text{m}^3$ on average) relative to WHO's recommended level of $10 \mu\text{g}/\text{m}^3$ leads to \$42 billion (with a 95% confidence interval of \$2.8 - \$81.8 billion) added healthcare spending each year in China alone.

Our analysis suggests that OECD (2016) underestimates the health cost from air pollution, potentially up to an order of magnitude for developing countries. This could be due to: (1) the downward bias from endogeneity in the dose-response function; (2) the inherent differences in

³⁹China's health expenditure exceeded four trillion *yuan* (\$615 billion) in 2015 (National Health Commission, 2016).

the dose-response function across countries; and (3) the monetization of the disease incidences. The discrepancy highlights the importance of empirical studies using data on health spending from developing countries.

The morbidity cost of air pollution includes both the direct healthcare cost and the value of lost time from the illnesses (such as hospital visits and sick days). Our database recorded 670 million health-related transactions in 2015, which accounted for 50% of private health spending. As such, our estimate implies 35.5 million additional trips to healthcare facilities from a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$. To monetize the lost time, we assume that each trip takes three hours and the value of time (VOT) is 100% of the hourly wage, which is an upper-end estimate of VOT in the literature (Small, 2012; Wolff, 2014). The total value of the lost time from additional trips to healthcare facilities amounts to 2.3 billion *yuan* in 2015, compared to 59.6 billion *yuan* in additional healthcare spending from a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$. This suggests that the direct healthcare cost is the dominant component of the overall morbidity cost.

5.2 Comparing Morbidity and Mortality Cost

The current literature on the burden of disease from air pollution is based primarily on mortality. A common perception is that relative to mortality, the morbidity cost is a minor component of the overall cost of pollution. To put our estimates on healthcare cost (the primary component of morbidity) into perspective, we calculate the mortality cost based on the empirical analysis of Ebenstein et al. (2017). Using detailed mortality data by gender, age cohort, and disease types in 161 representative counties across China, they estimate that a $10 \mu\text{g}/\text{m}^3$ increase of PM_{10} would increase the cardiorespiratory mortality rate by 8% on average and the impact varies across age cohorts but not across gender.

We take two steps to monetize the literature's mortality estimate. There are no national-level estimates on the Chinese population's VSL. Narain and Sall (2016) suggest a transfer elasticity (or income elasticity) of 1.2 for transferring the U.S. VSL estimate to a developing country. China's per capita income is about an eighth of that in the U.S. At the elasticity of 1.2, the VSL for the Chinese population is 9.27% of that for the U.S. population. Using Ashenfelter and Greenstone (2004)'s estimate of \$2.27 million (in 2015\$) for the U.S. population – adopted by Deschênes et al. (2017) to quantify the mortality cost of NO_x emissions reductions – the Chinese population's VSL is \$0.21 million in 2015.⁴⁰

Second, we use estimates in Murphy and Topel (2006) to adjust the VSL for each age group. The VSL is at the full value for people less than 40 years old but reduces to 40% of its full value

⁴⁰Hoffmann et al. (2017) use data from 3 major cities in China and find a VSL of \$0.615 million in 2016 dollars. The average income in these cities doubles China's national average income level. Using a transfer elasticity of 1.2, the VSL at the national level would be \$0.25 million, close to our estimate of \$0.21 million.

by age 65 and 15% by age 80. Similar to what [Deschênes et al. \(2017\)](#) find, this adjustment is important because the age group 65 and above accounts for less than 9% of the total population but nearly 75% of the changes in mortality from air quality improvement.

This back-of-the-envelope analysis implies that a $10 \mu\text{g}/\text{m}^3$ increase of $\text{PM}_{2.5}$ would generate a mortality impact of \$13.4 billion in 2015 in China (Appendix Table E17). In comparison, our conservative estimate of the healthcare cost is 59.6 billion yuan, or \$9.2 billion, which constitutes 69% of the mortality cost. The implied ratio of healthcare costs to mortality costs is similar to that from [Deschênes et al. \(2017\)](#) in the context of the NO_x emissions reduction in the U.S.. Both estimates are substantially higher than the 10% ratio used in [WHO \(2015\)](#) to interpolate air pollution's economic impact.

The mortality cost is sensitive to the assumed VSL. There is considerable heterogeneity across published estimates of the VSL ([Kniesner et al., 2012](#)), ranging from under \$2 million ([Alberini et al., 2004](#); [Ashenfelter and Greenstone, 2004](#)) to EPA's estimate of \$8.7 million. If we were to use EPA's estimate of \$8.7 million, the healthcare cost we estimate from $\text{PM}_{2.5}$ is still 18% of the mortality cost. The ratio of morbidity over mortality could be higher once other components of morbidity are factored in, such as reduced productivity and the disutility of falling sick.

6 Conclusion

WHO's global air pollution database shows that the world's most polluted cities in 2016 were all from developing countries such as China, India, Iran, Pakistan, Philippines, and Saudi Arabia. In addition, 98% of cities in low- and middle-income countries with more than 100,000 residents do not meet WHO air quality guidelines. However, past research from epidemiology and economics going back several decades has focused on the impacts of air pollution on human health, particularly mortality, in developed countries. This study provides the first comprehensive analysis on the direct healthcare cost of $\text{PM}_{2.5}$ in a developing country context based on the high-resolution data from the world's largest payment network.

To address potential endogeneity in the pollution exposure measure, we develop an air quality prediction model in the spirit of the US EPA's source-receptor matrix that allows us to isolate exogenous variations in local air quality using the spatial spillovers of $\text{PM}_{2.5}$. We propose a flexible distributed flexible distributed lag model to estimate the temporal effect on healthcare spending and use a data-driven method to construct powerful IVs. Our results suggest that a $10 \mu\text{g}/\text{m}^3$ decrease in $\text{PM}_{2.5}$ would lead to at least a \$9.2 billion reduction in healthcare spending annually, or 1.5% of China's national annual healthcare expenditure. The estimated healthcare cost exceeds two-thirds of the mortality cost based on the recent literature. China's elevated $\text{PM}_{2.5}$ level relative to the WHO's annual standards entails at least \$42 billion added healthcare expenditure in

2015. Together, these results indicate that the recent report by [OECD \(2016\)](#) may have significantly underestimated the worldwide impact of air pollution on health expenditure (\$10 billion for all non-OCED countries including China).

In estimating the healthcare cost of air pollution in China, our analysis offers an alternative approach to the commonly used benefit-transfer approach for developing countries. The air pollution level in urban centers in developing countries is often an order of magnitude higher than that observed in developed countries. As urbanization continues and development pressure rises, air pollution could get worse before it gets better. The aggregate impact of air pollution on economic growth, including factors such as human capital accumulation, productivity, talent loss due to migration, and foreign direct investment, is an interesting and important area for future research.

References

- Agency for Healthcare Research and Quality**, “Medical Expenditure Panel Survey. Mean expenditure per person with care by condition and event type, United States, 2017,” 2017. Generated interactively: Fri May 22 2020. https://www.meps.ahrq.gov/mepstrends/hc_cond_icd10/.
- Alberini, Anna, Maureen Cropper, Alan Krupnick, and Nathalie B. Simond**, “Does the value of a statistical life vary with age and health status? Evidence from the US and Canada,” *Journal of Environmental Economics and Management*, 2004, 48 (1), 769–792.
- Almon, Shirley**, “The distributed lag between capital appropriations and expenditures,” *Econometrica*, 1965, pp. 178–196.
- Anderson, Michael L**, “As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality,” *Journal of the European Economic Association*, 2020, 18 (4), 1886–1927.
- Andrews, Isaiah**, “Valid two-step identification-robust confidence sets for GMM,” *Review of Economics and Statistics*, 2018, 100 (2), 337–348.
- , **James H Stock, and Liyang Sun**, “Weak instruments in instrumental variables regression: Theory and practice,” *Annual Review of Economics*, 2019, 11, 727–753.
- Arceo, Eva, Rema Hanna, and Paulina Oliva**, “Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City,” *Economic Journal*, 2015, 126, 257–280.
- Ashenfelter, Orley and Michael Greenstone**, “Using Mandated Speed Limits to Measure the Value of a Statistical Life,” *Journal of Political Economy*, 2004, 112, S226–67.

- Bagnall, John, David Bounie, Kim P Huynh, Anneke Kosse, Tobias Schmidt, Scott D Schuh, and Helmut Stix**, “Consumer cash usage: A cross-country comparison with payment diary survey data,” *ECB Working Paper*, 2014.
- Barwick, Panle Jia, Shanjun Li, Ligu Lin, and Eric Zou**, “From Fog to Smog: the Value of Pollution Information,” 2020.
- Bayer, Patrick, Nate Keohane, and Christopher Timmins**, “Migration and Hedonic Valuation: The Case of Air Quality,” *Journal of Environmental Economics and Management*, 2009, 58, 1–14.
- Browne, Andrew**, “Chinese Doctors Tell Patients To Pay Upfront, or No Treatment,” *The Wall Street Journal*, December 2005.
- Brunekreef, Bert and Stephen T Holgate**, “Air Pollution and Health,” *Lancet*, 2002, 360, 1233–1242.
- Cameron, A. Colin and Pravin K. Trivedi**, *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.
- Central Bank of China**, “General Report on the Operations of Chinese Payment System, 2015 Q3,” Technical Report, Central Bank of China 2015.
- Chay, Kenneth Y. and Michael Greenstone**, “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession,” *Quarterly Journal of Economics*, 2003, 118, 1121–1167.
- Chen, Yuyu, Avraham Ebenstein, Michael Greenstone, and Hongbin Li**, “Evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River policy,” *Proceedings of the National Academy of Sciences*, 2013, 110, 12936–12941.
- Choi, Weng I., Honghao Shi, Ying Bian, and Hao Hu**, “Development of Commercial Health Insurance in China: A Systematic Literature Review,” *BioMed Research International*, 2018, 2018 (1), 1–18.
- Clay, Karen, Joshua Lewis, and Edson Severnini**, “Canary in a Coal Mine: Impact of Mid-20th Century Air Pollution on Infant Mortality and Property Values,” 2016. Working Paper.
- Corradi, Corrado**, “Smooth Distributed Lag Estimators and Smoothing Spline Functions in Hilbert Spaces,” *Journal of Econometrics*, 1977, 5 (1), 211–219.
- **and G. Gambetta**, “The Estimation of Distributed Lags by Spline Functions,” *Empirical Economics*, 1976, 1 (1), 41–51.
- Currie, Janet and Matthew Neidell**, “Air Pollution and Infant Health: What Can We Learn from

- California's Recent Experience," *Quarterly Journal of Economics*, 2005, 120, 1003–1030.
- **and W. Reed Walker**, "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal: Applied Economics*, 2011, 3, 65–90.
- Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif**, "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, December 2019, 109 (12), 4178–4219.
- Deschênes, Olivier and Michael Greenstone**, "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US," *American Economic Journal: Applied Economics*, October 2011, 3 (4), 152–185.
- , — , **and Joseph Shapiro**, "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program," *American Economic Review*, 2017, 107 (10), 2958–89.
- Díaz, RV and Elena Rosa Dominguez**, "Health risk by inhalation of PM2.5 in the metropolitan zone of the City of Mexico," *Ecotoxicology and Environmental Safety*, 2009, 72 (3), 866–871.
- Dockery, Douglas W.**, "Health Effects of Particulate Air Pollution," *Annals of Epidemiology*, 2009, 19, 257–263.
- Dolney, Timothy J. and Scott C. Sheridan**, "The relationship between extreme heat and ambulance response calls for the city of Toronto, Ontario, Canada," *Environmental Research*, 2006, 101 (1), 94–103.
- Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou**, "New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy," *Proceedings of the National Academy of Sciences*, 2017, 114, 10384–10389.
- Einav, Liran, Dan Knoepfle and Jonathan Levin, and Neel Sundaesan**, "Sales Taxes and Internet Commerce," *American Economic Review*, 2014, 104 (1), 1–26.
- EPA**, *The benefits and costs of the Clean Air Act from 1990 to 2020*, Washington, DC: United States Environmental Protection Agency, 2011.
- Greenstone, Michael and Rema Hanna**, "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India," *American Economic Review*, 2014, 104, 3038–3072.
- He, Guojun, Maoyong Fan, and Meigong Zhou**, "The Effect of Air Pollution on Mortality in China: Evidence from the 2008 Beijing Olympic Games," *Journal of Environmental Economics and Management*, 2016. forthcoming.
- Hoffmann, Sandra, Alan Krupnick, and Ping Qin**, "Building a set of internationally comparable

- value of statistical life studies: estimates of Chinese willingness to pay to reduce mortality risk,” *Journal of Benefit-Cost Analysis*, 2017, 8 (2), 251–289.
- Ito, Koichiro and Shuang Zhang**, “Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China,” *Journal of Political Economy*, 2018. forthcoming.
- Jha, Ashish**, “The People’s Hospital: Cost, transparency, and hospital care in the middle of China,” *Harvard T.H. Chan School of Public Health Blog*, January 2014.
- Joksić, Jasminka D, Milena Jovašević-Stojanović, Alena Bartonova, Mirjana Radenković, Karl-Espen Yttri, Snežana Matić-Besarabić, and Ljubisa Ignjatovic**, “Physical and chemical characterization of the particulate matter suspended in aerosols from the urban area of Belgrade,” *Journal of the serbian chemical society*, 2009, 74 (11), 1319–1333.
- Kapron, Zennon and Michelle Meertens**, “Social networks, e-commerce platforms, and the growth of digital payment eco-systems in China: What it means for other countries,” 2017. Better Than Cash Alliance Research Series, Case Study, 2017.
- Kniesner, Thomas J, W Kip Viscusi, Christopher Woock, and James P Ziliak**, “The value of a statistical life: Evidence from panel data,” *Review of Economics and Statistics*, 2012, 94 (1), 74–87.
- Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders**, “Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health,” *Review of Economics and Statistics*, 2015, 98, 350–366.
- Landrigan, Philip, Richard Fuller, Nereus J R Acosta, Olusoji Adeyi, Maureen Cropper, Alan Krupnick, Michael Greenstone, and et al.**, “The Lancet Commission on pollution and health,” *The Lancet*, 2018, 391 (10119), 462–512.
- Lelieveld, J., J. S. Evan, M. Fnais, D. Giannadaki, and A. Pozzer**, “The contribution of outdoor air pollution sources to premature mortality on a global scale,” *Nature*, 2015, 525, 367–371.
- Lelieveld, Jos and Ulrich Poschl**, “Chemists can help to solve the air-pollution health crisis,” *Nature*, 2017, 551, 291–293.
- Meng, Qingyue and Hongwei Yang**, *People’s Republic of China health system review*, WHO Press, 2015.
- Mu, Quan and Junjie Zhang**, “Air pollution and defensive expenditures: evidence from particulate-filtering face masks,” *Journal of Environmental Economics and Management*, 2016.
- Murphy, Kevin M. and Robert H. Topel**, “The Value of Health and Longevity,” *Journal of Political Economy*, 2006, 114 (5), 871–904.

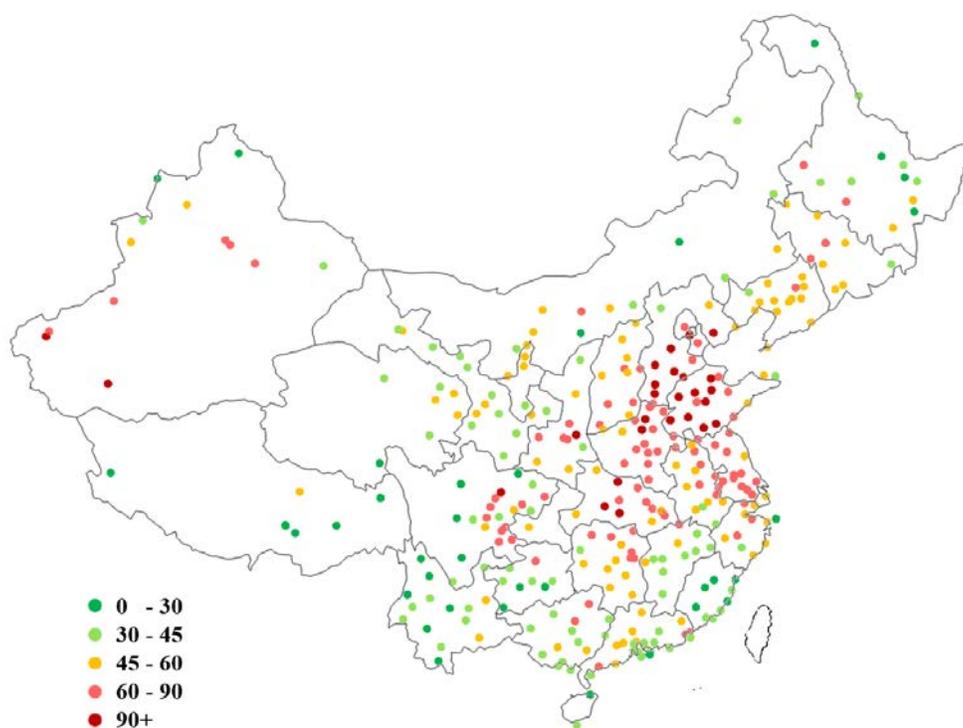
- Narain, Urvashi and Chris Sall**, *Methodology for Valuing the Health Impacts of Air Pollution: Discussion of Challenges and Proposed Solutions*, The World Bank, 2016.
- National Health Commission**, “China Public Health Statistical Yearbook (2012),” 2013.
- , “Statistic Bulletin on Health Development 2015,” 2016. http://www.gov.cn/shuju/2016-07/21/content_5093411.htm.
- Neidell, Matthew**, “Air Pollution, Health, and Social-economic Status: the Effect of Outdoor Air Quality on Childhood Asthma,” *Journal of Health Economics*, 2004, 23, 1209–1236.
- OECD**, “The Economic Consequences of Air Pollution,” 2016. <http://www.oecd.org/env/air-pollution-to-cause-6-9-million-premature-deaths-and-cost-1-gdp-by-2060.htm>.
- Olea, José Luis Montiel and Carolin Pflueger**, “A robust test for weak instruments,” *Journal of Business & Economic Statistics*, 2013, 31 (3), 358–369.
- Poirier, Dale J.**, “On the Use Of Bilinear Splines in Economics,” *Journal of Econometrics*, 1975, 3 (1), 23–34.
- Pope, C. Arden and Douglas W. Dockery**, “Health Effects of Fine Particular Air Pollution: Lines that Connect,” *Journal of the Air and Waste Management Association*, 2012, 56, 709–742.
- Pope, CA III**, “Respiratory disease associated with community air pollution and a steel mill, Utah Valley,” *American Journal of Public Health*, 1989, 79 (5), 623–628.
- Riva, DR_, CB Magalhaes, A_A Lopes, T Lancas, T Mauad, O Malm, SS Valenca, PH Saldiva, DS Faffe, and WA Zin**, “Low dose of fine particulate matter (PM_{2.5}) can induce acute oxidative stress, inflammation and pulmonary impairment in healthy mice,” *Inhalation toxicology*, 2011, 23 (5), 257–267.
- Schlenker, Wolfram and W. Reed Walker**, “Airports, Air Pollution, and Contemporaneous Health,” *Review of Economic Studies*, 2016, 83, 768–809.
- Schwartz, Joel**, “The distributed lag between air pollution and daily deaths,” *Epidemiology*, 2000, 11 (3), 320–326.
- Seibert, P and A Frank**, “Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode,” *Atmos. Chem. Phys. Discuss*, 2003, 3, 4515–4548.
- Skeels, Christopher and Frank Windmeijer**, “On the Stock–Yogo Tables,” *Econometrics*, 2018, 6 (4), 44.
- Small, Kenneth A**, “Valuation of travel time,” *Economics of transportation*, 2012, 1 (1), 2–14.
- Stock, James H and Motohiro Yogo**, “Testing for weak instruments in Linear Iv regression,” in

- “Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,” Cambridge University Press, 2005, pp. 80–108.
- Straney, L., J. Finn, M. Dennekamp, A. Bremner, A. Tonkin, and I. Jacobs**, “Evaluating the impact of air pollution on the incidence of out-of-hospital cardiac arrest in the Perth Metropolitan Region: 2000-2010,” *Journal of Epidemiology and Community Health*, 2014, 68 (1), 6–12.
- Sun, Cong, Siqi Zheng, and Matthew E. Kahn**, “Self-protection investment exacerbates air pollution exposure inequality in urban China,” *Ecological Economics*, 2017, 131, 468–474.
- Tai, APK, LJ Mickley, DJ Jacob, EM Leibensperger, L Zhang, JA Fisher, and HOT Pye**, “Meteorological modes of variability for fine particulate matter (PM. sub. 2.5) air quality in the United States: implications for PM. sub. 2.5 sensitivity to climate change,” *Atmospheric Chemistry and Physics*, 2012, 12 (6), 3131–3131.
- Wang, Jianhua and Susumu Ogawa**, “Effects of meteorological conditions on PM2. 5 concentrations in Nagasaki, Japan,” *International journal of environmental research and public health*, 2015, 12 (8), 9089–9101.
- Wang, Lili, Zirui Liu, Yang Sun, Dongsheng Ji, and Yuesi Wang**, “Long-range transport and regional sources of PM2.5 in Beijing based on long-term observations from 2005 to 2010,” *Atmospheric Research*, 2015, 157, 37–48.
- Wang, Weilin, Suli Zhao, Limin Jiao, Michael Taylor, Boen Zhang, Gang Xu, and Haobo Hou**, “Estimation of PM2. 5 Concentrations in China using a spatial back propagation neural network,” *Scientific reports*, 2019, 9 (1), 1–10.
- Wang, Xiaoyan, Robert E Dickinson, Liangyuan Su, Chunlüe Zhou, and Kaicun Wang**, “PM2.5 pollution in China and how it has been exacerbated by terrain and meteorological conditions,” *Bulletin of the American Meteorological Society*, 2018, 99 (1), 105–119.
- Wang, Xin, Xizhuo Sun, Stephen Birch, Fangfang Gong, Pim Valentijn, Lijin Chen, Yong Zhang, Yixiang Huang, and Hongwei Yang**, “People-centred integrated care in urban China,” *Bulletin of the World Health Organization*, 2018, 96 (12), 843–852.
- WHO**, *Economic cost of the health impact of air pollution in Europe: Clean air, health and wealth*, Copenhagen: WHO Regional Office for Europe, 2015.
- Williams, Austin M. and Daniel J. Phaneuf**, “The Impact of Air Pollution on Medical Expenditures: Evidence from Spending on Chronic Respiratory Conditions,” 2016. Working Paper.
- Wolff, Hendrik**, “Value of time: Speeding behavior and gasoline prices,” *Journal of Environmental Economics and Management*, 2014, 67 (1), 71–88.

- World Bank**, *Cost of Pollution in China: Economic Estimates of Physical Damages*, The World Bank, 2007.
- Yang, Changyuan, Ailan Chen, Renjie Chen, Yongqing Qi, Jianjun Ye, Shuangming Li, Wanglin Li, Zijing Liang, Qing Liang, Duanqiang Guo, Haidong Kan, and Xinyu Chen**, “Acute effect of ambient air pollution on heart failure in Guangzhou, China,” *International Journal of Cardiology*, 2014, *177* (2), 436–441.
- Yassin, Mohamed F, Bothaina EY AlThaqeb, and Eman AE Al-Mutiri**, “Assessment of indoor PM_{2.5} in different residential environments,” *Atmospheric environment*, 2012, *56*, 65–68.
- Yu, Hao**, “Universal health insurance coverage for 1.3 billion people: What accounts for China’s success?,” *Health Policy*, 2015, *119* (1), 1145–1152.
- Zanobetti, A, MP Wand, J Schwartz, and LM Ryan**, “Generalized additive distributed lag models: quantifying mortality displacement,” *Biostatistics*, 2000, *1* (3), 279–292.
- Zanobetti, Antonella, Joel Schwartz, Evi Samoli, Alexandros Gryparis, Giota Touloumi, Richard Atkinson, Alain Le Tertre, Janos Bobros, Martin Celko, Ayana Goren et al.**, “The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement,” *Epidemiology*, 2002, pp. 87–93.
- Zhang, Lin, Licheng Liu, Yuanhong Zhao, Sunling Gong, and Xiaoye Zhang**, “Source attribution of particulate matter pollution over North China with the adjoint method,” *Environmental Research Letters*, 2015, *10* (8).
- Zhong, Nan, Jing Cao, and Yuzhu Wang**, “Traffic Congestion, Ambient Air Pollution, and Health: Evidence from Driving Restrictions in Beijing,” *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (3), 821–856.
- Zhou, Xuehua, Zhaoyu Cao, Yujie Ma, Linpeng Wang, Ruidong Wu, and Wenxing Wang**, “Concentrations, correlations and chemical species of PM_{2.5}/PM₁₀ based on published data in China: potential implications for the revised particulate standard,” *Chemosphere*, 2016, *144*, 518–526.
- Zou, Eric**, “Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality,” 2018. Working Paper.

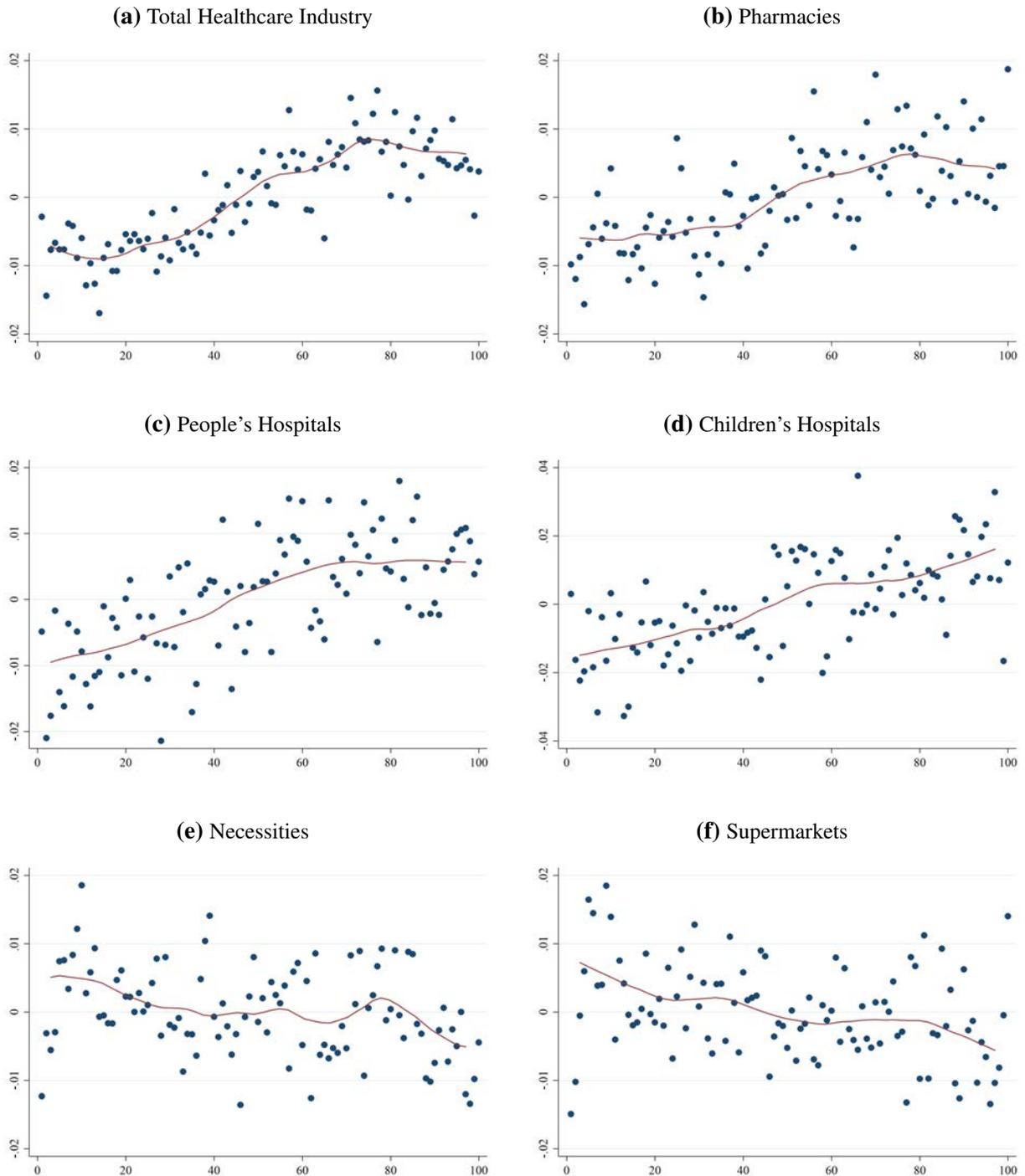
Figure 1: Three-Year Average PM_{2.5} Concentration

Jan. 2013 - Dec. 2015, $\mu\text{g}/\text{m}^3$



Notes: Each dot represents a city. There are 329 cities in total.

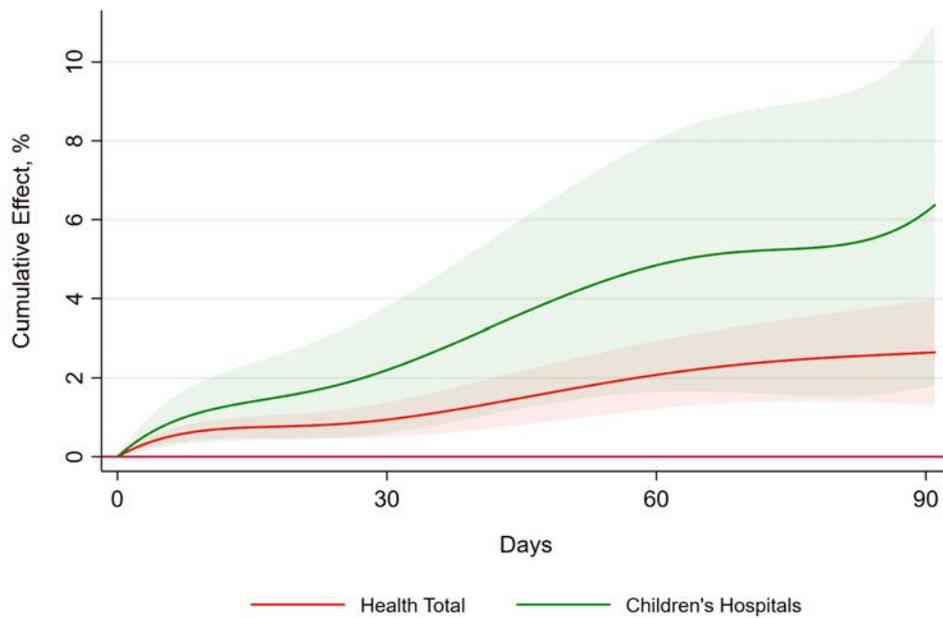
Figure 2: Residuals of Log Number of Transactions v. Residual of $PM_{2.5}$ Concentration (Binned)



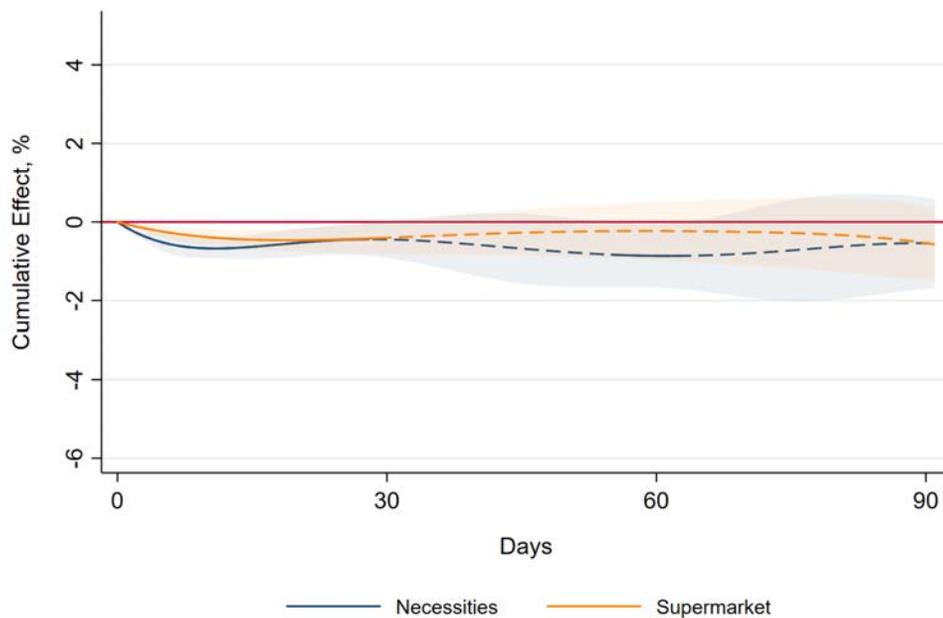
Notes: Each dot denotes the in-group average residuals, partialing out city FEs, weekly FEs, city-specific time trends, city-specific seasonality, day-of-week FEs, dummies for holidays and working weekends, and weather controls (temperature, precipitation, wind speed). Groups are binned by percentiles of $PM_{2.5}$, which is depicted by the x-axis.

Figure 3: Impact of Air Pollution on Number of Transactions from IV Regressions with 90 Lags

(a) Health-related Consumption

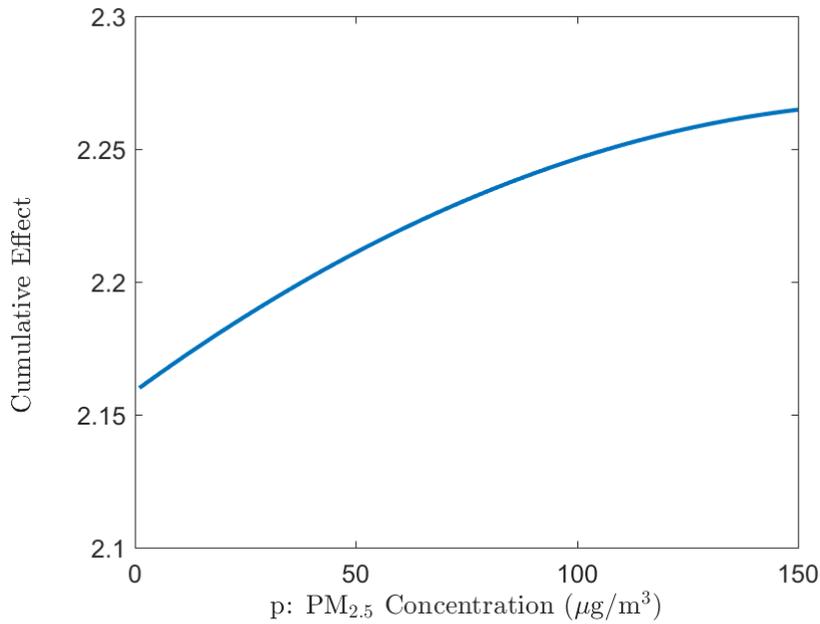
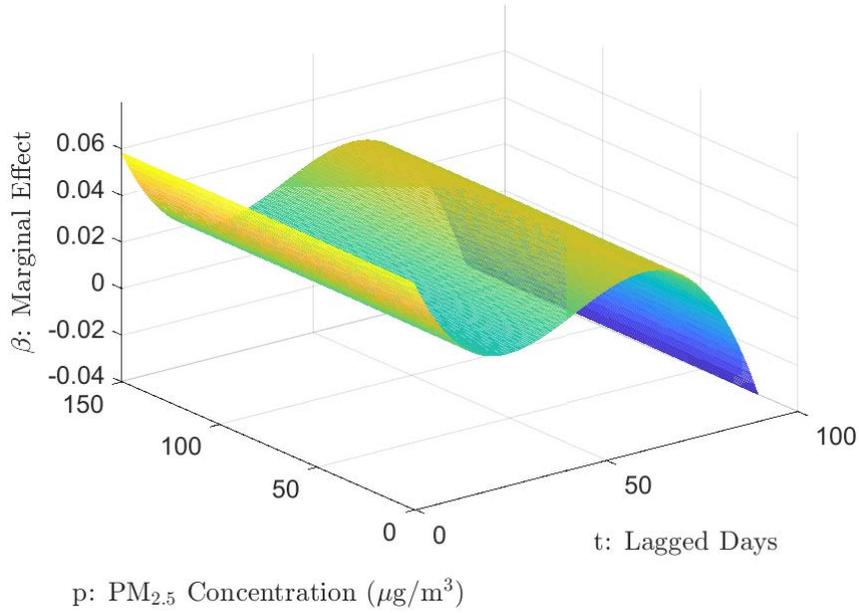


(b) Non-Health Consumption



Notes: the figure plots $\sum_{\tau=0}^k \beta_{\tau}$, the percentage change in the number of transactions for a given consumption category as a result of a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration over the past k days as indicated by the x-axis. On the x-axis, 0 refers to the current day, 30 refers to the past 30 days, etc. For example, a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ concentration over the past 28 days leads to 2.12% more transactions in Children's hospitals but 0.41% fewer transactions in supermarkets. Solid lines (and solid segments) indicate significance at the 5% level. Dashed lines indicate that the impact is statistically insignificant at the 5% level. Shaded areas are 95% confidence intervals.

Figure 4: Nonlinear Impacts of Air Pollution



Notes: the top panel plots β_τ , the (marginal) percentage change in today's healthcare transactions for a temporary 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} on day $t - \tau$ at a specific level of pollution concentration. On the τ -axis, 0 refers to day t , 50 refers to day $t - 50$, etc. The bottom panel plots $\sum_{\tau=0}^{90} \beta_\tau$, the percentage change in today's total healthcare transactions as a result of a 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} over the past 90 days, at different pollution levels as denoted by the x-axis. For example, a 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} over the past 90 days raises total healthcare transactions by 2.21% when the PM_{2.5} concentration is at 50 $\mu\text{g}/\text{m}^3$. Based on parameter estimates reported in Appendix Table D14.

Table 1: Summary Statistics

	Mean	Std. Dev.	Min.	Max.	N
Pollution					
PM _{2.5} Concentration, $\mu g/m^3$	56.3	46.4	0	985.2	198,246
Number of Transactions, Daily					
Healthcare Industry, Total	7,229.2	21,308.6	0	330,974	211,318
All Hospitals	4,122.7	14,503.9	0	237,525	210,539
People's Hospitals	1,060.6	2,800.4	0	40,332	203,407
Children's Hospitals	464.7	1,290.5	0	18,227	158,637
Pharmacies	2,245.3	7,063.3	0	96,336	210,001
Non-health Spending, <i>from 1% card sample</i>					
Daily Necessities	233.3	628.6	0	10,865	211,318
Supermarkets	393.4	990.3	0	15,224	210,493
Total Value of Transactions, Daily, thousand yuan					
Healthcare Industry, Total	6,701.8	17,818.9	0	301,108.7	211,318
All Hospitals	5,556.5	15,066.8	0	275,883.0	210,539
People's Hospitals	1,588.1	3,401.2	0	56,856.9	203,407
Children's Hospitals	363.9	843.3	0	10,324.3	158,637
Pharmacies	407.4	1,109.5	0	16,735.1	210,001
Non-health Spending, <i>from 1% card sample</i>					
Daily Necessities	236.9	551.3	0	9,532.4	211,318
Supermarkets	232.8	643.4	0	14,404.7	210,493
Weather					
Mean Temperature, $^{\circ}F$	60.1	18.9	-27.5	101.6	211,317
Precipitation, <i>inch</i>	0.1	0.4	0	15.6	211,318
Mean Wind Speed, <i>mph</i>	5.5	3.1	0	48.7	211,296
Wind Direction, <i>navigational bearing</i>	-	-	0	360	211,263

Notes: Data sources include China's Ministry of Environmental Protection, UnionPay, Integrated Surface Database (ISD), and Global Surface Summary of the Day (GSOD) Database. Data for health spending are from the full sample of bank cards. Data for non-health spending are based on a randomly selected 1% of bank cards. Children's hospital category has fewer observations because some small cities do not have a Children's hospital. UnionPay's data quality control process treats certain transactions as fraudulent, which leads to missing data in a few cases. The arithmetic mean and standard deviation of wind directions do not have statistical meaning and are left out in the table.

Table 2: Cumulative Effect of Pollution, OLS with 90 Lags

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current Day	0.03*** (0.01)	0.04*** (0.01)	0.05*** (0.01)	0.04*** (0.01)	0.06*** (0.02)	-0.03*** (0.01)	-0.02** (0.01)
Current + Past 3d	0.12*** (0.03)	0.11*** (0.03)	0.18*** (0.04)	0.13*** (0.04)	0.19** (0.08)	-0.11*** (0.03)	-0.07** (0.03)
Current + Past 7d	0.19*** (0.05)	0.16*** (0.06)	0.32*** (0.07)	0.21*** (0.06)	0.25* (0.15)	-0.16*** (0.05)	-0.11*** (0.04)
Current + Past 14d	0.25*** (0.08)	0.16 (0.10)	0.49*** (0.10)	0.30*** (0.08)	0.20 (0.28)	-0.16** (0.07)	-0.13** (0.06)
Current + Past 28d	0.38*** (0.13)	0.18 (0.15)	0.80*** (0.16)	0.39*** (0.14)	0.12 (0.50)	-0.15 (0.12)	-0.09 (0.11)
Current + Past 56d	0.66*** (0.19)	0.27 (0.20)	1.42*** (0.29)	0.47** (0.24)	0.57 (0.74)	-0.27 (0.21)	0.03 (0.18)
Current + All Lags	0.86*** (0.27)	0.34 (0.28)	1.81*** (0.42)	0.59* (0.36)	0.38 (1.14)	-0.08 (0.27)	0.02 (0.21)
N	141,794	141,657	141,567	137,853	110,259	141,770	141,652

Notes: The dependent variable is log(number of transactions) for a given consumption category in city i on day t . Column (1) includes all healthcare facilities. Columns (2)-(5) include all hospitals, pharmacies, people's hospitals, and children's hospitals, respectively. Columns (6)-(7) include necessities following United Nations' COICOP classification and supermarkets, respectively. Each row reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the corresponding period, $\sum_{\tau=0}^k \beta_{\tau}$, estimated using the OLS version of the flexible distributed lag model with 90 lags. For example, the third row reports the cumulative effect of a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ in the current day and the past week. The controls are city FEs, week FEs, city-specific time trends, city-specific seasonality, day-of-week FEs, dummies for holidays and working weekends, and weather controls (temperature, precipitation, wind speed). Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 3: Cumulative Effect of Pollution, IV with 90 Lags

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current Day	0.12*** (0.02)	0.12*** (0.03)	0.07* (0.04)	0.14*** (0.04)	0.19*** (0.07)	-0.14*** (0.03)	-0.06*** (0.02)
Current + Past 3d	0.40*** (0.07)	0.40*** (0.08)	0.23* (0.12)	0.47*** (0.13)	0.65*** (0.23)	-0.45*** (0.09)	-0.21*** (0.07)
Current + Past 7d	0.61*** (0.10)	0.62*** (0.12)	0.39** (0.18)	0.75*** (0.19)	1.04*** (0.36)	-0.64*** (0.13)	-0.34*** (0.10)
Current + Past 14d	0.74*** (0.14)	0.75*** (0.16)	0.57*** (0.21)	0.97*** (0.22)	1.40*** (0.50)	-0.63*** (0.16)	-0.45*** (0.12)
Current + Past 28d	0.91*** (0.22)	0.90*** (0.25)	0.99*** (0.30)	1.24*** (0.27)	2.12*** (0.79)	-0.44* (0.23)	-0.41** (0.21)
Current + Past 56d	1.97*** (0.42)	1.71*** (0.47)	2.31*** (0.54)	2.01*** (0.46)	4.65*** (1.56)	-0.85** (0.41)	-0.23 (0.36)
Current + All Lags	2.65*** (0.68)	2.18*** (0.71)	2.80*** (0.89)	2.13*** (0.75)	6.37*** (2.33)	-0.55 (0.58)	-0.57 (0.47)
N	141,794	141,657	141,567	137,853	110,259	141,770	141,652
First-stage F	38.35	38.36	38.37	39.69	47.79	38.29	38.29

Notes: The dependent variable is log(number of transactions) for a given consumption category in city i on day t . Column (1) includes all healthcare facilities. Columns (2)-(5) include all hospitals, pharmacies, people's hospitals, and children's hospitals, respectively. Columns (6)-(7) include necessities following United Nations' COICOP classification and supermarkets, respectively. Each row reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the corresponding period, $\sum_{\tau=0}^k \beta_{\tau}$, estimated via the IV version of the flexible distributed lag model with 90 lags. Same controls as in Table 2. The IVs are interactions of pollution transported from distant source cities (150km away) and meteorological conditions in the source and destination cities as defined in Equation (6) and Section 3.2.2. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table 4: IV Cumulative Effect of Pollution on Health Transactions: Different Fixed Effects

	Total Number of Healthcare-related Transactions							
	No Controls	City FE	+ Week FE	+ City×Trend	+ Holiday	+ Day-of-week	+ Weather	+ City×Season
Current Day	0.11 (0.31)	0.34*** (0.03)	0.05* (0.03)	0.12*** (0.02)	0.14*** (0.02)	0.15*** (0.02)	0.12*** (0.02)	0.12*** (0.02)
Current + Past 3d	0.61 (0.99)	1.00*** (0.10)	0.18* (0.11)	0.39*** (0.07)	0.47*** (0.07)	0.48*** (0.07)	0.39*** (0.07)	0.40*** (0.07)
Current + Past 7d	1.57 (1.48)	1.26*** (0.14)	0.29* (0.16)	0.61*** (0.11)	0.72*** (0.11)	0.73*** (0.11)	0.61*** (0.10)	0.61*** (0.10)
Current + Past 14d	3.50** (1.74)	0.73*** (0.15)	0.39* (0.20)	0.76*** (0.13)	0.87*** (0.13)	0.87*** (0.13)	0.75*** (0.12)	0.74*** (0.14)
Current + Past 28d	6.12*** (1.87)	-0.93*** (0.25)	0.56* (0.32)	0.96*** (0.20)	1.04*** (0.20)	1.03*** (0.20)	0.92*** (0.19)	0.91*** (0.22)
Current + Past 56d	6.87*** (2.03)	-3.27*** (0.31)	1.15* (0.62)	1.80*** (0.38)	1.88*** (0.38)	1.86*** (0.38)	1.70*** (0.36)	1.97*** (0.42)
Current + All Lags	11.43*** (3.15)	-6.33*** (0.46)	1.15 (1.06)	2.45*** (0.61)	2.53*** (0.61)	2.50*** (0.61)	2.28*** (0.59)	2.65*** (0.68)
N	146,427	146,427	146,427	146,427	146,427	146,427	141,794	141,794
First-stage F	50.50	55.15	29.71	38.30	38.35	38.34	38.39	38.35
City FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE			Yes	Yes	Yes	Yes	Yes	Yes
City×Trend				Yes	Yes	Yes	Yes	Yes
Holiday					Yes	Yes	Yes	Yes
Day of the Week						Yes	Yes	Yes
Weather							Yes	Yes
City×Season FE								Yes

Notes: The dependent variable is log(total number of healthcare transactions) in city i on day t . Each row reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the corresponding period, $\sum_{\tau=0}^k \beta_{\tau}$, estimated via the IV version of the flexible distributed lag model with 90 lags. Same IVs as in Table 3. Each column progressively adds more controls, as denoted in the bottom panel. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table 5: IV Cumulative Effects of Pollution: Different Number of Lags and Segments

Number of Segments z	Number of Lags k				
	30 days	60 days	90 days	120 days	150 days
1	1.18*** (0.25)	2.12*** (0.51)	2.42*** (0.69)	2.60*** (0.98)	2.58* (1.48)
2	1.41*** (0.25)	2.26*** (0.52)	2.67*** (0.69)	2.80*** (0.95)	2.62* (1.43)
3	1.28*** (0.25)	2.16*** (0.49)	2.65*** (0.68)	2.74*** (0.93)	2.41* (1.40)

Notes: The dependent variable is $\log(\text{total number of healthcare transactions})$ in city i on day t . Each cell reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the period as indicated by the column heading, estimated via the IV version of the flexible distributed lag model that uses the number of B-spline segments as indicated by the row heading. For example, the cell in the first column and first row reports the percentage change in all health-related transactions in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 30 days, estimated via the IV version of the flexible distributed lag model with one B-spline segment. Same IVs and controls as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table 6: IV Cumulative Effects of Pollution: Different Buffer Zone Radii

	Radius for the Buffer Zone				
	100 km	150 km	200 km	250 km	300 km
<i>First Stage Regression</i>					
R^2	0.486	0.474	0.467	0.464	0.462
First-stage F	46.69	38.35	34.14	35.36	35.33
<i>IV Regression</i>					
Current + All Lags	2.42*** (0.60)	2.65*** (0.68)	2.86*** (0.71)	2.86*** (0.72)	2.88*** (0.70)

Notes: The dependent variable is $\log(\text{total number of healthcare transactions})$ in city i on day t . The number of obs is 141,794. Each column uses the buffer zone radius as indicated by the column heading in constructing the instruments. The top panel reports the first-stage results and the bottom panel reports the percentage change in all health-related transactions in response to a medium-run $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 90 days. Same set of IVs (except for the buffer zone radius) and controls as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table 7: IV Cumulative Effects of Pollution: Additional Controls

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A: Controlling for other pollutants</i>							
Current + All Lags	2.55*** (0.69)	2.07*** (0.72)	2.73*** (0.91)	2.01*** (0.76)	6.21*** (2.34)	-0.55 (0.58)	-0.69 (0.46)
First-stage F	39.76	39.85	39.75	41.61	50.98	39.71	39.71
<i>Panel B: Controlling for economic spillover</i>							
Current + All Lags	2.62*** (0.68)	2.15*** (0.72)	2.76*** (0.89)	2.12*** (0.76)	6.37*** (2.34)	-0.56 (0.59)	-0.56 (0.47)
First-stage F	37.53	37.49	37.54	38.91	45.28	37.49	37.48
<i>Panel C: Controlling for card adoption</i>							
Current + All Lags	2.60*** (0.69)	2.14*** (0.73)	2.75*** (0.90)	2.10*** (0.74)	6.31*** (2.37)	-0.56 (0.56)	-0.59 (0.46)
First-stage F	38.01	38.02	38.03	39.33	47.38	37.95	37.95

Notes: The dependent variable is log(number of transactions) for a given consumption category in city i on day t . Each cell reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 90 days, $\sum_{\tau=0}^{90} \beta_{\tau}$, estimated via the IV version of the flexible distributed lag model. Same IVs as in Table 3. In addition to controls in Table 3, Panel A includes the daily average concentration levels of O_3 , SO_2 , NO_2 and CO , Panel B includes the average pollution level in cities outside of the buffer zone but within the same region, and Panel C includes log(number of cards used) and log(number of POS terminals) at the city-year level. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table 8: IV Cumulative Effects of Pollution: Flexible Weather Controls

	Health-related Consumption				Non-health Spending		
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
<i>Panel A: Baseline regression with linear terms of local weather</i>							
Current + All Lags	2.65*** (0.68)	2.18*** (0.71)	2.80*** (0.89)	2.13*** (0.75)	6.37*** (2.33)	-0.24 (0.44)	-0.55 (0.58)
First-stage F	38.35	38.36	38.37	39.69	47.79	38.35	38.29
<i>Panel B: Include 2nd-order polynomials of local weather</i>							
Current + All Lags	2.39*** (0.67)	1.85*** (0.69)	2.66*** (0.90)	1.94*** (0.74)	5.54** (2.34)	-0.41 (0.57)	-0.74* (0.45)
First-stage F	37.28	37.25	37.30	38.11	46.46	37.21	37.20
<i>Panel C: Include bins of local weather variables</i>							
Current + All Lags	2.16*** (0.68)	1.58** (0.70)	2.54*** (0.91)	1.63** (0.72)	5.49** (2.37)	-0.37 (0.58)	-0.74 (0.46)
First-stage F	36.25	36.21	36.23	36.89	46.90	36.19	36.22
<i>Panel D: Include current and lagged local weather variables</i>							
Current + All Lags	1.98*** (0.59)	1.06* (0.58)	3.01** (1.24)	1.81** (0.87)	4.29 (3.65)	0.25 (0.76)	-0.61 (0.60)
First-stage F	24.88	24.82	24.81	26.33	33.75	24.86	24.88
<i>Panel E: Include lagged weather variables at both source cities and destination city</i>							
Current + All Lags	1.80*** (0.60)	0.68 (0.61)	3.03** (1.24)	1.37 (0.90)	4.02 (3.74)	0.36 (0.77)	-0.55 (0.59)
First-stage F	23.77	23.81	23.64	24.44	32.16	23.76	23.73

Notes: The dependent variable is log(number of transactions) for a given consumption category in city i on day t . Each cell reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 90 days, $\sum_{\tau=0}^{90} \beta_{\tau}$, estimated via the IV version of the flexible distributed lag model. Panel A replicates the bottom row of Table 3 which controls for linear terms of weather variables (temperature, precipitation and average wind speed). Panel B includes 2nd-order polynomials in weather variables. Panel C includes bins of temperature, precipitation and average wind speed as well as all the interactions between the bins. Temperature is grouped into 10 bins ($< 10F^{\circ}$, $10 - 20F^{\circ}$, ..., $> 90F^{\circ}$). Precipitation and average wind speed are grouped into 6 bins each. Panel D includes the following weather lags at the destination city: the weather of day $(t - 1)$, $(t - 2)$, up to day $(t - 7)$; the average weather between day $(t - 8)$ and $(t - 14)$, the average weather between $(t - 15)$ and $(t - 28)$, the average weather between day $(t - 29)$ and $(t - 56)$ and the average weather between $(t - 57)$ and $(t - 90)$. Panel E includes the same set of weather lags as in Panel D for both the source and destination cities. The remaining controls and IVs are the same as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table 9: Cumulative Effects of Pollution: Robustness of IVs

	Health-related Consumption				Non-health Spending		
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
<i>Panel A: IV constructed using source cities' time-invariant pollution</i>							
Current + All Lags	2.69*** (0.64)	2.17*** (0.58)	3.83*** (1.22)	2.16** (0.90)	2.41 (1.99)	-1.74*** (0.67)	-0.56 (0.57)
First-stage F	26.96	27.02	26.95	26.88	29.34	26.79	26.93
<i>Panel B: IV constructed without destination cities' weather variables</i>							
Current + All Lags	2.91*** (0.71)	2.49*** (0.77)	2.79*** (0.90)	2.41*** (0.80)	6.66*** (2.56)	-0.41 (0.62)	-0.31 (0.49)
First-stage F	29.62	29.67	29.60	30.57	37.92	29.55	29.57
<i>Panel C: Drop major cities</i>							
Current + All Lags	2.25*** (0.55)	1.75*** (0.51)	2.75*** (0.92)	2.21*** (0.76)	5.50*** (2.03)	-0.48 (0.60)	-0.51 (0.48)
First-stage F	37.00	37.03	37.06	38.16	47.25	36.98	36.93

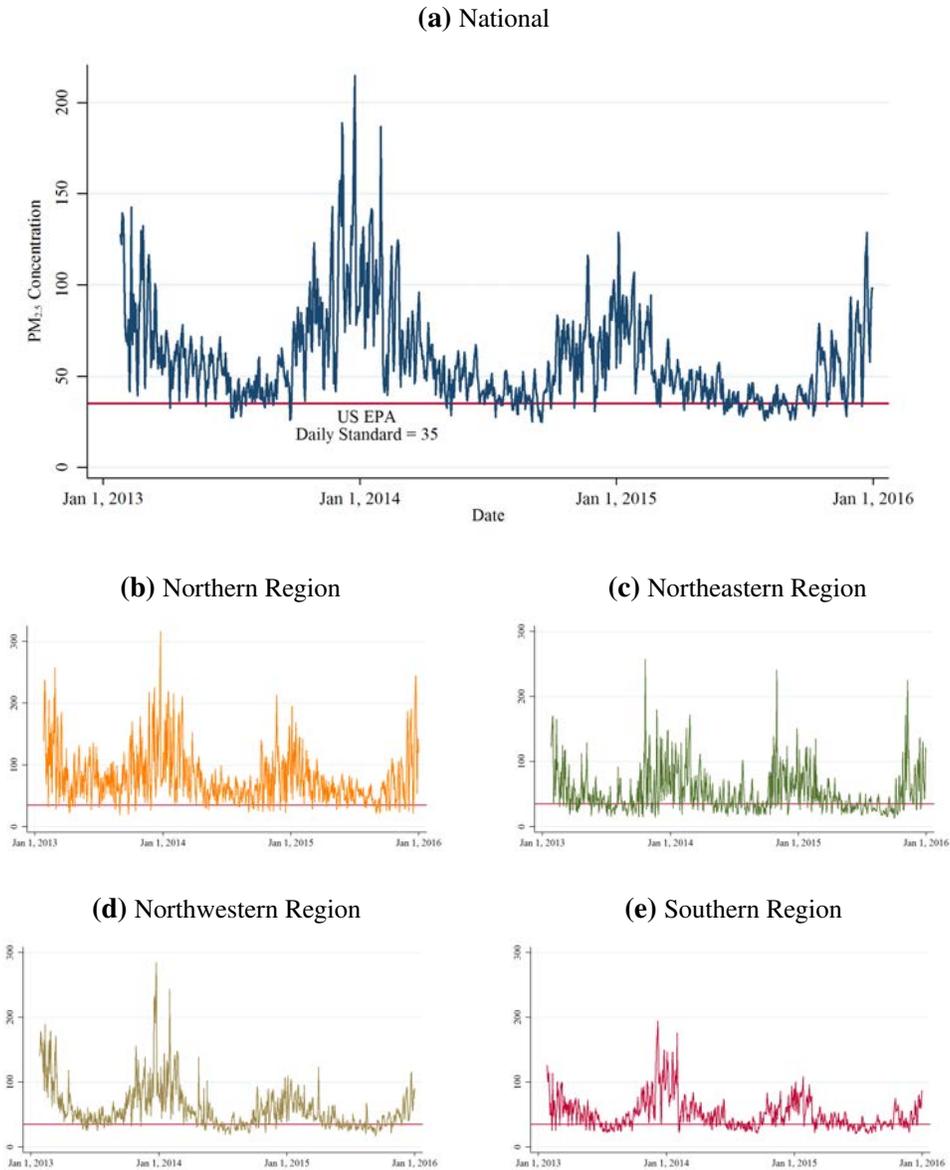
Notes: The dependent variable is log(number of transactions) for a given consumption category in city i on day t . Each cell reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 90 days, estimated via the IV version of the flexible distributed lag model. Same controls as in Table 3. Panel A constructs the IVs using the *average* (time-invariant) level of $\text{PM}_{2.5}$ in cities more than 150 km away. Panel B drops IVs constructed using destination cities' weather so that none of the IVs uses local information. Panel C drops the following large cities: Beijing, Shanghai, Guangzhou, Shenzhen, Wuhan, Chongqing, Chengdu, and Nanjing. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Appendices

Online Appendix. Not for Publication.

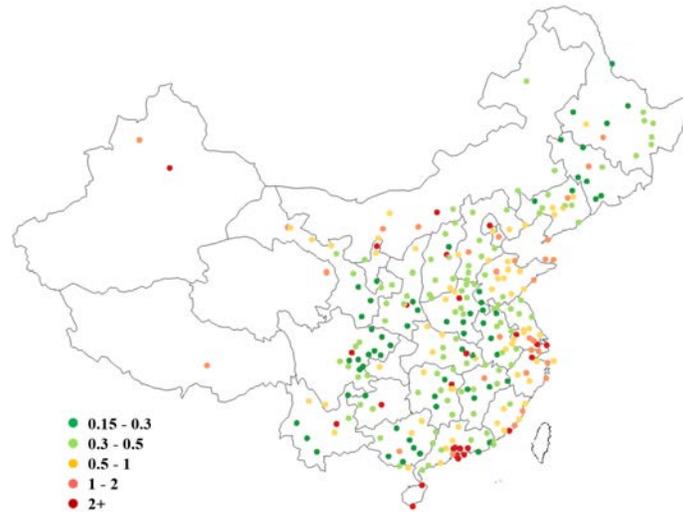
A Descriptive Data Pattern

Figure A1: National and Regional Average Daily PM_{2.5} Concentration ($\mu\text{g}/\text{m}^3$) 2013-2015



Notes: This figure reports the national and regional average daily PM_{2.5} concentration ($\mu\text{g}/\text{m}^3$) during 2013-2015. The Red line in all subfigures indicates the daily standard set by the US EPA: $35 \mu\text{g}/\text{m}^3$. Daily averages are across all monitoring stations in the respective region.

Figure A2: The Number of Active Bank Cards per Capita, 2015



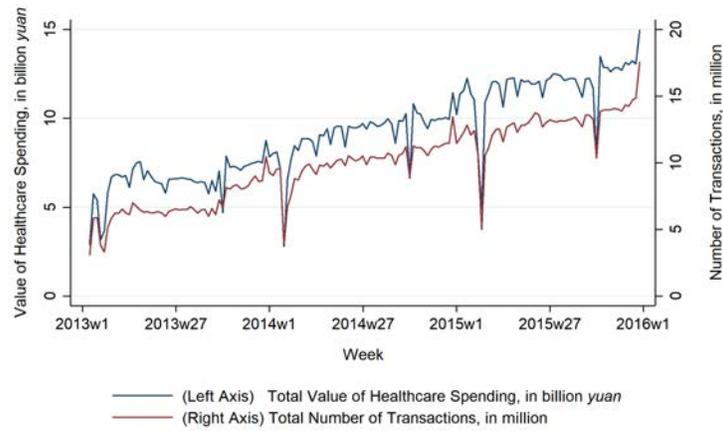
Notes: Bank cards include debit and credit cards. Active bank cards are defined as cards that have been used at least once in a given year. Each card is assigned to one primary city based on the location of its most frequent usage. Population measure is year-end registered population of each city.

Table A1: Coverage of Bank Cards from UnionPay in 2015

	Log(No. of cards per capita)	
	(1)	(2)
log(household income)	1.556*** (0.093)	1.362*** (0.126)
Years of education	0.156*** (0.041)	0.327*** (0.055)
Average age	-0.040*** (0.012)	0.005 (0.014)
Constant	-13.00*** (0.662)	-13.82*** (0.983)
Province fixed effects	No	Yes
No. of obs.	287	287
R^2	0.682	0.831

Notes: The unit of observation is a city. The dependent variable is the log of number of active bank cards per capita in 2015 as shown in Figure A2. The city-level demographics (income, education, and age) are from the 2005 Census.

Figure A3: Weekly Healthcare Spending, 2013 - 2015

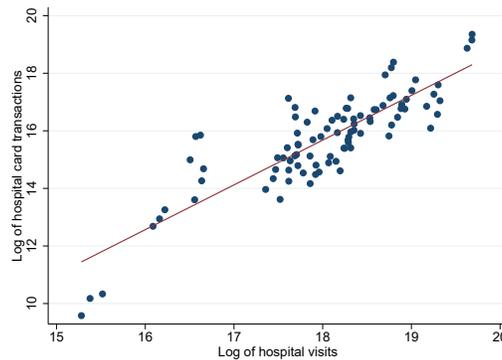


Notes: bank card transactions in all healthcare facilities aggregated to the national-week level from 2013 to 2015.

A.1 Correlation between Card Spending and Health Outcomes

Our data on card transactions in healthcare facilities do not identify specific diseases associated with the spending. Figure A4 illustrates the high correlation between the log number of Unionpay card transactions in hospitals against the log number of NBS reported hospital visits at the province-year level for our sample period. The remaining part of this section provides further evidence on the strong correlation between our spending data and health outcomes based on two confidential micro-level health data.

Figure A4: Annual Hospital Card Transactions vs. Hospital Visits

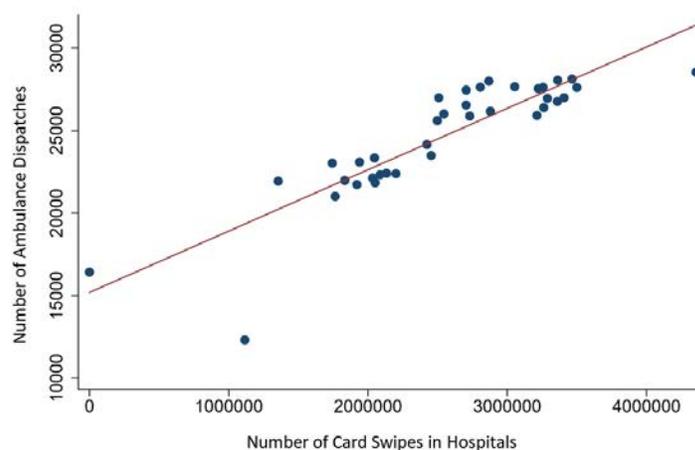


Notes: The figure plots the logarithm of bank card transactions in hospitals (our data) against the logarithm of total hospital visits (from the National Bureau of Statistics) at the province-year level from 2013 to 2015.

Beijing: Emergency Ambulance Dispatches (EAD) Our first micro-level evidence on the validity of using card transactions in healthcare facilities as a measure of health outcomes is based on the universe of daily emergency ambulance dispatches (EAD) data in Beijing from 2013 to 2015, which was used in [Zhong et al. \(2017\)](#). EAD has been used in the medical literature as a measure of health outcomes ([Yang et al., 2014](#); [Straney et al., 2014](#); [Dolney and Sheridan, 2006](#)). In China, an ambulance is dispatched whenever someone calls 120, the public phone number for emergency medical services. Due to the low private vehicle ownership and the highly subsidized fees for emergency ambulances (50 Yuan (about \$7) within 3km and 7 Yuan (about \$1) for each additional km), calling 120 is very common in case of a medical emergency.

Figure A5 shows a strong positive correlation between the number of ambulance dispatches in Beijing and the number of card transactions in hospitals at the monthly level. The correlation coefficient between these two variables is 0.55 at the daily level and 0.88 at the monthly level. Table A2 presents the OLS regressions of daily card transactions in hospitals on daily emergency ambulance dispatches in Beijing with and without controlling for various fixed effects. These fixed effects capture seasonalities in diseases that are common underlying factors behind card transactions and ambulance dispatches. The positive correlation between these two variables persists after a rich set of time fixed effects, is very precisely estimated (significant at the 1% level), and remains stable from Column 3 onwards. The results imply that the correlation between the two data series is driven not only by seasonalities (captured by the fixed effects) but also by idiosyncratic factors (such as weather and air pollution). The R-squared value is high, considering the different nature of these two times series.

Figure A5: Monthly Card Transactions vs. Ambulance Dispatches in Beijing 2013-2015



Notes: The figure plots the number of ambulance dispatches in Beijing (y-axis) against the number of card transactions at hospitals in Beijing (x-axis) at the monthly level from 2013 to 2015.

Table A2: Daily Card Transactions and Ambulance Dispatches in Beijing 2013 - 2015

		Number of Card Transactions in Hospitals						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number of Emergency Calls	227.5*** (10.60)	213.0*** (7.804)	89.64*** (15.37)	92.49*** (15.38)	91.84*** (15.46)	85.56*** (11.90)	86.13*** (11.89)
Fixed Effects	Day of the Week		Yes	Yes	Yes	Yes	Yes	Yes
	Year-week			Yes	Yes	Yes	Yes	Yes
	Linear Trend				Yes	Yes	Yes	Yes
	Season					Yes	Yes	Yes
	Holidays						Yes	Yes
	Spring Festival							Yes
R^2		0.300	0.625	0.849	0.850	0.851	0.912	0.912

Notes: We regress the number of card transactions in hospitals in Beijing on the number of hospital ambulance dispatches in Beijing at the daily level from 2013 to 2015. The number of observations is 1,078. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

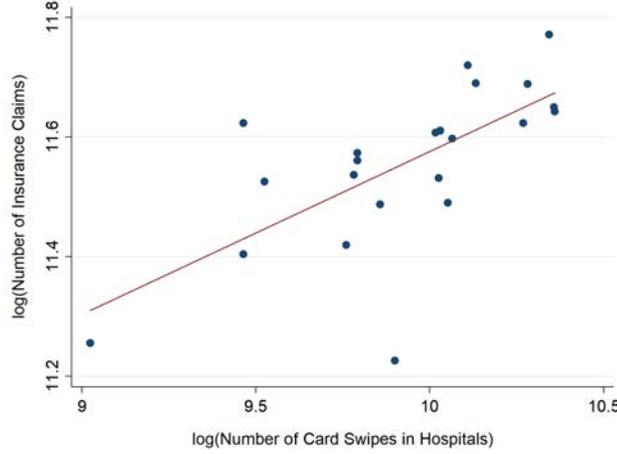
Ganzhou City: Health Insurance Claims Our second micro-level evidence is based on the universe of health insurance claims by urban employees from the Urban Employee Health Insurance program (UEHI) in the city of Ganzhou in Jianxi province. UEHI is one of the three major public health insurance programs in the country. Ganzhou is a medium sized city with a population of 8.4 million, 75% of whom lives in rural areas. Its GDP per capita is 15,000 *yuan* in 2011, less than half of the national average. Bank card penetration in these rural areas is lower than in urban cities. We use Ganzhou to evaluate the correlation between card transactions and health outcomes in rural areas, which helps to address concerns of sample selection.

The data contain the total number of health insurance claims at the daily level from January 2012 to September 2013, which overlaps with the earlier part of our sample period.⁴¹ As the case with Beijing, there is a strong and positive correlation between the number of health insurance claims and hospital related card transactions (Figure A6 and Table A3). The correlation coefficient is 0.39 at the daily level and 0.69 at the monthly level.⁴² The R-squared value is high and the regression coefficients are statistically significant, with Column (7) implying a 1% increase in insurance claims is associated with a 0.1% increase in the number of hospital card transactions. These results suggest there is a strong correlation between card transactions and insurance claims after partialling out a rich set of time fixed effects.

⁴¹Card penetration prior to 2013 was lower than that in our sample period. We expect the correlation between card transactions and health insurance claims during our sample period to be higher than the numbers reported here.

⁴²The date reported in the claims data refers to the date when the patient visited the hospital, rather than the date when the claim was filed.

Figure A6: Monthly Hospital Card Transactions vs. Insurance Claims in Ganzhou



Notes: The figure plots the logarithm of the number of insurance claims in Ganzhou (y-axis) against the logarithm of the number of card transactions at hospitals in Ganzhou (x-axis). The data is aggregated to the monthly level from January 2012 to September 2013.

Table A3: Daily Hospital Card Transactions and Insurance claims in Ganzhou

		log (Number of Card Transactions in Hospitals)						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
	log (Number of Insurance Claims)	0.571*** (0.057)	0.553*** (0.056)	0.198*** (0.048)	0.193*** (0.048)	0.192*** (0.048)	0.118** (0.048)	0.103** (0.048)
Fixed Effects	Day of Week		Yes	Yes	Yes	Yes	Yes	Yes
	Year-week			Yes	Yes	Yes	Yes	Yes
	Linear Trend				Yes	Yes	Yes	Yes
	Season					Yes	Yes	Yes
	Holidays						Yes	Yes
	Spring Festival							Yes
R^2		0.140	0.177	0.651	0.652	0.652	0.677	0.680

Notes: We regress the logarithm of the number of card transactions in hospitals in Ganzhou on the logarithm of the number of insurance claims in Ganzhou. The data is at the daily level, ranging from January 2012 to September 2013. The number of observations is 623. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B B-Spline and IV Construction

B.1 B-Spline Construction

Table B4 illustrates that using daily lagged pollution leads to imprecise and oscillating coefficients as a result of high serial correlations among lagged pollution. To address this issue, we extend Almon (1965) and specify the impact of past pollution exposure on today's spending as basis func-

tions to flexibly capture pollution’s dynamic impact on health spending.

Table B4: IV Estimates for the Effect of Lagged PM_{2.5} on Health Spending

	(1)	(2)	(3)	(4)	(5)	(6)
$p_{i,t}$	0.65*** (0.0891)	0.213 (0.141)	1.639*** (0.277)	1.282*** (0.227)	1.032*** (0.210)	1.012*** (0.204)
$p_{i,t-1}$		0.343** (0.132)	-1.708*** (0.367)	-1.062*** (0.294)	-0.646* (0.268)	-0.505* (0.255)
$p_{i,t-2}$			1.136*** (0.220)	0.572** (0.206)	0.270 (0.206)	-0.00588 (0.217)
$p_{i,t-3}$				0.152 (0.115)	0.256 (0.150)	0.604** (0.194)
$p_{i,t-4}$					-0.0467 (0.0953)	-0.413** (0.141)
$p_{i,t-5}$						0.231* (0.104)
N	192586	191,598	190,786	190,068	189,401	188,750
First-stage F	61.93	29.24	13.03	10.06	11.19	12.16

Notes: The dependent variable is log(total number of healthcare transactions) in city i on day t . Same controls as in Table 3. The IVs are interactions of pollution transported from distant source cities (150km away) and meteorological conditions in the source and destination cities as defined in Equation (6) in the main text. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Recall our baseline specification:

$$y_{it} = \sum_{\tau=0}^k \beta_{\tau} p_{i,t-\tau} + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \quad (\text{B.1})$$

where $\{\beta_{\tau}\}_{\tau}$ denotes the impact of pollution exposure τ days in the past on today’s spending. We assume that β_{τ} can be approximated by a set of B-spline basis functions of τ : $\beta_{\tau} = \sum_m \gamma_m B_m(\tau)$, where γ_m are unknown parameters to be estimated and B_m are basis functions. Section 3.1 discusses the example of a cubic B-spline with one segment, which amounts to a simple 3rd order polynomial function of τ . We now describe how to extend this to the more general case where the basis function $B_m(\tau)$ is an r -th order B-spline in τ with z segments.

To do so we introduce some new notation. Let the support of τ be $[0, \bar{s}]$. We divide the support into z sub-intervals by a vector of $z + 1$ knots $\mathbf{t} = [t_0, t_1, \dots, t_z]$, where $t_0 = 0$ and $t_z = \bar{s}$. The r -th order B-spline, which is equivalent to a piecewise polynomial of order $r - 1$ (enforcing C^{r-2} continuity),

can be constructed from a set of basis functions:

$$B_{m,r}(\tau|\mathbf{t}) = (t_{m+r} - t_m) \sum_{k=0}^r \left[\prod_{0 \leq h \leq r, h \neq k} (t_{m+h} - t_{m+k}) \right]^{-1} (\tau - t_{m+k})_+^{r-1}$$

where

$$(\tau - t_{m+k})_+^{r-1} = \mathbb{1}(\tau > t_{m+k}) \cdot (\tau - t_{m+k})^{r-1}$$

Since there are z subintervals and the order of the spline is r , there will be $z + r - 1$ such B-splines.

We can now define β_τ as a linear combination of these B-splines:

$$\beta_\tau = \sum_{m=1-r}^{z-1} \gamma_m B_{m,r}(\tau|t_m, \dots, t_{m+r}) \quad (\text{B.2})$$

Plug equation (B.2) into equation (B.1) and rewrite the distributed lag model as:

$$\begin{aligned} y_{it} &= \sum_{\tau=0}^k \beta_\tau p_{i,t-\tau} + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \\ &= \sum_{\tau=0}^k \sum_{m=1-r}^{z-1} \gamma_m B_{m,r}(\tau|\mathbf{t}) p_{i,t-\tau} + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \\ &= \sum_{m=1-r}^{z-1} \gamma_m \left[\sum_{\tau=0}^k B_{m,r}(\tau|\mathbf{t}) p_{i,t-\tau} \right] + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \\ &= \sum_{m=1-r}^{z-1} \gamma_m v_{m,it} + \mathbf{x}_{it} \alpha + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \end{aligned}$$

where $v_{m,it} = \sum_{\tau=0}^k B_{m,r}(\tau|\mathbf{t}) p_{i,t-\tau}$, a weighted sum of past pollution exposure with the B-spline basis terms $B_{m,r}(\tau|\mathbf{t})$ as weights.

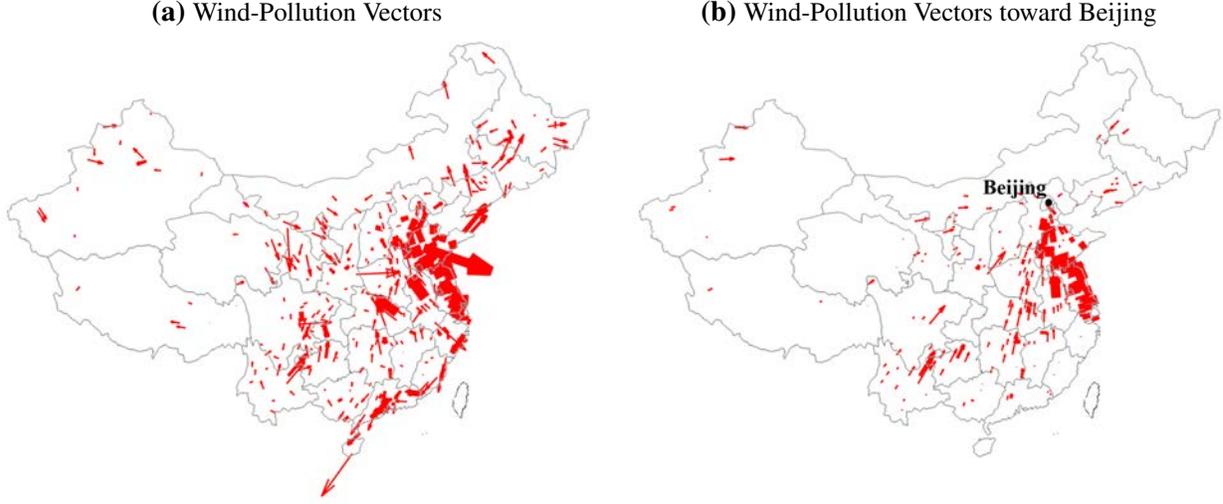
In practice, the econometrician chooses both the order of the spline, $r - 1$, and the number of segments, z . In our benchmark estimates, we use cubic B-splines ($r = 4$) with three segments ($z = 3$), which leads to six key regressors $\{v_{1,it}, \dots, v_{6,it}\}$. The cubic B-spline is a popular choice and is equivalent to a piece-wise cubic polynomial with smoothness constraints at each knot up to the 2nd order derivative (twice continuously differentiable). We choose three segments so that the time series pattern of the marginal impact β_τ for each past month is characterized by a separate cubic polynomial.

B.2 IV Construction

As discussed in Section 3.2.2, we exploit the long-range transport property of PM_{2.5} to construct instrumental variables for PM_{2.5}. To provide a graphical example, panel (a) of Figure B7 plots the

wind-pollution vectors for over 300 cities on Dec. 5, 2013. Each arrow's length indicates the wind speed, rescaled to match the exact distance the arrow can travel in a day. The arrow width indicates the level of PM_{2.5} concentration at the source city. Panel (b) shows all pollution subvectors that are blown towards Beijing on the same day. The amount of pollution that Beijing imports is the sum of pollutants carried through *all* subvectors that reach Beijing at time t after originating from other cities in previous days.

Figure B7: Wind-Pollution Vector Decomposition



Notes: Panel (a) depicts the wind-pollution vector fields on Dec. 5, 2013. Each vector's length indicates the wind speed (re-scaled to match the distance traveled per day). Its width indicates the PM_{2.5} concentration level in the source city. Panel (b) plots the decomposed subvectors pointing towards Beijing.

Our instruments aim to proxy for the amount of pollution imported from cities outside of the buffer zone, \hat{p}_{it}^{far} :

$$\begin{aligned}
 \hat{p}_{it}^{far} &= \sum_{j:d_{ij}>r} p_{j \rightarrow i, t}^+ = \sum_{j:d_{ij}>r} \max[\cos \Phi_{ji}, 0] \cdot p_{j, t-s_{ijt}} \cdot \sum_l^L \gamma_l u_l(d_{ij}, w_{j, t-s_{ijt}}, w_{i, t}) \\
 &= \sum_l^L \gamma_l \sum_{j:d_{ij}>r} \max[\cos \Phi_{ji}, 0] \cdot p_{j, t-s_{ijt}} \cdot u_l(d_{ij}, w_{j, t-s_{ijt}}, w_{i, t}) \\
 &= \sum_l^L \gamma_l Z_{it}^l, \text{ where } Z_{it}^l = \sum_{j:d_{ij}>r} \max[\cos \Phi_{ji}, 0] \cdot p_{j, t-s_{ijt}} \cdot u_l(d_{ij}, w_{j, t-s_{ijt}}, w_{i, t}), \quad l = 1, \dots, L
 \end{aligned}
 \tag{B.3}$$

The instruments for p_{it} is the set of $\{Z_{it}^l\}_{l=1}^L$. These are valid instruments since they only depend on weather in city i , which we control for in our regressions, and on pollution and weather variables in cities outside of the buffer zone, which are uncorrelated with city i 's spending shocks by our

identification assumption. In our baseline specification, we use 15 second-order polynomial terms $u_l(\cdot)$ for a flexible approximation of the decay function: 1) constant, the inverse distance, and origin city's weather (wind speed, precipitation, temperature) (5 terms); 2) the quadratic terms of the inverse distance, and the quadratic terms of origin city's weather (4 terms); 3) the product of the inverse distance and the origin city's weather (3 terms); 4) the destination city's weather (wind speed, precipitation, temperature) (3 terms).

Baseline IVs In summary, we use fifteen instruments $\{Z_{it}^l\}_{l=1}^L$ for current-day pollution p_{it} , where the IVs are defined in equation (B.3). The IVs for lagged pollution $p_{i,t-\tau}$ are lagged instruments $\{Z_{i,t-\tau}^l\}_{l=1}^L$.

We now describe how to construct instruments for our flexible distributed lag model. As shown in Section 3.1 and in Appendix B.1, the lagged distributed model can then be written as:

$$\begin{aligned} y_{it} &= \sum_{m=1-r}^{z-1} \gamma_m \left[\sum_{\tau=0}^k B_{m,r}(\tau|\mathbf{t}) p_{i,t-\tau} \right] + \mathbf{x}_{it} \boldsymbol{\alpha} + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \\ &= \sum_{m=1-r}^{z-1} \gamma_m v_{m,it} + \mathbf{x}_{it} \boldsymbol{\alpha} + \theta_i \cdot t + \xi_i + \eta_w + \varepsilon_{it} \end{aligned}$$

where the main regressors, $v_{m,it} = \sum_{\tau=0}^k B_{m,r}(\tau|\mathbf{t}) p_{i,t-\tau}$, are weighted sums of lagged pollution $p_{i,t-\tau}$. To construct instruments for these regressors, we similarly take the weighted sum of the lagged exogenous variables $Z_{i,t-\tau}^l$, where the weights are the same B-spline terms:

$$W_{m,it}^l = \sum_{\tau=0}^k B_{m,r}(\tau|\mathbf{t}) Z_{i,t-\tau}^l, \quad l = 1, \dots, L$$

Our L instruments for $\{v_{m,it}\}_m$ are therefore $\{W_{m,it}^l\}_{l=1}^L$. In our baseline specification with cubic B-splines ($r = 4$) and three segments ($z = 3$), we have 6 endogenous regressors $\{v_{m,it}\}_{m=-3}^{m=2}$ and 90 instruments.

C Robustness Analysis

C.1 Additional Results for the Baseline Specification

Table C5: First-Stage Regressions

		Endogenous variable: PM _{2.5} , current day			
		(1)	(2)	(3)	
Included IVs	Daily weather	Temperature	-4.57*** (0.44)	-4.21*** (0.43)	-6.47*** (0.64)
		Precipitation	-48.52*** (2.32)	-47.30*** (2.29)	-23.35*** (4.12)
		Wind speed	-2.58*** (0.49)	-3.93*** (0.49)	-5.61*** (0.76)
	Day controls	Holiday	-47.41*** (4.06)	-45.75*** (4.09)	-42.72*** (4.06)
		Working weekend	15.46* (6.23)	15.95* (6.32)	17.58** (6.15)
		Spring festival	293.7*** (17.67)	273.6*** (17.37)	253.9*** (16.80)
City FEs, week FEs, city-specific time trends, city-specific seasonality, day-of-week FEs		Yes	Yes	Yes	
Excluded IVs	Sum of PM _{2.5} from distant cities:				
		Unweighted	0.01*** (0.002)		
		Weighted by inverse of distance		0.46*** (0.03)	
		Weighted by wind speed of source city		0.001 (0.001)	Yes
		Weighted by temperature of source city		-0.09*** (0.006)	
		Weighted by precipitation of source city		-0.02*** (0.005)	
		Weighted by destination city's weather variables, <i>squares</i> of inverse distance and source cities' weather variables, and interactions between inverse distance and source cities' weather variables.			Yes
R ²		0.46	0.46	0.47	
Number of excluded IVs		1	4	15	
Effective F-statistic for excluded IVs		36.02	161.06	112.21	
Critical value for weak IVs		37.41	28.84	28.89	

Notes: The endogenous variable is current-day PM_{2.5}. The number of observation is 192,586. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** p < 0.01, ** p < 0.05, and * p < 0.10. Effective F-statistics for excluded IVs following [Olea and Pflueger \(2013\)](#) are reported in the bottom panel. Critical value for weak IVs reports the upper 5% quantile for the noncentral χ^2 distribution from [Olea and Pflueger \(2013\)](#), allowing for a 5% bias with one single endogenous variable.

Table C6: OLS Estimates of the Pollution Impact on Health Spending: Contemporaneous Effects

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PM _{2.5} , Current Day	0.11*** (0.02)	0.11*** (0.02)	0.12*** (0.02)	0.13*** (0.02)	0.18*** (0.05)	-0.06*** (0.02)	-0.03 (0.02)
N	192,586	191,814	191,277	185,773	146,224	192,035	191,766

Notes: the dependent variable is log(number of transactions) for a given consumption category in city i on day t . Each column reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in PM_{2.5} in the current day. Same controls as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table C7: IV Estimates of the Pollution Impact on Health Spending: Contemporaneous Effects

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PM _{2.5} , Current Day	0.65*** (0.09)	0.73*** (0.11)	0.60*** (0.15)	0.77*** (0.13)	1.13*** (0.37)	-0.09 (0.15)	-0.10 (0.12)
N	192,586	191,814	191,277	185,773	146,224	192,035	191,766
First-stage F	112.21	111.37	111.17	104.91	86.13	111.71	111.77
Critical value for weak IVs	28.89	28.88	28.89	28.85	29.01	28.88	28.89

Notes: the dependent variable is log(number of transactions) for a given consumption category in city i on day t . Each column reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in PM_{2.5} in the current day. Same controls as in Table 3. The IVs are interactions of pollution transported from distant source cities (150km away) and meteorological conditions in the source and destination cities as defined in Equation (6). Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The First-stage F for excluded IVs is the effective F-statistics following [Olea and Pflueger \(2013\)](#). Critical value for weak IVs reports the upper 5% quantile for the noncentral χ^2 distribution from [Olea and Pflueger \(2013\)](#), allowing for a 5% bias with one single endogenous variable.

Table C8: IV Estimates of Pollution Impacts on the Value of Transactions with 90 Lags

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current Day	0.07*** (0.02)	0.07** (0.03)	0.01 (0.05)	0.10** (0.05)	0.01 (0.09)	-0.12** (0.05)	-0.04 (0.05)
Current + Past 3d	0.23*** (0.08)	0.25*** (0.08)	0.04 (0.15)	0.34** (0.15)	0.05 (0.30)	-0.38** (0.17)	-0.16 (0.17)
Current + Past 7d	0.36*** (0.11)	0.38*** (0.13)	0.04 (0.21)	0.54** (0.23)	0.13 (0.45)	-0.55** (0.24)	-0.36 (0.25)
Current + Past 14d	0.42*** (0.15)	0.46*** (0.17)	0.03 (0.24)	0.69** (0.27)	0.33 (0.59)	-0.56** (0.27)	-0.68** (0.31)
Current + Past 28d	0.43* (0.25)	0.44 (0.29)	0.29 (0.34)	0.79*** (0.30)	1.09 (0.88)	-0.34 (0.34)	-0.99** (0.43)
Current + Past 56d	1.04** (0.47)	0.83 (0.54)	1.64*** (0.61)	1.15** (0.47)	4.07** (1.72)	-0.32 (0.58)	-0.95 (0.75)
Current + All Lags	1.47** (0.70)	1.08 (0.78)	1.96** (0.96)	1.20 (0.83)	6.12** (2.61)	0.54 (0.87)	-0.66 (1.04)
N	141,794	141,656	141,566	137,854	110,257	141,757	141,641
First-stage F	38.35	38.38	38.37	39.68	47.79	38.26	38.30

Notes: the dependent variable is $\log(\text{value of transactions})$ for a given consumption category in city i on day t . Each row reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the corresponding period, $\sum_{\tau=0}^k \beta_{\tau}$, estimated via the IV version of the flexible distributed lag model with 90 lags. Same controls and IVs as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

C.2 Additional Robustness Analysis

Table C9: IV Estimates of Pollution Impacts on the Number of Transactions per capita

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current Day	0.11*** (0.02)	0.12*** (0.03)	0.05 (0.04)	0.14*** (0.04)	0.19*** (0.07)	-0.16*** (0.03)	-0.07*** (0.02)
Current + Past 3d	0.36*** (0.08)	0.38*** (0.08)	0.19 (0.12)	0.48*** (0.12)	0.64*** (0.24)	-0.51*** (0.09)	-0.26*** (0.06)
Current + Past 7d	0.56*** (0.11)	0.58*** (0.12)	0.34* (0.18)	0.76*** (0.18)	1.03*** (0.36)	-0.73*** (0.13)	-0.41*** (0.10)
Current + Past 14d	0.70*** (0.14)	0.70*** (0.16)	0.54** (0.22)	0.98*** (0.22)	1.40*** (0.50)	-0.73*** (0.16)	-0.53*** (0.12)
Current + Past 28d	0.91*** (0.22)	0.86*** (0.25)	1.04*** (0.30)	1.22*** (0.27)	2.13*** (0.79)	-0.54** (0.23)	-0.46** (0.21)
Current + Past 56d	1.99*** (0.42)	1.71*** (0.47)	2.34*** (0.53)	1.98*** (0.45)	4.71*** (1.54)	-0.95** (0.42)	-0.25 (0.36)
Current + All Lags	2.70*** (0.69)	2.22*** (0.72)	2.85*** (0.89)	2.26*** (0.75)	6.53*** (2.33)	-0.59 (0.61)	-0.61 (0.47)
N	135,297	135,162	135,291	131,699	107,424	135,283	135,294
First-stage F	39.12	39.16	39.16	40.85	52.47	39.09	39.14

Notes: the dependent variable is log(number of transactions per capita) for a given consumption category in city i on day t . Each row reports the percentage change in the dependent variable in response to a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the corresponding period, $\sum_{\tau=0}^k \beta_{\tau}$, estimated via the IV version of the flexible distributed lag model with 90 lags. Same controls and IVs as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table C10: IV Estimates of the Impact of Average Lagged Pollution on Spending

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Contemporaneous effect (Table C7)							
Current Day PM _{2.5}	0.65*** (0.09)	0.73*** (0.11)	0.60*** (0.15)	0.77*** (0.13)	1.13*** (0.37)	-0.09 (0.15)	-0.10 (0.12)
First-stage F	112.21	111.37	111.17	104.91	86.13	111.71	111.77
Panel B:							
7-day Average PM _{2.5}	0.68*** (0.13)	0.73*** (0.16)	0.63*** (0.17)	0.83*** (0.15)	1.58** (0.60)	-0.22 (0.13)	-0.08 (0.12)
First-stage F	109.3	109.1	108.7	104.2	83.43	109.2	109.3
Panel C:							
30-day Average PM _{2.5}	1.17*** (0.28)	1.25*** (0.33)	1.31*** (0.38)	1.39*** (0.32)	3.33** (1.27)	-0.10 (0.27)	0.37 (0.25)
First-stage F	61.74	61.71	61.46	59.28	45.12	61.73	61.80
Panel D:							
60-day Average PM _{2.5}	1.97*** (0.53)	1.67** (0.60)	2.21** (0.69)	1.79*** (0.52)	5.50* (2.22)	-0.53 (0.45)	0.17 (0.35)
First-stage F	37.97	37.92	37.91	36.51	29.34	37.97	37.96
Panel E:							
90-day Average PM _{2.5}	2.15** (0.76)	1.50 (0.79)	2.44* (1.02)	1.61* (0.73)	6.53* (2.64)	-0.44 (0.63)	-0.11 (0.44)
First-stage F	27.69	27.63	27.66	26.96	24.43	27.68	27.71

Notes: the dependent variable is log(number of transactions) for a given consumption category in city i on day t . Each panel is a separate set of regressions, where the key regressor is PM_{2.5} averaged over the corresponding time window. For example, in Panel B, the regressor '7-day average PM_{2.5}' is the average PM_{2.5} from day $t - 7$ to day t . Remaining controls (fixed effects, time trends and weather controls) and IVs are the same as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F for excluded IVs is the effective F-statistic following [Olea and Pflueger \(2013\)](#). Critical value for weak IVs, allowing for a 5% bias with one endogenous variable, ranges from 27.3 to 29.0 depending on the specification.

Table C11: Placebo Exercise – IV Estimates Using Randomized Wind Direction and Speed

	(1)	(2)	(3)
Only keep IVs that interact with wind	No	Yes	Yes
Randomize wind direction + speed (placebo)	No	No	Yes
Current Day	0.12*** (0.02)	0.11 (0.07)	-0.06* (0.03)
Current + Past 3d	0.40*** (0.07)	0.37* (0.21)	-0.15 (0.12)
Current + Past 7d	0.61*** (0.10)	0.54* (0.31)	-0.10 (0.22)
Current + Past 14d	0.74*** (0.14)	0.60 (0.37)	0.20 (0.41)
Current + Past 28d	0.91*** (0.22)	0.71* (0.40)	0.80 (0.71)
Current + Past 56d	1.97*** (0.42)	2.07*** (0.55)	1.13 (1.17)
Current + All Lags	2.65*** (0.68)	2.29*** (0.83)	-1.69 (1.82)
# IVs	90	24	24
First-stage F	38.35	15.09	7.24

Notes: the dependent variable is $\log(\text{total number of healthcare transactions})$ in city i on day t . The number of observations is 141,794. Same controls as in Table 3. Column (1) replicates the first column in Table 3. Column (2) keeps IVs that interact pollution transported from distant source cities (150km away) with wind speed and wind direction and drops IVs that use variation in temperature and rainfall. Column (3) replicates Column (2), except that the actual wind direction and speed are replaced with randomly generated wind direction and speed (a placebo exercise). Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

Table C12: Estimates of Pollution Impacts: Alternate IVs Using Wind Direction Following [Deryugina et al. \(2019\)](#)

	(1)	(2)
Baseline IVs	Yes	
Wind direction interacted with spatial group FE as IVs		Yes
Current Day	0.12*** (0.02)	0.06 (0.08)
Current + Past 3d	0.40*** (0.07)	0.19 (0.29)
Current + Past 7d	0.61*** (0.10)	0.30 (0.51)
Current + Past 14d	0.74*** (0.14)	0.39 (0.86)
Current + Past 28d	0.91*** (0.22)	0.56 (1.51)
Current + Past 56d	1.97*** (0.42)	1.22 (2.62)
Current + All Lags	2.65*** (0.68)	1.48 (3.54)
# IVs	90	900
First-stage F (Cragg-Donald)	254.91	14.73

Notes: The dependent variable is log(total number of healthcare transactions) in city i on day t . The number of observations is 141,794. Column (1) replicates the first column in Table 3. Column (2) follows [Deryugina et al. \(2019\)](#), where wind direction dummies are interacted with 50 group dummies classified via the K-means clustering algorithm, resulting in 900 IVs. Since the number of IVs exceeds the number of city clusters, the cluster-robust Kleibergen-Paap Wald rk F-stat cannot be computed. Instead, we report the Cragg-Donald F-statistic which assumes homoskedasticity. For reference, the [Stock and Yogo \(2005\)](#) critical value based on a maximal TSLS bias of 5% with three endogenous variables and 30 instruments is 20.27. It is computationally challenging to compute critical values with many endogenous regressors; simulation evidence indicates the critical value is around 20. As Column (2)'s IVs fail to pass the weak IV test, we follow [Andrews \(2018\)](#) and construct the identification-robust confidence intervals that are valid under weak IVs. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table C13: IV Estimates of Pollution Impacts, Limiting to Cities Where Pollution Was Monitored from 2013 onward

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Current Day	0.11*** (0.03)	0.12*** (0.03)	0.06 (0.04)	0.17*** (0.04)	0.18** (0.08)	-0.13*** (0.03)	-0.05*** (0.02)
Current + Past 3d	0.36*** (0.08)	0.39*** (0.09)	0.20 (0.13)	0.55*** (0.14)	0.60** (0.25)	-0.42*** (0.10)	-0.20*** (0.07)
Current + Past 7d	0.55*** (0.12)	0.58*** (0.14)	0.33* (0.20)	0.84*** (0.20)	0.95** (0.40)	-0.62*** (0.14)	-0.36*** (0.10)
Current + Past 14d	0.64*** (0.16)	0.63*** (0.19)	0.50** (0.24)	1.01*** (0.24)	1.30** (0.57)	-0.67*** (0.16)	-0.52*** (0.13)
Current + Past 28d	0.73*** (0.25)	0.63** (0.28)	0.86** (0.34)	1.10*** (0.28)	2.14** (0.89)	-0.47** (0.22)	-0.46** (0.19)
Current + Past 56d	1.55*** (0.48)	1.28** (0.52)	1.84*** (0.61)	1.54*** (0.46)	4.77*** (1.72)	-0.65* (0.37)	-0.17 (0.31)
Current + All Lags	2.08*** (0.77)	1.62** (0.80)	2.09** (0.98)	1.63** (0.79)	6.37** (2.54)	-0.19 (0.56)	-0.50 (0.45)
N	103,146	103,142	103,140	101,161	87,295	103,137	103,145
First-stage F	45.31	45.33	45.42	50.60	64.90	45.37	45.32

Notes: The dependent variable is log(number of transactions) for a given consumption category in city i on day t . The sample is limited to the 159 cities with hourly monitored pollution data since 2013. Same controls and IVs as in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.

D Heterogeneous Impact

This section examines the heterogeneity of pollution’s impact on health spending. To conserve the number of parameters, we use one-segment instead of three-segment B-splines, since cumulative effects are similar across different segments (Table 5). This corresponds to a simple third-order polynomial, as explained in Section 3.1.

To examine how pollution’s impact on health spending differs across polluted and less-polluted days and by income, we allow β_τ , the coefficient of pollution exposure on day $t - \tau$, to depend on the quadratic term of $\text{PM}_{2.5}$ (for Figure 4 in the main text) and income (for Appendix Figure D8) as well as the linear term. Specifically, the impact of past pollution $p_{i,t-\tau}$ is defined as:

$$\beta_\tau(\tau, w|\gamma, \sigma) = (\sigma_1 w + \sigma_2 w^2) + (\gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3) \quad (\text{D.1})$$

where $\sigma_1 w + \sigma_2 w^2$ captures heterogeneity and allows the intercept of the β_τ -curve to vary across different levels of w . Parameters σ ’s and γ ’s are coefficients to be estimated. If σ ’s are significant, then pollution’s impact on health spending exhibits heterogeneity across variable w .

We report coefficient estimates for equation (D.1) in Table D14. Column (1) reports estimates without heterogeneity, i.e., the specification $z = 1, k = 90$ in Table 5. Columns (2) and (3) report σ estimates that govern heterogeneity across different $\text{PM}_{2.5}$ concentration and per capita disposable income, respectively. We draw Figure 4 and Appendix Figure D8 based on estimates in Columns (2) and (3). Column (4) allows β_τ to depend on both pollution concentration and income.

To examine pollution’s heterogeneous impact across seasons, we use the following:

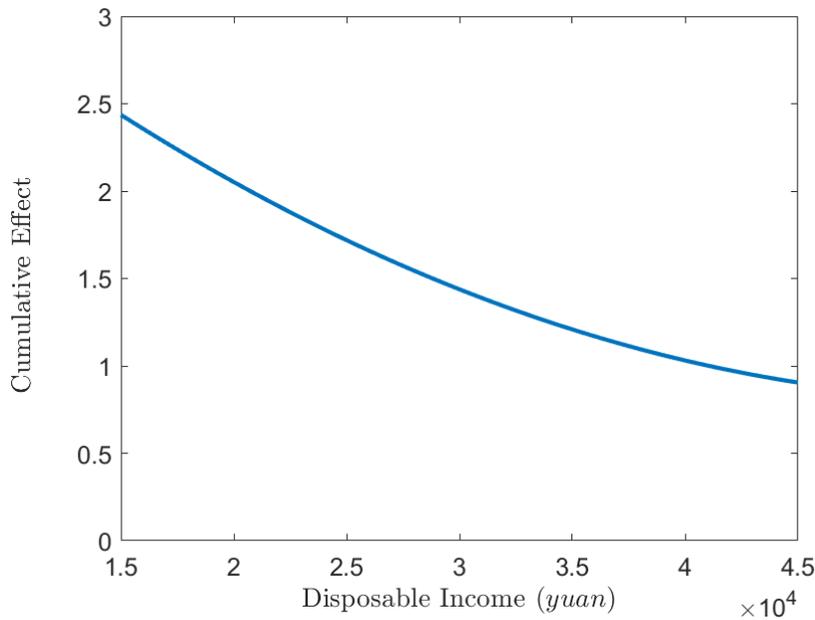
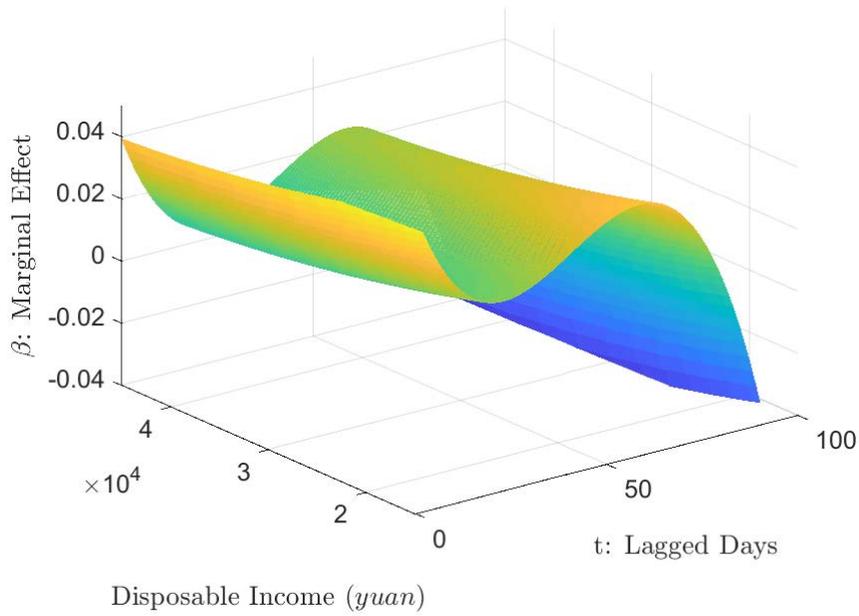
$$\beta_\tau(\tau, w|\gamma, \sigma) = (\sigma_1 \cdot \mathbb{1}\{\text{summer}\} + \sigma_2 \cdot \mathbb{1}\{\text{fall}\} + \sigma_3 \cdot \mathbb{1}\{\text{winter}\}) + (\gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3) \quad (\text{D.2})$$

where $\mathbb{1}\{\text{summer}\}$, $\mathbb{1}\{\text{fall}\}$, and $\mathbb{1}\{\text{winter}\}$ are dummy variables for different seasons. Coefficient estimates for σ ’s are reported in Table D15. Analysis for the heterogeneity across years is done analogously:

$$\beta_\tau(\tau, w|\gamma, \sigma) = (\sigma_1 \cdot \mathbb{1}\{\text{Year 2014}\} + \sigma_2 \cdot \mathbb{1}\{\text{Year 2015}\}) + (\gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3) \quad (\text{D.3})$$

Results are reported in Column 2 of Table D15.

Figure D8: Impact of Air Pollution Across Income Levels



Notes: the top panel plots the (marginal) percentage change in total healthcare transactions for a temporary $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ at a specific level of per-capita income and on a given day. On the t -axis, 0 refers to day t , 50 refers to day $t - 50$, etc. The bottom panel plots the percentage change in total healthcare transactions for a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 90 days at different levels of per capita disposable income as denoted by the x -axis. For example, a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ over the past 90 days raises today's total healthcare transactions by 2.44% when disposable income is 15,000 *yuan*. Based on parameter estimates reported in Appendix Table D14.

Table D14: Coefficient Estimates on Heterogeneity: PM_{2.5} and Disposable Income

		(1)	(2)	(3)	(4)
Heterogeneity (w)		No Heterogeneity	PM _{2.5}	Disposable Income	PM _{2.5} & Income
Cubic Polynomial	γ_1	0.06*** (0.01)	0.06*** (0.01)	0.07* (0.04)	0.08** (0.04)
	γ_2	-4.11E-03*** (1.41E-03)	-3.86E-03*** (1.38E-03)	-3.52E-03*** (1.25E-03)	-3.91E-03*** (1.32E-03)
	γ_3	1.17E-04*** (3.79E-05)	1.11E-04*** (3.74E-05)	1.07E-04*** (3.18E-05)	1.16E-04*** (3.47E-05)
	γ_4	-9.23E-07*** (2.89E-07)	-8.77E-07*** (2.88E-07)	-8.76E-07*** (2.36E-07)	-9.30E-07*** (2.60E-07)
Heterogeneity	σ_1 - PM _{2.5}		0.13** (0.06)		0.14** (0.06)
	σ_2 - PM _{2.5}		-0.37*** (0.11)		-0.40*** (0.12)
	σ_1 - Income			-1.24E-02 (2.53E-02)	-1.53E-02 (2.49E-02)
	σ_2 - Income			1.14E-03 (3.66E-03)	1.62E-03 (3.60E-03)
Max./Min. Point = $\frac{-\hat{\sigma}_1}{2\hat{\sigma}_2}$			179 $\mu\text{g}/\text{m}^3$	54,700 <i>yuan</i>	

Notes: the table reports coefficient estimates on equation D.1:

$$\beta_\tau(\tau, w|\gamma, \sigma) = (\sigma_1 w + \sigma_2 w^2) + (\gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3).$$

where $\beta_\tau(\tau, w|\gamma, \sigma)$ denotes the percentage change in healthcare expenditure in response to a 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} on day $t - \tau$. Column (3) uses each city's average annual per capita disposable income from 2013 to 2015. Column (4) allows β_τ to depend on both pollution concentration and income. PM_{2.5} concentration is rescaled to mg/m^3 and disposable income is rescaled to 10,000 *yuan*. The maximum/minimum points are defined by $\frac{\partial \beta}{\partial w} = 0$, or $w^* = \frac{-\hat{\sigma}_1}{2\hat{\sigma}_2}$, which corresponds to the maximum/minimum point of the quadratic term ($\sigma_1 w + \sigma_2 w^2$). Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. $\gamma_2, \gamma_3, \gamma_4$'s in Columns (2), (3) and (4) are insignificantly different from those in Column (1).

Table D15: Coefficient Estimates on Heterogeneity: Season and Year

		(1)	(2)
Heterogeneity (w)		Season	Year
Cubic Polynomial	γ_1	0.04*** (0.02)	0.05* (0.02)
	γ_2	-2.94E-03** (1.30E-03)	-4.59E-03*** (1.29E-03)
	γ_3	9.36E-05*** (3.47E-05)	1.34E-04*** (3.58E-05)
	γ_4	-7.80E-07*** (2.66E-07)	-1.1E-06*** (2.81E-07)
Heterogeneity	$\mathbb{1}\{\text{spring}\}$	-	-
	$\mathbb{1}\{\text{summer}\}$	-8.13E-05 (1.82E-03)	
	$\mathbb{1}\{\text{fall}\}$	2.13E-03 (2.65E-03)	
	$\mathbb{1}\{\text{winter}\}$	7.66E-03*** (1.60E-03)	
	$\mathbb{1}\{\text{Year 2013}\}$		-
		-	
	$\mathbb{1}\{\text{Year 2014}\}$		0.01 (0.02)
	$\mathbb{1}\{\text{Year 2015}\}$		0.02 (0.03)

Notes: the table reports coefficient estimates based on equation (D.2) and (D.3). $\mathbb{1}\{\cdot\}$ is a dummy variable. This analysis allows β_τ , the percentage change in healthcare expenditure in response to a 10 $\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ on day $t - \tau$, to differ across seasons and years. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

E Healthcare Costs

According to China's [National Health Commission \(2016\)](#), national health expenditure which includes both private and public spending, was more than four trillion *yuan* (\$615 billion) in 2015. Bank card transactions in the Unionpay system account for half of the total private spending (which is roughly 30% of the aggregate health spending). In order to interpolate our estimates of the health impact to the entire population, we need to make assumptions on health spending that is not covered by Unionpay. There are several considerations that suggest our analysis is likely to underestimate the population impact. First, elderly are more vulnerable to air pollution. In the U.S., the elderly

accounts for 15% of the population but 34% of health spending in 2014. The elderly population has few cards per person and is less likely to use credit and debit cards on average. Second, low-income residents experience a more severe health impact from pollution (Figure D8), but bank card penetration is lower in low-income areas. As a result, our analysis is likely to underestimate the population impact.

One might be concerned that individuals using credit and debit cards are likely to have better insurance coverage than individuals without cards, and hence consume more healthcare when pollution increases. While the extent of moral hazard could be different between card users and other consumers, this concern is mitigated by the institutional features of China's healthcare system, in particular the long waiting time (which discourages over-usage), time lags in getting reimbursed for insurance, and high co-pays. Moreover, while individuals with better insurance coverage may use more healthcare than individuals with lower coverage, the morbidity cost (which includes lost productivity and reduced quality of life due to sickness in addition to the healthcare cost) of the latter group is not necessarily lower. Indeed, the inability of an individual to seek treatment upon falling sick (due to limited insurance coverage) may lead to worsening of their health condition, thus increasing their morbidity cost relative to individuals with better insurance coverage.

Our main analysis indicates that health spending (in the Unionpay system) increases by 1.5% in value and 2.65% in transaction frequency in response to a $10 \mu\text{g}/\text{m}^3$ increase of $\text{PM}_{2.5}$ over 90 days. Table E16 benchmarks our results with the findings in the related literature. To estimate the healthcare cost from elevated $\text{PM}_{2.5}$ for the entire population, we assume that the health impact is the same for both bank-card and non-bank-card spending. Based on China's national healthcare spending in 2015, the 1.5% impact from a $10 \text{ mg}/\text{m}^3$ increase in $\text{PM}_{2.5}$ translates to 59.6 billion yuan (\$9.2 billion).

To compare our healthcare estimate (part of the morbidity cost) with the mortality impacts in the literature, we monetize the mortality estimate from Ebenstein et al. (2017). Specifically, Ebenstein et al. (2017) find that a $10 \mu\text{g}/\text{m}^3$ increase of PM_{10} would increase the cardiorespiratory mortality rate by 8% on average. In the main text, we assume that the mortality cost of a $10 \mu\text{g}/\text{m}^3$ increase of $\text{PM}_{2.5}$ is the same as that of a $10 \mu\text{g}/\text{m}^3$ increase of PM_{10} . This translates to a mortality cost of \$13.4 billion. Alternatively, since $\text{PM}_{2.5}$ accounts for 60% of PM_{10} concentration (Zhou et al., 2016), another assumption could be that the mortality cost of a $10 \mu\text{g}/\text{m}^3$ increase of PM_{10} is equivalent to the mortality cost of a $6 \mu\text{g}/\text{m}^3$ increase of $\text{PM}_{2.5}$. Under this alternative assumption, a $10 \mu\text{g}/\text{m}^3$ increase of $\text{PM}_{2.5}$, or equivalently a $16.7 \mu\text{g}/\text{m}^3$ increase in PM_{10} , would inflict a mortality cost of \$22.3 billion (as opposed to \$13.4 billion).

Table E16: Summary of the Dose-Response Relationship from Literature

Source	Dose, additional	Response
Mu and Zhang (2016)	100-point AQI	54.5% increase in mask purchases, 70.6% increase in anti-PM _{2.5} mask purchases
Williams and Phaneuf (2016)	1 std. dev. of PM _{2.5} (3.78 $\mu\text{g}/\text{m}^3$)	8.3% more spending on asthma and COPD
Schlenker and Walker (2016)	1 std. dev. of pollution	17% more asthma and other respiratory incidences, 9% more heart incidences
Arceo et al. (2015)	1 $\mu\text{g}/\text{m}^3$ PM ₁₀ 1 ppb CO	0.23 per 100,000 increase in infant mortality 0.0046 per 100,000 increase in infant mortality
He et al. (2016)	10 $\mu\text{g}/\text{m}^3$ PM ₁₀ (roughly 10%)	8.36% increase in all-cause mortality rate 285,000 more premature deaths each year
Chay and Greenstone (2003)	1% TSP	0.35% increase in infant mortality rate nationwide
Our estimation	10 $\mu\text{g}/\text{m}^3$ PM _{2.5}	2.65% increase in healthcare transactions 1.5% increase in healthcare expenditure

Table E17: Mortality Cost Calculation

Age group	Urban population	Rural Population	Urban mortality rate (per 100,000)	Rural mortality rate (per 100,000)	VSL in million (2015\$)	Mortality impact in percentage
20-24	73,195,616	58,048,857	4.31	3.81	0.2106	10
25-29	59,414,692	44,637,171	5.47	5.54	0.2106	11
30-34	57,695,497	42,364,156	8.07	7.07	0.2106	14
35-39	66,981,015	54,594,597	13.71	12.48	0.2106	10
40-44	65,704,887	62,801,076	26.02	23.90	0.1895	12
45-49	55,242,460	53,527,870	42.25	46.27	0.1684	13
50-54	40,364,926	40,756,761	65.87	71.27	0.1474	13
55-59	38,563,476	45,194,486	105.52	125.79	0.1263	12
60-64	26,819,982	33,611,729	209.62	255.81	0.1053	12
65-69	18,448,986	23,900,786	402.25	459.16	0.0842	11
70-74	15,221,689	18,742,359	880.11	1092.46	0.0632	9
75-79	10,848,240	13,721,250	1744.92	1998.33	0.0421	7
80-84	5,936,146	7,839,253	3632.06	4316.95	0.0316	5
85 and above	3,370,721	4,474,484	9685.26	13128.58	0.0211	3

Notes: this mortality cost calculation follows closely [Deschênes et al. \(2017\)](#). The population data are for 2015. The mortality rates per 100,000 are only for cardiorespiratory diseases and are from the 2015 National Health Statistics. Based on the transfer elasticity of 1.2 and the 2.27 million (in 2015\$) estimate for the U.S. population's VSL from [Ashenfelter and Greenstone \(2004\)](#), the estimated VSL for the Chinese population is \$0.2106 million for a prime age person. The age adjustment is based on [Murphy and Topel \(2006\)](#). The estimated mortality impact (last column) of a 10 $\mu\text{g}/\text{m}^3$ increase in PM_{10} on the cardiorespiratory mortality rate during life cycle is from Table S6 in [Ebenstein et al. \(2017\)](#).

F Avoidance Behavior

This section provides a brief description on evidence that households engage in avoidance behavior to mitigate their pollution exposure. A key insight of this analysis is that when consumers engage in avoidance behavior, expectations of *future* pollution levels should affect current consumption. For example, if consumers expect pollution to improve in the near future, they may postpone their consumption to avoid current exposure. On the other hand, an expectation of worse air quality tomorrow may encourage them to shift future consumption to today.

To investigate avoidance behavior, we assume that consumers can perfectly foresee the air quality on the following day and adjust their spending accordingly. In other words, we include the following day’s pollution as an additional regressor in the main specification. The perfect-foresight assumption is partly driven by the lack of systematic data on pollution forecasts that individuals had access to when making decisions. Admittedly, this is a strong assumption. However, the issue of air quality was highly salient during our sample period and China’s Ministry of Environmental Protection (MEP) had just launched a nationwide pollution monitoring-and-disclosure program (Barwick et al., 2020). Real-time forecasts of air quality were available to consumers both from government websites and smartphone apps. To account for measurement errors and possible omitted variable bias, we instrument for future $PM_{2.5}$ using 1-day leads of our instruments for today’s $PM_{2.5}$.

The results are illustrated in Panel A of Table F18. A $10 \mu g/m^3$ increase in $PM_{2.5}$ on the next day is associated with a 0.70% contemporaneous increase in transactions in the aggregate health care sector. The impact is larger for pharmacies than for hospitals, consistent with the fact that hospital visits are often scheduled in advance and less substitutable intertemporally. Spending in supermarkets also increases when next-day pollution is expected to deteriorate. The estimated cumulative impact on healthcare spending that is associated with a $10 \mu g/m^3$ increase of $PM_{2.5}$ over the past 90 days is 2.51%, slightly lower than but comparable to when we do not control for avoidance. In Panels B and C, we replace pollution the following day with the average pollution in the next 3 days and 7 days respectively. The results are similar. While the assumption of perfect foresight is stronger for longer time horizons, our IV approach can potentially address the concern of measurement errors that arise as a result of not observing the forecasts that consumers use.

We have carried out additional analyses where we relax the assumption of perfect foresight and assume that individuals form an expectation of $PM_{2.5}$ on day $(t + 1)$ based on the information they have available at day t . Specifically, we regress $p_{i,t+1}$ on current day $PM_{2.5}$, current weather, weather on day $(t + 1)$ (assuming meteorological forecasts are available and accurate), as well as $PM_{2.5}$ that is expected to arrive on day $(t + 1)$ from surrounding cities, and use the predicted value from this regression as the forecasted $PM_{2.5}$ on the next day. The results are quantitatively similar, though slightly noisier.

Table F18: IV Cumulative Effects of Pollution: Avoidance Behavior

	Health-related Consumption					Non-health Spending	
	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
<i>Panel A: Avoidance depends on tomorrow's pollution</i>							
Pollution tomorrow	0.70*** (0.12)	0.67*** (0.11)	0.86*** (0.20)	0.61*** (0.17)	0.85*** (0.32)	-0.02 (0.14)	0.31*** (0.11)
Current + All Lags	2.51*** (0.70)	2.11*** (0.74)	2.57*** (0.90)	2.01*** (0.75)	6.25*** (2.42)	-0.56 (0.59)	-0.71 (0.48)
First-stage F	44.95	45.07	45.41	45.86	51.04	45.01	45.13
<i>Panel B: Avoidance depends on average pollution in the next 3 days</i>							
Average pollution, next 3 days	0.89*** (0.16)	0.88*** (0.16)	1.10*** (0.27)	0.80*** (0.25)	1.10** (0.51)	-0.02 (0.19)	0.47*** (0.15)
Current + All Lags	2.61*** (0.73)	2.27*** (0.77)	2.65*** (0.92)	2.16*** (0.75)	6.46** (2.55)	-0.52 (0.60)	-0.68 (0.49)
First-stage F	38.89	39.07	39.15	38.89	46.44	38.92	39.12
<i>Panel C: Avoidance depends on average pollution in the next 7 days</i>							
Average pollution, next 7 days	0.75*** (0.26)	0.86*** (0.24)	1.02** (0.46)	0.72* (0.40)	1.08 (0.81)	-0.30 (0.28)	0.62*** (0.23)
Current + All Lags	2.69*** (0.76)	2.31*** (0.81)	2.75*** (0.95)	2.11*** (0.76)	6.72** (2.79)	-0.54 (0.64)	-0.65 (0.52)
First-stage F	36.45	36.00	36.69	35.15	41.91	36.48	36.75

Notes: the dependent variable is log(number of transactions) for a given consumption category in city i on day t . We assume that individuals perfectly foresee pollution in the next few days and adjust spending accordingly. The instrument for future pollution is the same instrument for today's pollution (defined in equation (6) in the main text), leading 1 day (Panel A), 3 days (Panel B), and 7 days (Panel C). The remaining IVs and controls are identical to those in Table 3. Standard errors are in parentheses, clustered at the city level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The first-stage F-statistics are Kleibergen-Paap Wald rk F-stat that are robust to heteroskedasticity and clustered at the city level.