The Healthcare Cost of Air Pollution: Evidence from the World's Largest Payment Network

Panle Jia Barwick Shanjun Li Deyu Rao Nahim Bin Zahur*

December 2023

Abstract

This paper exploits the universe of credit- and debit-card transactions in China during 2013-2015 and provides the first nationwide analysis of the healthcare cost of $PM_{2.5}$ for a developing country. We leverage spatial spillovers of $PM_{2.5}$ from long-range transport to generate exogenous variation in local pollution and employ a flexible distributed lag model to capture semiparametrically the dynamic response of pollution exposure. We find significant impacts of $PM_{2.5}$ on healthcare spending in both the short and medium terms. A 10 μg/m^3 decrease in $PM_{2.5}$ would reduce annual healthcare spending by over $9.2 billion, about 1.5% of China's annual healthcare expenditure.

Keywords: Air Pollution, Consumer Spending, Morbidity, Healthcare Cost

JEL Classification Codes: D8, L1, L8, R2, R3

---

*Barwick: Department of Economics, University of Wisconsin-Madison and NBER (email: pbarwick@wisc.edu); Li: S.C. Jonshon College of Business, Cornell University, NBER and RFF (email: sl2448@cornell.edu); Rao: Department of Economics, School of Business and Management, The Hong Kong University of Science and Technology (email: dyrao@ust.hk); Zahur: Department of Economics, Queen's University (email: nz17@queensu.ca). 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 Minwei Tang, Meng Zhao, Xiaoting Zheng, and Nan Zhong for generous help with data access.

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] There is a limited understanding of the morbidity cost of air pollution from all health outcomes and a lack of a commonly agreed method to measure it (WHO, 2015). 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] 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).

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

[2] For example Pope (1989); Dockery (2009); Pope and Dockery (2012); Neidell (2004); Schlenker and Walker (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 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 includes 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 $PM_{2.5}$ levels 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. Because the pollution level could vary greatly across locations within a city, res-

[3] A 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 healthcare delivery in China is fragmented and hospital-centered, with little effective collaboration among institutions in different tiers of the system (Wang et al., 2018b), making it difficult to obtain consistent micro-level data on health outcomes for the whole country.

idents' pollution exposure should be measured ideally by the population-weighted local pollution. However, monitoring stations are located sparsely across space as is common in other countries, preventing us from constructing population-weighted averages at a fine geographic scale.

To deal with the endogeneity, 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 in constructing IVs, including wind patterns and other meteorological conditions in both the source and receptor cities, as well as geographic information such as distance.

Our instruments are 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 the time-invariant level of air pollution in source cities, IVs that only use wind direction in the destination city interacted with regional dummies as in Deryugina et al. (2019) and do not depend on local conditions, 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 μg/m^3 in $PM_{2.5}$ leads to 0.65% more healthcare transactions. A medium-run (i.e., three-months) increase of $PM_{2.5}$ by 10 μg/m^3 leads to 2.65% more healthcare transactions.[5] The impact of $PM_{2.5}$ differs across health facilities: spending in Children's hospitals is more than twice as responsive as spending in other types of health facilities. Non-healthcare spending experiences a negative impact of $PM_{2.5}$ in the short-term but no significant impact beyond a few weeks. In addition, the predicted worsening of air quality the next day increases current-day spending in both health and non-healthcare categories. These results provide evidence of avoidance behavior, whereby consumers reduce outdoor activities to mitigate pollution exposure.

We have conducted a host of robustness checks, including various parametric specifications of the medium-term impact, alternative approaches for constructing the instrumental variables and placebo tests, other identification strategies, more flexible controls of meteorological conditions, the inclusion of other pollutants such as CO, SO_2 and average $PM_{2.5}$ in cities in the same region, different buffer zones, alternative B-spline segments, 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 control for card penetration over time.

In monetary terms, a medium-run reduction of 10 μg/m^3 in daily $PM_{2.5}$ generates annual sav-

[4]This 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.

[5]The 90-day average $PM_{2.5}$ is 56 μg/m^3, with a standard deviation of 27μg/m^3 during our sample period.

ings in healthcare spending that exceed 59.6 billion *yuan*, or $9.2 billion.[6] This is equivalent to $22.4 per household per year. Reducing China's $PM_{2.5}$ to the World Health Organization's (WHO) annual standard of 10 μg/m^3 from the level observed in our sample period 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 in the literature? Ebenstein et al. (2017) examine the mortality impact of PM_{10} 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 PM_{10}. 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.[7]

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; 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 reduce 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

[6] We use an exchange rate of $1= 6.5 *yuan* throughout this analysis. The 95% confidence interval of the healthcare savings ranges from 4.0-115.2 billion *yuan*.

[7] 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.

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 potentially 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- 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.[8]

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 Description

Our analysis is based on three comprehensive, nationwide, micro-level datasets of air pollution, consumer spending by category, and meteorological conditions from January 2013 to December 2015, aggregated to the city by day level. These datasets enable us to evaluate the impact of air pollution on consumer 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.

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.[9] Improving air quality has become an important

[8] While semi-parametric distributed lag models have been more widely used in the epidemiology literature (e.g., Gasparrini et al. (2017)), our paper is the first to utilize this methodology in conjunction with instrumental variables to address endogeneity.

[9] Lelieveld et al. (2015) estimate that air pollution led to 1.3 million premature deaths in China in

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 (Barwick et al., 2022). The Ministry of Environmental Protection (MEP) publishes hourly measures of $PM_{2.5}$, CO, SO_2, NO_2, and O_3. 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 have 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 the 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.

Appendix Figure J1 plots the three-year average of $PM_{2.5}$ from 2013 to 2015 across cities. China's nationwide average during this period is 56 μg/m^3 (with a standard error of 46 μg/m^3), which is much higher than the annual standard of 12 μg/m^3 set by the U.S. Environmental Protection Agency and also higher than the annual standard of 35 μg/m^3 by China's MEP.[10] 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 μg/m^3 or higher. The less-developed regions in the

2010 (40% of the global total). World Bank (2007) puts the health cost of air pollution at 1.2-3.8% of China's GDP in 2003.

[10]U.S. EPA's daily standard is 35 μg/m^3 and the annual standard is 12 μg/m^3. China's MEP set limits on $PM_{2.5}$ for the first time in 2012 to take effect in 2016: the daily standard is 75 μg/m^3 and the annual standard is 35 μg/m^3.

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 J2 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 μg/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 the southern region, largely because of the coal-fired central heating systems north of the Huai River (Ebenstein et al., 2017). The pollution level has been decreasing in all regions, partly driven by tighter regulations and changes in China's industry structure (Greenstone et al., 2021).

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.[11,12] 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

[11] Of the 1.1 billion individuals above 15 (the minimum age for bank cards), 72% hold at least one bank card, with a total of 800 million bank-card holders in 2015.

[12] Merchants are classified by 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.

Pay and AliPay) was limited. The share of mobile payments in China's total retail consumption was only 8% in 2015, compared to 44% for bank cards (Kapron and Meertens, 2017).[13]

Appendix Figure J3 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. Appendix Table I1 correlates the cross-sectional card adoption rate with city demographics. Adoption is higher in cities with a higher household income and education and a younger population.

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).[14] 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 transactions in 2015. Our data account for 31% of total private healthcare spending in 2013 and the coverage rose to 51% in 2015, similar to the share of bank card transactions in other sectors.

In addition to healthcare spending, we also analyze spending in non-healthcare categories, such as daily necessities. We follow the United Nations' Classification of Individual Consumption According to Purpose (COICOP) in defining necessity goods.[15] Relative to healthcare spending,

---

[13]In comparison, spending from bank cards accounts for 55% of U.S. consumer spending in 2012 (Bagnall et al., 2014).

[14]200,000 *yuan* ($30,770) is the 99$^{\text{th}}$ percentile of transaction values across all categories. Larger transactions are excluded due to UnionPay's data protocol that aims to remove fraudulent transactions (a practice called "cash out").

[15]United Nations' COICOP defines necessity goods as 1) food and non-alcoholic beverages, 2)

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. Spending in supermarkets is over four times as large as healthcare spending in value and five times as frequent in 2015.

To graphically illustrate the relationship between pollution and spending, we plot the log number of transactions against contemporaneous $PM_{2.5}$ in Figure 1, after partialling out all other controls (weather, city trend, etc.). We group $PM_{2.5}$ (residuals) 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 different healthcare and non-healthcare categories. $PM_{2.5}$ has a positive relationship with spending in all health categories and a negative relationship with non-health spending across nearly all quantiles of $PM_{2.5}$. This suggests that 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 and low coverage ceilings that vary across insurance programs, healthcare facilities,

alcoholic beverages, tobacco and narcotics, 3) clothing and footwear, 4) recreation and culture, and 5) restaurants and hotels. We exclude supermarkets from necessity spending because they sell a large variety of goods other than necessities.

and cities (Meng and Yang, 2015). Only drugs on the National Reimbursement Drug List (maintained by the Ministry of Human Resources and Social Security) 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's 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.[16] 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). 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.

Bank card transactions do not identify disease diagnoses or treatments associated with the spending. This may raise concerns over how well the healthcare spending data correspond 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 each province from the annual China Statistical Yearbook published by the National Bureau of Statistics. This allows us to examine the correlation between our healthcare spending data and national-level healthcare statistics. Appendix Figure J4 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 Tables I2 and I3, and Figures J5 and J6 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

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

our spending data and micro-level health outcomes. 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 meteorological data from the Integrated Surface Database (ISD) 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 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.[17] 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 μg/m^3, with the inter-quartile range from 27 to 69 μg/m^3. The maximum recorded daily $PM_{2.5}$ is 985 μg/m^3. Sixty-seven percent of these city-day observations record a concentration level that is above the U.S. daily standard of 35 μg/m^3. For healthcare spending, the average daily number of transactions is 7,229 per city, and the average 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.

[17]GSOD 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.

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:

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

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. We include a rich set of controls $\boldsymbol{x}_{it}$ such as weather conditions, holiday and day-of-week fixed effects, and seasonality. We also control for city fixed effect ξ_i, city-specific linear time trend $\theta_i \cdot t$, and week-of-sample fixed effect η_w. City fixed effects control for baseline differences across cities, as more polluted cities tend to have higher health spending. City-specific time trends capture heterogeneous card adoption rates across cities, given rising card penetration in our sample period. Week-of-the-sample fixed effects allow for nation-wide temporal variation in spending and air pollution.

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 characterized by β_0, which captures responses in healthcare spending to a contemporaneous increase in pollution concentration. The long-term or cumulative impact of pollution is characterized by $\sum_{\tau=0}^{k} \beta_\tau$, which reflects changes in healthcare spending as a result of persistent elevation in past pollution exposure.

Suppose for a moment that there is no measurement error in pollution exposure $p_{i,t-\tau}$ and no omitted variables, two important issues we return to in the next section. Then the DL model can be estimated using OLS. But linear estimation with a large number of lags is undesirable due to high autocorrelation among lagged pollution $p_{i,t-\tau}$. The parameter estimates tend to be imprecise with artificial oscillations, as shown in Appendix Table I7. Alternatively, one can use the average pollution during a time window (such as the past week or month) as in the following framework:

$$y_{it} = \beta \bar{p}_{it} + \boldsymbol{x}_{it}\alpha + \xi_i + \theta_i \cdot t + \eta_w + \varepsilon_{it}, \tag{2}$$

where $\bar{p}_{it} = \frac{1}{k+1}\sum_{\tau=0}^{k} p_{i,t-\tau}$. While this specification is easy to implement and addresses the issue of high autocorrelation, it imposes a strong restriction that all lags of pollution within the window have a constant impact on spending and does not allow for dynamic time-varying 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, following Corradi (1977).[18] 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:

$$\beta_\tau = \gamma_0 + \gamma_1\tau + \gamma_2\tau^2 + \gamma_3\tau^3, \tag{3}$$

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 $\beta_\tau = \gamma_0 + \gamma_1\tau + \gamma_2\tau^2 + \gamma_3\tau^3$. Plug (3) into (1) and rearrange terms, we have:

$$y_{it} = \gamma_0 v_{1,it} + \gamma_1 v_{2,it} + \gamma_2 v_{3,it} + \gamma_3 v_{4,it} + \boldsymbol{x}_{it}\alpha + \xi_i + \theta_i t + \eta_w + \varepsilon_{it}, \tag{4}$$

where $v_{1,it} = p_{it} + p_{i,t-1} + p_{i,t-2} + \ldots + p_{i,t-k}$, $v_{2,it} = p_{i,t-1} + 2p_{i,t-2} + \ldots + kp_{i,t-k}$, $v_{3,it} = p_{i,t-1} + 2^2 p_{i,t-2} + \ldots + k^2 p_{i,t-k}$, and $v_{4,it} = p_{i,t-1} + 2^3 p_{i,t-2} + \ldots + k^3 p_{i,t-k}$, 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, $v_{2,it}, \ldots, v_{4,it}$, 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 obtain OLS or IV estimates and standard errors of γ's, we can recover β_τ's using equation

[18] Almon (1965) first proposed approximating the lag coefficients with polynomial functions. Poirier (1975) and Corradi (1977) suggested using spline functions, which impose weaker restrictions on the lag coefficients while keeping the number of parameters small. Zanobetti et al. (2000) and Schwartz (2000) apply these methods to estimate the non-linear impact of pollution on mortality. Appendix B.3 further discusses related literature.

(3), and use the delta method to estimate standard errors for β_τ's. 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.

In summary, the flexible distributed lag model transforms a series of many lagged pollution variables $\{p_{i,t-\tau}\}_{\tau=0}^{k}$ 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 geometric decay models. Third, it is straightforward to impose additional restrictions that are generated by economic theory or reflect prior knowledge of the data generating process. For example, if tomorrow's pollution should not affect current spending, then $\beta_{-1} = 0$. If pollution prior to k lags has no effect, then $\beta_{k+\tau} = 0, \forall \tau \in \mathbb{N}$. These constraints can be imposed individually or jointly and tested as linear restrictions. 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 includes 90 daily lags ($k = 90$) and characterizes the the marginal impact β_τ in each month by a separate cubic polynomial. This corresponds to a cubic B-spline with three segments, which leads to six regressors $\{v_{1,it}, \ldots, v_{6,it}\}$ and six γ parameters to be estimated. We examine robustness to different choices of lags 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 μg/m^3 in 2014, a sizable gap given the daily average of 87 μg/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 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, 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 with more local pollution in densely populated areas, the difference between the population's pollution exposure and the simple average pollution will be larger.

Another source of endogeneity is the presence of unobservables correlated with pollution. Despite our rich set of controls for weather and local conditions (e.g., city specific time trend and seasonality), there are sources of temporal variation 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, may affect both air quality and healthcare spending (and consumer activities in general).[20] These unobservables that are not absorbed by our location

[19] Satellite 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, 2021). However, there are a lot of missing values at the daily level due to cloud coverage.

[20] An 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).

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 (Zhang et al., 2015; Wang et al., 2018a). 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 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 C for more details). The pollution level of city i in time t, p_{it}, is a function of past pollution and pollution from other cities:

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

where θ_1 captures the amount of pollution that is carried over from the previous day (which is affected by local meteorological conditions), $p^{+}_{j\to i,\ t}$ 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 $\sum_{j\neq i, d_{ij}\leq r} p^{+}_{j\to i,\ t}$ (pollution imported from cities within the buffer zone) and $\sum_{j\neq i, d_{ij}> r} p^{+}_{j\to i,\ t}$ (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:

[21] Meteorological conditions play a key role in $PM_{2.5}$ diffusion (Seibert and Frank, 2003 and Wang et al., 2019).

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

where $p^{+}_{j\to i,\,t}$ 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 pollution is generated ($w_{j,t-s_{ijt}}$), and weather conditions in the destination city when pollution enters its atmosphere ($w_{i,t}$). To account for 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 $v_{j,t-s_{ijt}}$ 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 $p^{+}_{j\to i,\,t}$ is zero if $\cos(\Phi_{ji})$ is negative: when the wind blows away from city i, pollution from the source city j should not affect city i. 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 J7 illustrates graphically all subvectors of pollutants that were 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 L polynomial functions $\{u_l(d_{ij}, w_{j,t-s_{ijt}}, w_{i,t})\}_{l=1}^{L}$. The total amount of pollution imported from cities outside the buffer zone, $\hat{p}^{far}_{it}$, is the following:

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

where

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

Our instruments for current day pollution p_{it} is the set of $\{Z^{l}_{it}\}_{l=1}^{L}$. These are valid instruments since they only depend on the weather in city i at time t, which we control for in our regressions, and on pollution and weather variables in cities outside 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 the 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 a function of $p_{j,t-s_{ijt}}$, where s_{ijt} is the number of days it takes for the pollution generated in city j to arrive in city i. As such, the calculation of Z_{it} properly dates the relevant pollution source in the 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. 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 in our main specification.[22] Appendix C 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 variations in local wind direction to estimate the health impacts of particulate matter pollution

[22] A cubic B-spline with three segments has six B-spline terms and hence six endogenous variables $\{v_{1,it},\dots,v_{6,it}\}$.

(Deryugina et al., 2019). Although changes in local wind direction are more plausibly exogenous and well suited to 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 the buffer zone are uncorrelated with local shocks to spending. This assumption is violated if spending shocks (e.g., the 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.

We assume that pollution imported from regions outside the buffer zone is uncorrelated with measurement error in local pollution exposure. As discussed earlier, the difference between population-weighted average pollution and the city-wide average pollution leads to measurement error in our independent variable. Thus, changes in the measurement error over time mainly arise from within-city variation in local sources of air pollution. For example, the distribution of vehicular emissions across a city varies over time due to changing traffic patterns, leading to variation in the difference between the simple average of air pollution and the population-weighted average. By contrast, the variation in imported pollution over time from faraway, non-local sources is a function of the average pollution at the source cities and weather conditions determining the diffusion of pollutants and is unlikely to be correlated with measurement error in local pollution. Empirical evidence provided at the end of Section 4.2 corroborates this assumption.

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 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 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 Ebenstein et al. (2017) 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 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 the 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 Section 2.3, the number of transactions is a good proxy for visits to healthcare facilities. In Appendix I, 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 city-level trends in economic growth 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 Section 3.2. To assess the strength of instruments, we follow the best practice as suggested in the weak IV literature. When there is one endogenous regressor, we follow Andrews et al. (2019) and report the effective F-statistic of Olea and Pflueger (2013) (which is robust to heteroskedasticity). 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 regressions with multiple endogenous regressors.

Appendix Table I4 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 $PM_{2.5}$

[23] Weekends 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*.

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. We provide further evidence in Appendix Section C that variation in the instrumental variables systematically explains changes in the average PM2.5 levels in destination cities. For example, Figure J8 shows that for the coastal city of Shanghai whose polluting cities are located to its west, the instrumental variables correctly predict that pollution is higher on days when wind blows from the west to the east.

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.

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 I5 and I6 report the OLS and IV estimates of the short-term impacts, respectively. According to the IV estimates, a 10 μg/m^3 increase in current-day $PM_{2.5}$

[24] The 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.

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 are several times as large as their OLS counterparts, which is common in this literature (Schlenker and Walker, 2016; Ebenstein et al., 2017; Deryugina et al., 2019).[25] 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 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. Appendix Table I7 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 smooth path.

Appendix Table I8 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. A one-day surge of 10 μg/m^3 in $PM_{2.5}$ concentration increases today's transactions in all healthcare facilities by 0.03%. A medium-

[25] Schlenker and Walker (2016) use runway congestion at airports on the US East Coast as exogenous variation to examine the contemporaneous health impact of air pollution exposure for communities near large airports in California. Their 2SLS estimates are 6-10 times larger than 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 PM_{10} on mortality. Their RD estimates are 2-3 times as large as the OLS estimates. Deryugina et al. (2019) study the mortality and medical costs of $PM_{2.5}$ in the US and find the IV estimates 6 to 17 times larger than OLS estimates.

run (three-month) elevation of 10 μg/m^3 raises the number of transactions by 0.86%, eight times as large as the effect reported in Appendix Table I5 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 2. 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 μg/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 healthcare category. 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. Finally, the robust F-statistic varies from 38 to 48, suggesting that weak identification is unlikely to be a concern in our setting.

To examine how the impact on spending changes over time, Figure 2 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 aggregate health spending and spending in Children's hospitals. Consistent with Table 2, the cumulative effect increases over the 90-day window and is stronger (in percentage terms) for spending in Children's hospitals. For aggregate health spending, 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 becomes imprecise past one month. One explanation for the short-term reduction in non-health spending is budget constraints: if consumers have to spend more on heath 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 literature (Mu and Zhang, 2016; Ito and Zhang, 2018; Sun et al., 2017). Appendix E provides additional evidence that individuals engage in avoidance and that expectations about future air pollution affect current healthcare and non-health spending.

Our results suggest that a 10 μg/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 I9). These 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 since illnesses due to air pollution likely cost less to treat than other diseases on average.[26] In our analysis in Section 5.1, we use the estimates on transaction value to bound the healthcare cost of pollution.

4.2 Robustness Checks

We conduct an extensive set of robustness checks to illustrate that the results discussed above hold across different empirical specifications and choices of IVs.

Choices of Instruments We have carried out a series of robustness analyses to examine the role of the instruments. Panel A of Table I10 constructs an alternative set of IVs using the historical

[26] Average hospital spending is 6140 *yuan* ($944) for all in-patient visits versus 4109 *yuan* ($632) for in-patients treated for respiratory diseases in 2013 (National Health Commission, 2013).

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 depends on the destination city's weather. Panel B in Table I10 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%. Panel C of Table I10 drops 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.

Measurement Error and further tests of IV validity Our identification assumption is that the IVs are uncorrelated with the measurement error, which arises from local sources of pollution. Here we provide evidence that the instrumental variables primarily reflect variation in distant and non-local sources of air pollution and are uncorrelated with the measurement error.

We first examine how the strength of the first-stage differs with wind speed. If the instruments are driven by local sources of air pollution, then our first-stage should be strongest on days with lower wind speeds, since PM2.5 is more likely to be blown away from the local city to other cities on days with high wind speeds. However, Appendix Table I11 shows that the opposite is the case: the first-stage F-statistic is larger on days with higher wind speed.[27]

[27]This is similar to what Deryugina et al. (2019) find using instruments based on changes in wind direction.

Next, we divide the cities in our sample into two different subgroups, based on their contribution towards their province's total dust emissions in 2012.[28] In cities where dust emissions are high (as a share of the provincial total), local sources likely account for a larger share of overall air pollution. Thus, if our instruments primarily reflected local sources, then the first-stage would be stronger in cities where dust emissions are higher. Instead, we find that the first-stage F-statistic is smaller for cities with high dust emissions than for cities with low dust emissions (Appendix Table I12).

Additional Controls The next set of robustness analysis includes various additional controls. Panel A in Table 3 reports estimates controlling for other pollutants including O_3, SO_2, NO_2, 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 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 2 for both healthcare and non-healthcare spending categories.[29]

To address potential spillovers in regional economic activities, Panel B in Table 3 includes as a regressor the average level of $PM_{2.5}$ in period t of cities in the same region but outside the buffer zone.[30] If regional economic activities have systematic spillovers 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 cities in the same region directly controls for economic activities in other cities and delivers similar results as those in the benchmark specifications. In Appendix F, we provide

[28]2012 is the latest year for which we have dust emissions data.

[29]The 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.

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

further evidence that our results are unlikely to be biased by spillovers in economic activities. There exists little correlation between local economic activity and economic activity in cities outside the buffer zone once we include the full set of controls and fixed effects. Our main estimates are also robust to controlling for economic activity in regions outside the buffer zone.

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. Panel C of Table 3 further 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 $PM_{2.5}$.

Finally, Appendix Table I13 shows that adjusting for population size by using the number of transaction per capita as the dependent variable leads to similar results.

Specifications Using Average Pollution Our flexible distributed lag model delivers smooth marginal impact estimates of past pollution on current-day spending. The robustness checks presented in Table I20 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; Schwartz, 2000). On the one hand, the impact of air pollution on the 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 2 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 over time.

Nonetheless, Appendix Table I14 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 2, 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. Second, restricting lagged pollution to have a constant impact on health spending could either overestimate or underestimate the true effect.

Additional Robustness Checks Appendix D discusses several additional robustness checks. Our results are robust to different choices of lag length, B-spline segments, and buffer zone radius in constructing the IVs. We conduct a placebo test that randomizes wind direction and speed, showing that without the exogenous variation provided by changes in wind direction and speed, the IVs become weak. We also report results using alternative IVs proposed in the literature that rely only on wind direction in destination cities. These IVs also suffer from a weak first-stage, which makes it difficult to identify the medium-term impact of air pollution. Finally, our results are robust to the inclusion of more flexible controls of meteorological conditions and that the effects of air pollution are similar for cities where pollution monitoring began earlier.

In sum, our extensive list of robustness checks confirms that the results in our main specification are not driven by particular functional form assumptions or a specific set of instruments. Rather, our results hold across a myriad of specifications we have examined. Next, we examine effect heterogeneity across different pollution and income levels.

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 in developed countries that are 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 (World Bank, 2007). Despite its important implications, there is a lack of empirical evidence on the nature of 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 $PM_{2.5}$ for a wide range of pollution levels.

To capture nonlinearity, we include the quadratic term of $PM_{2.5}$ in addition to its linear form.[31] Appendix Figure J9 plots the cumulative marginal effect ($\sum_\tau \beta_\tau$) over three months (as well as one and two months) 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 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 J10 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 to 53,000 *yuan* across cities, with an average of 25,530 *yuan*. Pollution's impact on healthcare is largest in poor cities and smaller in richer cities. This may be driven by 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 economically meaningful, the evidence is suggestive given the statistical insignificance of income coefficients (Appendix Table I15).

We have also estimated heterogeneity across seasons and years (Appendix Table I16). 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

[31] To 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). The impact of past pollution $p_{i,t-\tau}$ is defined as: $\beta_\tau(\tau, w|\boldsymbol{\gamma}, \boldsymbol{\sigma}) = (\sigma_1 w + \sigma_2 w^2) + (\gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \gamma_3 \tau^3)$ where $\sigma_1 w + \sigma_2 w^2$ captures heterogeneity and allows the intercept of β_τ to vary across different levels of w. The rest of the estimation follows the linear model specified in Equation 4. Appendix Table I15 reports coefficient estimates.

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. 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 to reduce exposure to air pollution, as shown in Appendix E. 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 I17. Our preferred specifications show that a 10 $\mu g/m^3$ increase in $PM_{2.5}$ would lead to a 2.65% increase in the number of health-related transactions (Table 2) and a 1.5% increase in transaction value (Table I9) 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-$PM_{2.5}$ masks. Using the piecewise linear relationship between $PM_{2.5}$ and AQI, this means that exposure to 10 $\mu g/m^3$ more $PM_{2.5}$ 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 $\mu g/m^3$) change in $PM_{2.5}$ leads to 8.3% more spending on asthma and COPD, which is equivalent to a 22% increase for 10 $\mu g/m^3$ more $PM_{2.5}$. According to China's National Health Commission

(2013), spending on respiratory diseases accounted 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 $PM_{2.5}$. Assuming that the health impact is the same for both bank-card and non-bank-card spending (see discussion in Appendix H), a 10 μg/m^3 increase in $PM_{2.5}$ translates to 59.6 billion *yuan* ($9.2 billion) additional healthcare spending in 2015, with a 95% confidence interval of 4.0 - 115.2 billion *yuan*.[32] This estimate is much larger than existing estimates in policy discussions. For example, OECD (2016) estimates that $PM_{2.5}$ and ground-level ozone are associated with a $20 billion direct cost on health expenditures worldwide due to morbidity 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 $PM_{2.5}$ (56 μg/m^3 on average) relative to WHO's recommended level of 10 μg/m^3 leads to $42 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) downward endogeneity bias in the dose-response function, (2) inherent differences in the dose-response function across countries, and (3) 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 μg/m^3 increase in $PM_{2.5}$. 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,

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

2012). 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 μg/m^3 increase in $PM_{2.5}$. 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 μg/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 mortality estimate. First, we rely on the benefit transfer approach to estimate the VSL for the Chinese population, due to the lack of a national-level estimate of VSL for China. Second, we adjust the VSL for each age group. Appendix H provides further details.

Our analysis implies that a 10 μg/m^3 increase of $PM_{2.5}$ would generate a mortality impact of \$13.4 billion in 2015 in China (Appendix Table I18). 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.

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 often focuses on the impacts of air pollution in developed countries. This study provides the first comprehensive analysis of the direct healthcare cost of $PM_{2.5}$ in a developing country context based on high-resolution data from the world's largest payment network.

To address potential endogeneity in the measurement of pollution exposure, 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 $PM_{2.5}$. We propose a 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 μg/m^3 decrease in $PM_{2.5}$ would lead to at least a $9.2 billion reduction in healthcare spending annually, or 1.5% of China's national annual healthcare spending. The estimated healthcare cost exceeds two-thirds of the mortality cost based on the recent literature. China's elevated $PM_{2.5}$ 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 global 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

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, 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.

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, Liguo Lin, and Eric Zou, "From Fog to Smog: the Value of Pollution Information," 2022. Conditionally accepted at *American Economic Review*.

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.

Brunekreef, Bert and Stephen T Holgate, "Air Pollution and Health," *Lancet*, 2002, *360*, 1233–1242.

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.

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.

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*, 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.

Dockery, Douglas W., "Health Effects of Particulate Air Pollution," *Annals of Epidemiology*, 2009, *19*, 257–263.

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, Jonathan Levin, and Neel Sundaresan, "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.

Gasparrini, Antonio, Fabian Scheipl, Ben Armstrong, and Michael G Kenward, "A penalized framework for distributed lag non-linear models," *Biometrics*, 2017, *73* (3), 938–948.

Greenstone, Michael and Rema Hanna, "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India," *American Economic Review*, 2014, *104*, 3038–3072.

— **, Guojun He, Shanjun Li, and Eric Yongchen Zou**, "China's War on Pollution: Evidence from the First Five Years," *Review of Environmental Economics and Policy*, 2021, (2), 281–299.

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.

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.

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.

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.

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.

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.

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.

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.

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," *At-*

mospheric 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.

World Bank, *Cost of Pollution in China: Economic Estimates of Physical Damages*, The World Bank, 2007.

Yu, Hao, "Universal health insurance coverage for 1.3 billion people: What accounts for China's success?," *Health Policy*, 2015, *119* (1), 1145–1152.

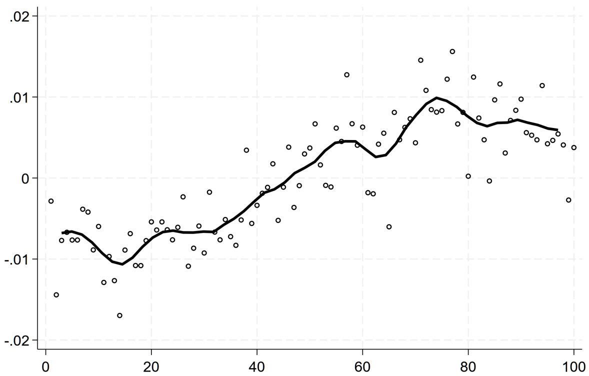
Zanobetti, A, MP Wand, J Schwartz, and LM Ryan, "Generalized additive distributed lag models: quantifying mortality displacement," *Biostatistics*, 2000, *1* (3), 279–292.

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).

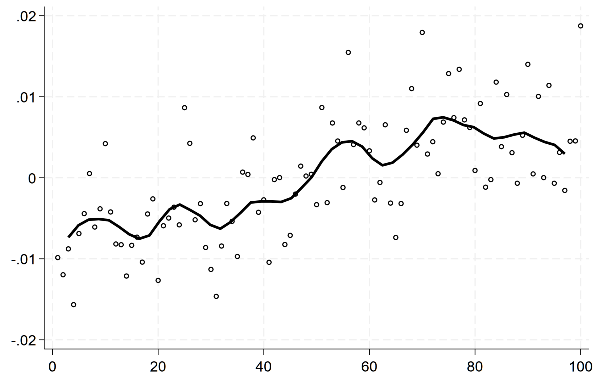
Zou, Eric, "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality," *American Economic Review*, 2021, *111* (7), 2201–26.

Figure 1: Residualized Plot of Log Number of Transactions v. Percentiles of $PM_{2.5}$ Concentration

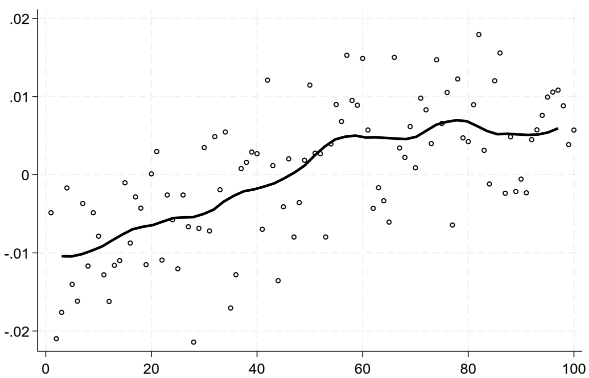
(a) Total Healthcare Industry



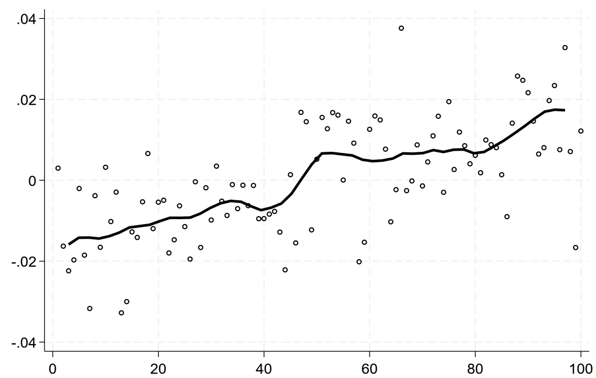
(b) Pharmacies



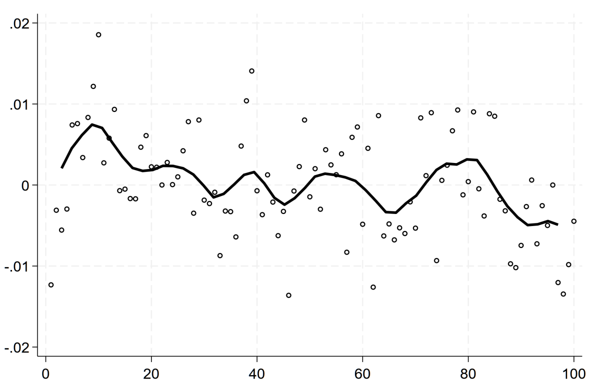
(c) People's Hospitals



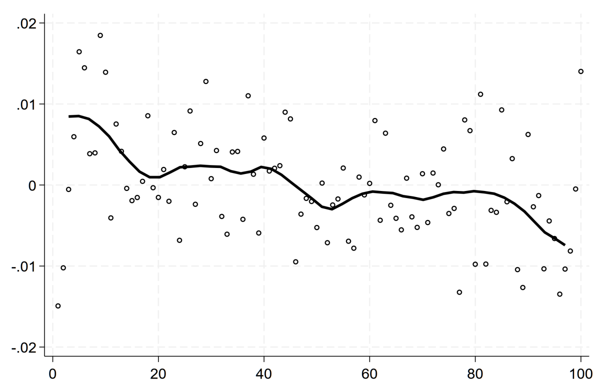
(d) Children's Hospitals



(e) Necessities



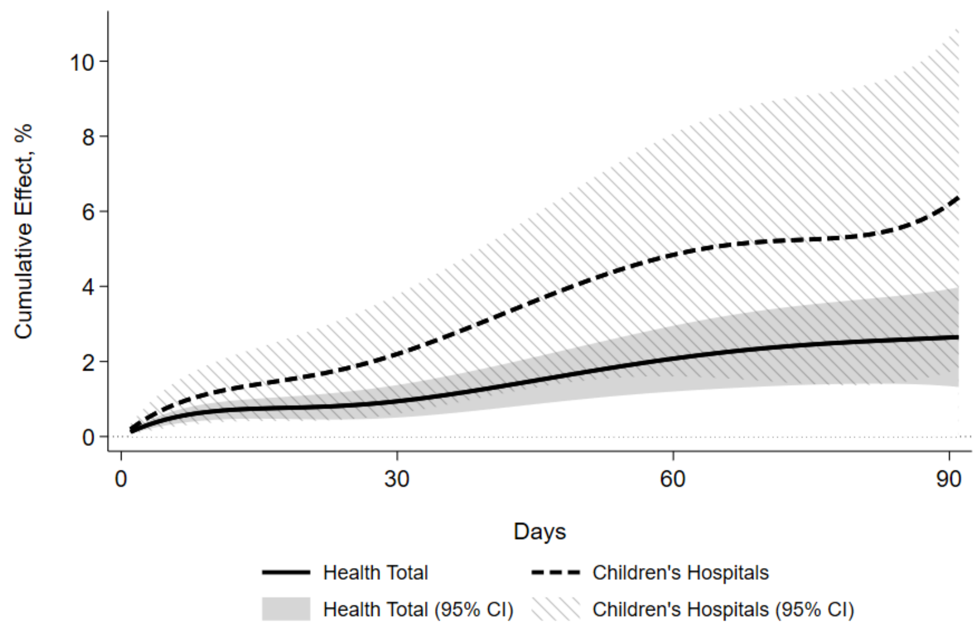
(f) Supermarkets



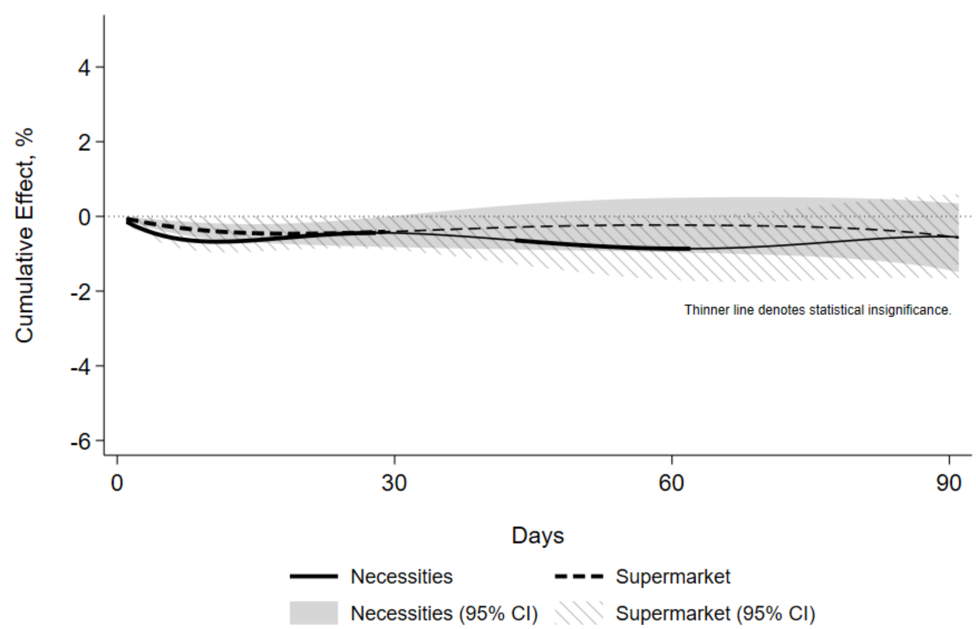
Notes: Each dot denotes the in-group average residuals, partialing out city FEs, weekly FEs, city-specific time trends, city-specific seasonality, day-of-week FEs, dummies for holidays and working weekends, and weather controls (temperature, precipitation, wind speed). Groups are binned by percentiles of $PM_{2.5}$ residuals, depicted by the x-axis. Figure J12 provides an alternative version of the graph by plotting the raw values of $PM_{2.5}$ residuals on the x-axis, which better demonstrates the linearity of the impacts.

Figure 2: 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 μg/m^3 increase in $PM_{2.5}$ 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 μg/m^3 increase in $PM_{2.5}$ concentration over the past 28 days leads to 2.12% more transactions in Children's hospitals but 0.41% fewer transactions in supermarkets. Thick lines (and thick segments) indicate significance at the 5% level. Thinner lines indicate that the impact is statistically insignificant at the 5% level. Shaded areas are 95% confidence intervals.

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, *from 1% card sample*					
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, *from 1% card sample*					
Daily Necessities	236.9	551.3	0	9,532.4	211,318
Supermarkets	232.8	643.4	0	14,404.7	210,493

Table 1: Summary Statistics

	Mean	Std. Dev.	Min.	Max.	N
Weather					
Mean Temperature, $^{\circ}F$	60.1	18.9	-27.5	101.6	211,317
Precipitation, *inch*	0.1	0.4	0	15.6	205,446
Mean Wind Speed, *mph*	5.5	3.1	0	48.7	211,296
Wind Direction, *navigational bearing*	-	-	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, 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***	0.12***	0.07*	0.14***	0.19***	-0.14***	-0.06***
	(0.02)	(0.03)	(0.04)	(0.04)	(0.07)	(0.03)	(0.02)
Current + Past 3d	0.40***	0.40***	0.23*	0.47***	0.65***	-0.45***	-0.21***
	(0.07)	(0.08)	(0.12)	(0.13)	(0.23)	(0.09)	(0.07)
Current + Past 7d	0.61***	0.62***	0.39**	0.75***	1.04***	-0.64***	-0.34***
	(0.10)	(0.12)	(0.18)	(0.19)	(0.36)	(0.13)	(0.10)
Current + Past 14d	0.74***	0.75***	0.57***	0.97***	1.40***	-0.63***	-0.45***
	(0.14)	(0.16)	(0.21)	(0.22)	(0.50)	(0.16)	(0.12)
Current + Past 28d	0.91***	0.90***	0.99***	1.24***	2.12***	-0.44*	-0.41**
	(0.22)	(0.25)	(0.30)	(0.27)	(0.79)	(0.23)	(0.21)
Current + Past 56d	1.97***	1.71***	2.31***	2.01***	4.65***	-0.85**	-0.23
	(0.42)	(0.47)	(0.54)	(0.46)	(1.56)	(0.41)	(0.36)
Current + All Lags	2.65***	2.18***	2.80***	2.13***	6.37***	-0.55	-0.57
	(0.68)	(0.71)	(0.89)	(0.75)	(2.33)	(0.58)	(0.47)

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

	Health	All Hospitals	Pharmacy	People's	Children's	Necessities	Supermarket
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 μg/m^3 increase in $PM_{2.5}$ over the corresponding period, $\sum_{\tau=0}^{k} \beta_\tau$, estimated via the IV version of the flexible distributed lag model with 90 lags. 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). The IVs are interactions of pollution transported from distant source cities (150 km away) and meteorological conditions in the source and destination cities as defined in Equation (6) and Section 3.2.2. Standard errors 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 3: 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)
Panel A: *Controlling for other pollutants*							
Current + All Lags	2.55***	2.07***	2.73***	2.01***	6.21***	-0.55	-0.69
	(0.69)	(0.72)	(0.91)	(0.76)	(2.34)	(0.58)	(0.46)
First-stage F	39.76	39.85	39.75	41.61	50.98	39.71	39.71
Panel B: *Controlling for economic spillover*							
Current + All Lags	2.62***	2.15***	2.76***	2.12***	6.37***	-0.56	-0.56
	(0.68)	(0.72)	(0.89)	(0.76)	(2.34)	(0.59)	(0.47)
First-stage F	34.02	34.09	34.01	35.59	45.96	33.94	34.00
Panel C: *Controlling for card adoption*							
Current + All Lags	2.60***	2.14***	2.75***	2.10***	6.31***	-0.56	-0.59
	(0.69)	(0.73)	(0.90)	(0.74)	(2.37)	(0.56)	(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 μg/m^3 increase in $PM_{2.5}$ 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 2. In addition to controls in Table 2, 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 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.