Trans-Boundary Air Pollution Spillovers:
Physical Transport and Economic Costs by Distance*

Shihe Fu
Xiamen University
fushihe@xmu.edu.cn

V. Brian Viard
Cheung Kong Graduate School of Business
brianviard@ckgsb.edu.cn

Peng Zhang
School of Management and Economics, The Chinese University of Hong Kong, Shenzhen
and
Shenzhen Finance Institute
zhangpeng@cuhk.edu.cn

This version: 12/8/2021

Abstract

The economic costs of trans-boundary pollution spillovers versus local effects is necessary to evaluate centralized versus decentralized environmental policies. Directly estimating these for air pollution is difficult because spillovers are high-frequency and vary with distance while economic outcomes are usually measured with low-frequency and local pollution is endogenous. We develop an approach to quantify local versus spillover effects as a flexible function of distance utilizing commonly-available pollution and weather data. To correct for the endogeneity of pollution, it uses a mixed two-stage least squares method that accommodates high-frequency (daily) pollution data and low-frequency (annual) outcome data and can improve efficiency. We estimate spillovers of particulate matter smaller than 10 micrograms (PM_{10}) on manufacturing labor productivity in China. A one μg/m^3 annual increase in PM_{10} locally reduces the average firm's annual output by CNY 45,809 (0.30%) while the same increase in a city 50 kilometers away decreases it by CNY 16,248 (0.11%). This effect declines rapidly to CNY 2,847 (0.02%) for an increase in a city 600 kilometers away and then slowly to zero at 1,000 kilometers. The results suggest the need for supra-provincial environmental policies or Coasian prices quantified under the approach.

JEL Codes: D62; Q51; Q53; R11
Key words: air pollution; spillovers; environmental costs and benefits, mixed two-stage least squares; regional coordination

* All authors contributed equally and are ordered alphabetically. We thank two anonymous referees, the editor, Michael Anderson, Richard Arnott, Tom Chang, Tatyana Deryugina, Andrew Foster, Qing Han, Ying Liu, Dick Morgenstern, Adriana Lleras-Muney, Ruixue Jia, Alberto Salvo, Nathan Schiff, Katja Seim, and seminar and conference participants at UC Riverside, 2019 AREUEA International Conference, Shanghai University of Finance and Economics, Nankai University, Jinan University, Wuhan University, Hunan University, 2019 Asian Meeting of Econometric Society, and National University Singapore for helpful comments. Quanyun Song, Jianghao Wang, Castiel Zhuang, and Shihan Shen provided excellent research assistance. Shihe Fu acknowledges financial support by the National Natural Science Foundation of China (Grant # 72173105 and Basic Scientific Center Project 71988101). Peng Zhang acknowledges financial support by the National Natural Science Foundation of China (Grant # 72033005 and 71822402).

1. Introduction

Since the seminal work of Oates (1972) on fiscal federalism, there has been a debate on whether centralized or decentralized policies achieve the most efficient outcome. Local authorities have better information about costs and benefits and can better tailor local policies than central authorities whose policies tend to be overly uniform. However, local jurisdictions generally ignore their policies' effects on other jurisdictions unless these are internalized administratively. Clear and enforceable assignment of property rights followed by Coasian bargaining can also solve these externalities even under decentralized control (Coase, 1960) but require quantification of the extra-territorial damages incurred as a function of distance.

Despite this question, we are not aware of any studies that quantify trans-boundary spillovers relative to local effects for any kind of pollution. Previous papers show that trans-boundary pollution spillovers exist and that they affect extra-territorial economic well-being but they do not quantify spillovers compared to local effects as a function of distance. Our paper aims to fill this gap by providing an approach for estimating an air pollution spillover gradient including local effects for endogenous economic outcomes.

Air pollution is a prototypical example of the fiscal federalism debate with serious welfare implications. High air pollution levels in developing countries have led to adverse effects on health, economic output, and physical and mental comfort. Ninety-two percent of all air pollution-related deaths are estimated to occur in low- and middle-income countries and ambient air pollution is estimated to have generated costs equal to 4.4% of global GDP in 2016 (Ostro, *et al.*, 2018). Air pollution levels exceed the social optimum because spillovers, including trans-boundary, are not internalized. Developed countries also recognize the need to manage cross-boundary pollution. For example, the U.S. Clean Air Act Section 126 allows a downwind state to petition the Environmental Protection Agency to take action against an upwind state that impedes its ability to comply with smog standards.[1]

Regardless of the method used to correct the externality, a necessary input is the magnitude of the spillovers by distance. Setting the geographic scope of centralized decision-making to internalize spillovers requires knowledge of how far spillovers extend at significant levels. Alternatively, assigning property rights and allowing for decentralized Coasian bargaining requires a method for the parties to estimate the damages from spillovers based on their origin. Quantifying air pollution spillovers requires estimating not just the quantity of pollution that drifts as a function of distance but also the economic costs that it imposes upon arrival. Finally, to establish

[1] Described at https://www.epa.gov/ground-level-ozone-pollution/ozone-national-ambient-air-quality-standards-naaqs-section-126.

the geographic reach of centralized control or to evaluate Coasian prices, it is necessary to quantify local effects which can be compared to spillover damages.

If daily pollution, weather, and outcome data are available, estimating the effect of spillovers on the outcome is straightforward: a reduced form estimate of imported pollution on local economic outcomes. However, many economic outcomes are measured at a low frequency (e.g., annual) and air pollution spillovers occur with high frequency (daily). Aggregating data to the annual level and relating economic outcomes to imported pollution is likely to involve significant efficiency losses as occurs in our application. In addition, reduced-form estimates do not quantify the effect of local pollution on local productivity. We develop an approach to overcome these obstacles and demonstrate it by estimating the effects of air pollution spillovers on annual manufacturing labor productivity in China.

Our approach is inspired by the fact that the reduced-form effect equals the intensity of treatment (the effect of nearby- on focal-city pollution) multiplied by the causal effect of focal-city pollution on focal-city productivity. Letting P^n represent nearby-city pollution, P^f focal-city pollution, and Y^f the focal-city outcome (in our case productivity):

$$spillover\ of\ P^n on\ Y^f = \\ (pollution\ decay\ function\text{:}\ effect\ of\ P^n\ on\ P^f) \times (causal\ effect\ of\ P^f\ on\ Y^f).^{2} \quad (1)$$

The right-hand side of Equation (1) contains the two determinants of the trans-boundary effect of pollution on an outcome: how much air pollution is physically transported across cities and the causal effect on the outcome when it arrives in the destination city.

We therefore proceed in two steps. In the first step, we estimate the pollution spillover (which we call the pollution decay function) of nearby- on focal-city pollution flexibly as a function of distance using *daily* data conditional on wind blowing toward the focal city. In the second step, we estimate the causal effect of focal-city air pollution on the economic outcome. Multiplying the spillover decay effects from the first step by the causal effect from the second step provides an estimate of spillover effects on the outcome[3] which can be compared to the local effect estimated in the second step. An advantage of separating the two steps is that

[2] This is because the causal effect estimated via 2SLS using nearby-city pollution as an instrument is the reduced-form effect divided by the first stage effect (Angrist and Pischke, 2015: 107):
$causal\ effect\ of\ P^f\ on\ Y^f = \frac{(spillove\quad of\ P^n\ on\ Y^f)}{(pollution\ decay\ function\text{:}\ effect\ of\ P^n\ on\ P^f)}$.

[3] Although the spillover decay function is estimated at the daily level, the effects can be interpreted as the annual effects of a sustained and uniform increase in nearby-city pollution on all days of the year if wind blew toward focal cities on all days. Since the wind blows toward focal cities roughly half the time on average, annual spillovers are roughly half the daily effect as we describe in our results.

the first step relating nearby- to focal-city pollution can involve highly nonlinear functions of pollution, wind patterns, and weather while preserving the linear relationship necessary for instrumenting in the second step.

When we estimate the causal effect of pollution in the second step, we instrument for the endogeneity of focal-city air pollution using the air quality of the nearest nearby city conditional on wind blowing toward the focal city. Although other instruments could be used, this instrument is convenient because the required data (daily pollution and wind measures) are commonly available and are already used to estimate the pollution decay function in the first step. The exogeneity of this instrument requires high-frequency data for two reasons. First, to capture wind direction shifts precisely enough and, second, to preclude confounding factors affecting both nearby-city pollution and focal-city economic outcomes that might occur over longer time periods (in particular inter-regional economic shocks).[4] We provide evidence that daily data are frequent enough but higher levels of aggregation are not.

To combine the daily instrumenting data with the annual outcome data, we employ mixed two-stage least squares (M2SLS) (Dhrymes and Lleras-Muney, 2006), a methodology for implementing 2SLS with different levels of aggregation in the two stages. While the daily instrumenting data can be annualized (conditional on wind direction) and Wald 2SLS applied, we show in our application that this results in inefficient estimates relative to M2SLS. The latter allows within-year controls to be included in the first stage generating more precise predicted values of the instrument.

We apply this approach to estimate the effect of trans-city drifts of particulate matter less than 10 micrograms in diameter (PM_{10}) on short-run manufacturing labor productivity in China using a large firm-level data set from 2001 to 2007. A one $\mu g/m^3$ annual increase in PM_{10} in a city within 50 kilometers decreases the annual labor productivity of an average firm in the focal city by CNY 16,248 (0.106%).[5] This effect declines quickly to CNY 2,847 (0.019%) for nearby cities at 550-600 kilometers after which it declines slowly to zero at about 1,000 kilometers. This compares to a local effect of CNY 45,809 (0.300%). The spillover is roughly 35.5% of the local effect at 50 kilometers, falling to 6.2% at 550 kilometers, and zero at 1,000 kilometers and beyond. While we demonstrate the estimation approach with PM_{10} and productivity, it can be easily tailored to estimate the spillovers for other pollutants and other annual outcomes such as GDP, morbidity, and mortality.

[4] Exogeneity also requires that wind direction is random with respect to nearby-city pollution conditional on control variables. We provide evidence that this is the case.
[5] This estimate is for the average city and average weather.

This paper contributes to three strands of literature. First, we quantify the magnitude of spillovers as a function of distance relative to local effects, a key input in choosing between centralized and decentralized environmental policies (Oates and Schwab, 1988; Ogawa and Wildasin, 2009; Banzhaf and Chupp, 2012; Eichner and Runkel, 2012; Williams, 2012; Fell and Kaffine, 2014). Extant work on trans-boundary spillovers either shows that trans-boundary pollution spillovers exist (Sigman, 2002; Sigman, 2005; Kahn *et al.*, 2015; Cai *et al.*, 2016; Lipscomb and Mobarak, 2017; He *et al.*, 2020; Wang and Wang, 2021) or that they affect extra-territorial economic well-being (Zheng *et al.*, 2014; Bošković, 2015; Altindag *et al.*, 2017; Sheldon and Sankaran, 2017; Jia and Ku, 2019) but do not quantify their extensiveness or size relative to local effects. Goodkind *et al.* (2019) estimate the pollution decay function for a different pollutant by a different method and use an exposure-response method to estimate health costs.

Second, we develop an approach based on M2SLS that allows high-frequency variation in wind direction to be used as an instrument for high-frequency air pollution in estimating its causal effect on low-frequency outcomes. There are two approaches to using wind direction as an instrument. One approach uses dominant wind direction alone without measures of non-local pollution sources (Deryugina *et al.*, 2019; Freeman *et al.*, 2019; Herrnstadt *et al.*, 2019; Anderson, 2020). This is convenient because the instrument is valid without the need to measure non-local pollution. The downside, as Deryugina *et al.* (2019) points out, is that the monitoring stations that measure local pollution must be geographically dense enough to avoid measurement error and confounding effects from local pollution sources.[6] The second approach combines wind direction with the extra identification from non-local pollution. The advantage of this is that it is not confounded by local pollution sources and can be used in the absence of a dense network of local monitoring stations. The downside is that non-local pollution sources must be measured and must be orthogonal to local sources. Previous papers that use this approach (Schlenker and Walker, 2016; Rangel and Vogel, 2019) use discrete, exogenous events that shift non-local pollution. Our paper adapts this approach to use a continuous measure of non-local pollution and allows for the instrument to be of higher frequency than the endogenous variable.

Since productivity is often serially correlated across days, we extend the approach to accommodate this in a way that does not require additional data. Controlling for

[6] As they explain, having a dense network of local monitors averages out the effects of local pollution sources so that they do not bias estimates. Slightly modifying their example (page 14) imagine a smokestack in the middle of a city. If there is a single monitor on the east side of the city then the monitor will detect the pollution from the smokestack when the wind is blowing from the west but not when it blows from the east and the wind direction instrument is correlated with local pollution. However, if there is a dense network of monitors on all sides of the smokestack then a local pollution measure averaged across all monitors will reduce, and in the limit, eliminate this correlation.

autocorrelation does not materially change the results. While we apply this extension to productivity, it can be applied in any setting in which the endogenous outcome variable exhibits high-frequency autocorrelation.

Third, our paper adds to the growing literature on estimating air pollution's effect on labor productivity (Graff Zivin and Neidell, 2012; Chang *et al*., 2016; Chang et *al*., 2019; He *et al*., 2019; Fu *et al*., 2021). These papers estimate the effect of an increase in local air pollution on local firms' productivity. In contrast to these papers, we distinguish the effect of local and imported pollution sources on productivity and show that spillovers can contribute significantly to productivity losses.

We find that pollution exerts a substantial negative effect on productivity even at relatively far distances. Twenty-two percent of PM_{10} produced from a city within 300 kilometers is imported into a focal city when the wind blows directly toward it. From a policy perspective, to internalize this would require centralized control of administrative areas that are 300 kilometers in radius or 283-thousand square kilometers in size. This is larger than many medium-sized provinces in China such as Hunan, Shaanxi, Hebei, Jilin, Hubei, and Guangdong (Ministry of Civil Affairs, 2017). This implies that environmental policies need to be coordinated at the supra-provincial level to internalize spillovers. The other major policy application of our method is in calculating Coasian prices as a decentralized solution to air pollution externalities. Our estimates allow a quantification of the compensation that one city must make to another to internalize inter-city pollution damage given the distance between the two cities, the annual wind-direction distribution, and annual levels of the economic outcome of interest.

The scientific literature uses an alternative approach for the first step of our procedure, chemical-transport models or CTMs, to relate source emissions to receptor concentrations (Moussiopoulos, *et al.* (1996); Seigneur and Moran (2004); Seigneur and Dennis (2011)). CTMs that estimate this relationship over long distances such as we do are models that relate locations defined by three-dimensional grids normally one kilometer or larger in size.[7] The relationships are based on mathematical models of atmospheric processes using detailed weather and emissions data. As an alternative for the first step of our procedure, CTMs offer advantages and disadvantages relative to our approach.

CTMs quantify the spillovers from emissions and is unaffected by their displacement unlike our approach which relies on concentrations. CTMs require highly disaggregated geographic data on weather and emissions which are often not available, especially in developing economies; while concentrations are more

[7] The other approach, source-specific models, identifies specific emissions sources that contribute to ambient concentrations but apply up to only about 150 kilometers.

broadly available. CTMs realistically model the processes of concentration formation and movement; however this greater complexity involves longer solutions times and more assumptions. In policy setting, agreeing upon these assumptions can require significant effort and resources.[8] In contrast, our approach can be estimated quickly and its transparency requires agreement on fewer assumptions.

Our results have specific implications for the role of China's governance system in managing air pollution spillovers. China's reforms have succeeded in part because of its regionally decentralized system in which the central government provides incentives to local governments based primarily on local GDP to the exclusion of other criteria (Jin *et al*., 2005; Li and Zhou, 2005; Xu, 2011). Our results imply that these incentives exacerbate the negative implications of air pollution spillovers on manufacturing productivity. This complements Jia (2017) which provides empirical evidence that these incentives result in more pollution. Including local environmental quality in local government officials' performance valuation is not enough; regional coordination of cross-boundary pollution spillovers must be considered.

The remainder of the paper proceeds as follows. The next section describes the data we use to illustrate the estimation approach and Section 3 the approach. Section 4 provides the results, and Section 5 concludes.

2. Data

We estimate pollution spillovers on labor productivity for manufacturing firms in China from 2001 to 2007 in two steps. The first step (estimating the pollution decay function) requires daily pollution and weather data. The second step (estimating the causal effect of air pollution on productivity) requires daily data for the instrument to address the endogeneity of pollution and accommodates annual data on the outcome variable (productivity).

2.1 Pollution data

The highest-frequency pollution data available with significant geographic coverage during our sample period is the daily Air Pollution Index (API) published by the Ministry of Ecology and Environment. This is available at the city level and only for larger cities. The number of cities reporting API data increases over time in the sample. The sample includes 60 unique cities (Appendix A shows their location).

The API ranges from 0 to 500 with higher values indicating higher pollution concentrations and more harmful health effects (Andrews, 2008). During the sample

[8] For example, the US EPA devotes significant resources in choosing models that meet their standards via conferences, technical analyses, and regulatory reports. A recent example is detailed in Federal Register (2017).

period, a city's daily API reports the worst of three pollutants: particulate matter (PM_{10}), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) whose concentrations are measured at monitoring stations within the city. Each is rescaled as an API measure to make them comparable and the pollutant with the maximum API is reported.[9] The identity of the maximal pollutant is reported if the API exceeds 50.

The API is potentially subject to manipulation by those who collect and report the data. Using 2001 to 2010 data, Ghanem and Zhang (2014) find a discontinuity in the API distribution around 100 which suggests that self-reported data is manipulated by local officials who are evaluated on the annual number of "Blue Sky" days (those below 100). Also consistent with this, Andrews (2008) finds that a significant number of days in 2006 and 2007 with reported API values between 96 and 100 would fall in the range 101 to 105 if calculated using the underlying monitoring station data. To avoid any possible bias in the estimates we exclude days when the API is between 95 and 105 in either the focal or nearby city in the main estimates but show that it is robust to including these.

We use PM_{10} in the analysis rather than the API index because we wish to use physical pollution levels in quantifying spillovers and PM_{10} is overwhelmingly the worst of the three pollutants (about 90% of days). We drop days in which PM_{10} is not the maximal pollutant and for the remaining days infer its value from the API based on the piecewise-linear relationship between PM_{10} and the API (Appendix B). Although we do not observe the worst pollutant when the API is below 50 we assume it is PM_{10} in the baseline estimates because at these low levels air quality is assumed to be safe regardless of pollutant. The results are robust to dropping these days.

2.2 Wind and weather data

We require daily wind data for estimating the spillover decay function and to instrument pollution when estimating its effect on productivity. We use station-level wind direction data from the World Weather Records Clearinghouse collected by the U.S. National Oceanic and Atmospheric Administration (NOAA).[10] The data provide a direction from which the wind is blowing stated in degrees clockwise from true North in each three-hour period of each day in each city. We use a "unit-vector" average method defined by the NOAA to arrive at an average daily wind direction for each city.[11] For wind direction we use data for the focal not the nearby city.

[9] Each monitoring station records the concentrations of the three pollutants multiple times a day. Each of these intra-day measurements is rescaled to an API index. A daily mean API for each pollutant across all stations in a city is then calculated and the maximum of these three means is the city-level API for that day.
[10] Data available at: http://www.ncdc.noaa.gov/data-access.
[11] In each three-hour period, we convert the direction for each monitoring station to a unit vector with coordinates $\langle u, v \rangle$. The u-component is the North-South wind direction and v the East-West. We average the

Regardless of the wind direction in the nearby city, pollution cannot be imported if the wind in the focal city is not blowing from the nearby city's direction.

To control for weather conditions that affect the transport of pollution and productivity we use daily weather (humidity, windspeed, and temperature) data downloaded from the Weather Underground.[12]

2.3 Firm productivity data

Our firm-level output and characteristics data are from annual surveys of manufacturing firms conducted by China's National Bureau of Statistics (NBS). The survey includes all state-owned enterprises (SOEs) regardless of size and all non-SOEs whose annual sales exceed CNY 5 million (USD 0.8 million).[13] Our results therefore do not necessarily extrapolate to non-SOE, below-scale firms. The survey also contains detailed information on firm location, accounting measures, and firm characteristics.[14] Before we match with the pollution data this captures 90.7% of China's total manufacturing output during the sample period (Brandt *et al*., 2012). We follow Brandt *et al.* (2012) in matching firms over time to form an unbalanced panel and in converting nominal into real values using industry-level price indices. To be consistent with the previous literature, we drop observations with missing or unreliable data (Cai and Liu, 2009; Brandt *et al*., 2012; Yu, 2014) and winsorize the top and bottom 0.5% of data based on each of the values of output, value added, employment, and capital (Cai and Liu, 2009).

We measure output as value added per worker which is common in the productivity (Syverson, 2011; Brandt *et al*., 2012) and temperature-productivity literature (Hsiang, 2010; Dell *et al*., 2012). Firms report value added directly in the data and it equals total production (including both sales and inventory) of all goods produced in the year valued at their market prices less the cost of all intermediate inputs employed in producing them. Using aggregate measures of productivity requires that prices do not reflect market power in either the primary or upstream input markets. We cannot guarantee this; however, nearby-city pollution is independent of firm-level market power in the focal city allowing us to consistently estimate pollution's effect on productivity via instrumented pollution. The mix of products is also not discernible from firm-level value added and may be correlated with local pollution levels. However, our instrumenting strategy also addresses this issue: nearby-city

two coordinates separately across the periods of each day and all stations to yield $\bar{u}$ and $\bar{v}$. We then translate the direction into a 0 to 360 degree scale based on the signs of $\bar{u}$ and $\bar{v}$: $180-\theta$ if $\bar{u}<0$ and $\bar{v}>0$, $\theta-180$ if $\bar{u}<0$ and $\bar{v}<0$, $360-\theta$ if $\bar{u}>0$ and $\bar{v}<0$, and θ if $\bar{u}<0$ and $\bar{v}>0$ where $\theta=(180/\pi)*arctan(\bar{u}/\bar{v})$. This is method 1 described at: http://www.ndbc.noaa.gov/wndav.shtml.

[12] Available at www.wunderground.com.

[13] A 2007 exchange rate of 7.6 is used throughout the paper.

[14] Over 95% of firm observations in the data are for single-plant firms so that location can be uniquely identified. Results are robust to excluding multi-plant firms which report a single address.

pollution is uncorrelated with the product-mix decisions of a firm in the focal city thereby removing any bias in the instrumented results.

As explained below, we impose a maximum distance of 1,800 kilometers in estimating the spillover decay function and 300 kilometers in the causal estimates of productivity effects. After merging the productivity, API, and weather data for the spillover estimates, the data include 60 focal cities that represent 26% of China's population. The total annual output of these cities is CNY 2.02 trillion (11.7% of China's annual GDP and 29% of China's manufacturing sector).[15] For the casual estimates, the data includes 88,716 firms in 47 focal cities with total annual output of CNY 1.35 trillion (7.8% of China's annual GDP and 20% of China's manufacturing sector). Although the sample of cities is not comprehensive these are major cities representing a significant fraction of manufacturing output and population.

3. Estimation

3.1 Overview of estimation approach

This section describes our two-step procedure. In the first step we estimate the pollution decay function. In the second step we employ M2SLS to estimate the causal effect of focal-city pollution on focal-city productivity using annual data, instrumenting daily focal-city pollution with daily nearby-city pollution conditional on wind blowing toward the focal city. This step estimates the local average treatment effect of pollution on productivity. We then multiply the estimates for the spillover decay function obtained in the first step by the instrumental variable coefficient from the second step to yield the spillover effect of nearby-city pollution on focal-city productivity according to Equation (1). We bootstrap to compute standard errors that account for estimation error across both steps. The spillover decay function is estimated at the city level because pollution is measured at that level while the causal effects of pollution on productivity are estimated at the firm level because productivity is measured and occurs at the firm level. The next subsection describes the first step of the approach (estimating the pollution decay function) and the following subsection the second step (estimating the causal effect).

3.2 Step one: estimating the pollution decay function

The pollution decay function isolates the physical transport of PM_{10} between nearby and focal cities. If wind direction is orthogonal to omitted factors that jointly affect both nearby- and focal-city pollution, relating the two during periods when wind blows toward the focal city identifies these spillovers. We offer evidence that wind

[15] China's average annual real GDP over the seven-year sample period is CNY 17.27 trillion. The manufacturing sector accounts for roughly 40% of China's GDP.

direction is orthogonal to these omitted factors when we present the results. It is also necessary to isolate time periods in which the wind blows toward the focal city versus away. In the sample, wind direction changes by more than 90 degrees in absolute value (and therefore blows in the opposite direction) from day-to-day on more than 25% of days (Appendix C shows the full distribution of the change in wind direction across days). Averaging over a longer time period risks mingling periods in which the wind blows toward and away from the focal city. Thus, it is imperative to use daily data to isolate imported from local pollution.

We follow the concentric rings approach from the urban economics literature to estimate the pollution decay function.[16] This approach estimates the spillover between a location and each of several concentric rings radiating outward from that location. We use a piecewise linear regression to implement this, allowing the slope and intercept to differ for each of the concentric rings. We define rings at every 50 kilometers indexed by $b = 1,2,3, \ldots, B$ and identify all the nearby cities within each ring (if at least one exists) for each focal city. That is, all nearby cities within 0 to 50, 50 to 100, . . . , $(B - 1)$*50 to B*50 kilometers. We expand B far enough to ensure the decay function has plateaued or hit zero ($B = 36$ or 1,800 kilometers).

Having identified these focal-nearby city pairs, we then estimate the impact of nearby city n's PM_{10} on focal city f's PM_{10} level on day d of month m in year t by estimating the following equation conditional on the wind blowing from the nearby to the focal city:

$$P_{td}^{f} = \mathrm{I}_b\left[\lambda_{1b} + \lambda_{2b} abs\left[cos\left(\theta_{td}^{fn}\right)\right] P_{td}^{n}\right] + \lambda_3 W_{td}^{f} + \omega_f + \kappa_{rtm} + \varepsilon_{td}^{fn},$$

$$\forall f, n \in \mathcal{F}, n \neq f, \forall b = 1, \cdots B, \quad (2)$$

where $\mathcal{F}$ is the set of all cities in the data, P_{td}^{f} and P_{td}^{n} are the pollution levels of focal city f and nearby city n on day d of year t, and W_{td}^{f} are daily weather controls that affect pollution in the focal city. The indictor variable I_b is set to one for distance band b if nearby city n is within distance band b. λ_{1b} allows the intercept to vary for each distance band. λ_{2b} are the coefficients of interest and capture the average physical transport of nearby-city pollution to the focal city within each band. An observation in this regression is a focal-nearby city pair on a particular day. We form all possible pairings of focal and nearby city cities within 1,800 kilometers. Since each focal city may have more than one nearby city across or even within bands this is a stacked regression with potentially multiple observations per focal city.

[16] Examples are Rosenthal and Strange (2003); Fu (2007); Henderson (2007); Arzaghi and Henderson (2008); Rosenthal and Strange (2008).

We follow Schlenker and Walker (2016) in weighting nearby-city pollution by the absolute value of the cosine of the angle.[17] This angle $\left(\theta_{td}^{fn}\right)$ is the difference between the wind direction and the direction of the ray from the nearby to the focal city on day d of year t. For example, in Figure 1 where the focal city lies at an angle of 21° from the nearby city, if the wind is blowing at $-19°$ then $\theta_{td}^{fn} = -40°$ or if the wind is blowing at 43° then $\theta_{td}^{fn} = 22°$. We include a day in estimation as long as the wind blows within a 90° arc on either side of the ray connecting the nearby to the focal city. This is illustrated in the shaded area of Figure 1 for the example in which the focal city lies at an angle of 21° from the nearby city. In this example a day is included as long as $-69° < \theta_{td}^{fn} < 111°$. The pollution decay function is therefore identified from variation along two dimensions: distance between focal and nearby city and wind direction angle.

[Insert Figure 1 here]

W_{td}^{f} includes daily averages of relative humidity and wind speed, daily total precipitation, and temperature bins as described below. We include focal-city fixed effects (ω_f) to control for any time-persistent unobserved factors affecting the pollution drift to a focal city. Region-by-year-by-month fixed effects (κ_{rtm}) control for seasonal factors that affect pollution drift in a region such as wind patterns. We follow Zhang *et al.* (2018) in grouping the provinces into each of seven regions as described in Appendix D. The error term (ε_{td}^{fn}) captures any unobserved factors affecting drift between the focal-nearby city pair on day d of year t. We cluster standard errors at the focal-city level to allow for serial correlation across time within a focal city. This also allows for heteroscedasticity introduced by focal cities having different numbers of nearby cities.

3.3 Step two: estimating causal effect of pollution on productivity

In the second step we estimate the causal effect of focal-city pollution on focal-city productivity. In the short run, high air pollution concentrations can lead to decreased lung function, irregular heartbeat, increased respiratory problems, nonfatal heart attacks, and angina.[18] Long-run cumulative exposure may lead to cardiopulmonary diseases, respiratory infections, lung cancer (EPA, 2004), and asthma (Neidell, 2004) that can surface in the short run. All of these health

---

[17] We weight by the angle because more nearby-city pollution is imported the more directly wind blows toward the focal city. Using data for $-90° \leq \theta \leq 90°$ for the nearest nearby-city within 300 kilometers, the correlation between $cos(\theta)$ and residuals from regressing focal-city pollution on nearby-city pollution and focal-city weather is 0.046 significant at better than the 0.01% level. This means that if nearby-city pollution is increased by one μg/m^3 while θ is moved from 90° (perpendicular to the focal city) to 0° (directly toward the focal city), imported pollution increases by 0.046 μg/m^3 (21% of the total 0.216 μg/m^3 spillover at 300 kilometers shown in Appendix G).

[18] See the EPA website: https://www.epa.gov/pm-pollution.

conditions may decrease physical stamina and lead to missed work days. Workers may also be absent from work to care for the young and elderly affected by pollution (Chay and Greenstone, 2003; Hanna and Oliva, 2015; Deryugina *et al*., 2019; Aragón *et al*., 2017). Increased mortality (Chen *et al*., 2013; Ebenstein *et al*., 2017) can lead to experienced workers being replaced by less experienced ones. Air pollution can also have psychological effects including lowering cognitive ability, altering emotions, and increasing anxiety (Levinson, 2012; Lavy *et al*., 2014; Pun *et al*., 2016; Chen *et al*., 2018) which would affect both physical and mental performance. While the estimates are unable to distinguish between these various channels they capture the effect of all of them.

3.3.1 Step two: identification

We focus here on identification issues related to productivity but the identification arguments apply to endogeneity issues that arise from outcomes more broadly. OLS estimates are subject to simultaneity and omitted variable biases. Even without any effect of pollution on productivity, cities with more output will produce more pollution. If pollution does lower productivity, the lower productivity will result in less pollution. Firms may also respond to the lowered labor productivity by substituting from labor to alternative inputs.

Omitted-variable biases due to local, time-varying conditions are also possible (firm fixed effects absorb any time-invariant effects). For example, high-productivity firms may implement advanced, lower-polluting technologies over time while low-productivity firms do not. Spatial sorting could introduce spurious correlations. Firms may choose to enter in or relocate to cities with less severe pollution because it will raise their productivity or in cities with more severe pollution because they have lax environmental regulations and impose fewer costs (Becker and Henderson, 2000; Greenstone, 2002; Brunnermeier and Levinson, 2004). Governments may force firms to relocate and pollution inflow from other cities may affect these decisions (for example, moving firms away from areas that are typically upstream of densely-populated areas). Firm exit may be endogenous due to the reduced productivity that pollution brings. Workers may also systematically sort across cities. High-skilled workers generally have a higher willingness-to-pay for clean air which would lead to low-skilled workers being located disproportionately in dirtier cities (Chen *et al*., 2017; Lin, 2017). The inclusion of firm fixed effects means that only migrations of firms or workers during the sample period will bias the results.

We address these issues using nearby-city pollution that drifts to the focal city as an instrumental variable to identify the causal effect of local pollution on local productivity. To ensure exogeneity, we condition on the wind blowing from the nearby to the focal city.[19] Exogeneity also requires that wind direction timing is random with respect to nearby-city air pollution, conditional on controls, which we confirm below.

The inclusion restriction requires that the nearby city is close enough that significant amounts of pollution can drift from it to the focal city. To ensure this, we include only focal cities that have a nearby city sufficiently close. We consider maximum distance cutoffs ranging from 150 to 300 kilometers (our pollution decay function estimates confirm significant transport at these distances). There is a tradeoff in increasing the distance: it increases the available data but weakens the instrument's power. To also increase the instrument's power we include only the nearest nearby city for each focal city. As a result, even with a maximum distance of 300 kilometers the average distance between focal and nearby cities is only 106.5 kilometers.

The exogeneity condition requires that unobserved determinants of focal-city productivity are uncorrelated with the nearby city's pollution. This requires high-frequency data for two reasons. First, periods in which the wind imports pollution from outside must be isolated from those when it does not. To ensure this, in the instrumenting equation we condition on the wind blowing from the nearby to the focal city on a particular day. We offer evidence when we present the results that daily data succeeds in isolating periods when wind blows toward the focal city. Conveniently, this high-frequency instrument is already available as it is required to estimate the pollution decay function.

Second, high-frequency data is required to ensure that common shocks do not affect both focal- and nearby-city output. Positive regional shocks to productivity could raise both cities' output thereby increasing nearby-city pollution as well. Alternatively, if focal- and nearby-city production are substitutes in output markets then output growth in a focal city will reduce nearby-city output and pollution. While common regional shocks are likely to induce correlated actions across cities over a long time period, they are unlikely to do so over a short time frame due to lags in shock propagation and delays in responses to those shocks. With the use of daily data lower-frequency controls can be included and violating the exogeneity condition would require that shocks affect focal- and nearby-city productivity on a *daily* basis.

This addresses each of the potential endogeneity biases. Nearby-city pollution is uncorrelated with focal-city output in the absence of common regional shocks that are propagated and responded to on a daily basis. Trends in pollution and

[19] When the wind blows toward the nearby city its pollution is not exogenous because greater focal-city output increases the nearby city's air pollution. We address the possibility of serial correlation in output when we present our results.

productivity would need to be correlated across the focal and nearby city on a daily basis to bias the estimates. Substitution away from labor and toward other inputs in response to imported pollution would need to occur on a daily basis. Similarly, firm entry, exit, or relocations and worker migrations in response to imported pollution would need to occur on a daily basis.[20]

This instrumenting strategy can be implemented using either M2SLS with daily data in the first stage or Wald 2SLS with annual averages in the first stage (in either case conditioning on wind direction). Appendix E shows formally that either approach produces unbiased estimates in the presence of a common shock to focal- and nearby-city output as long as it is of lower than daily frequency. However, there are two important differences between the two approaches as shown formally in Appendix E. M2SLS produces unbiased estimates in the first stage because intra-year common regional shocks to pollution (as opposed to output) can be controlled for using fixed effects while Wald 2SLS may produce biased estimates.[21] Second, M2SLS produces more efficient second-stage estimates as we demonstrate below.

In the results we assess the effects of aggregating the instrument to lower and lower frequencies. Consistent with the theoretical predictions in Appendix E, the first-stage coefficient becomes increasingly biased at lower frequencies due to common shocks to focal- and nearby-city pollution and the second-stage coefficient become less precise.

3.3.2 Step two: implementation

The outcome that we wish to estimate (productivity) is measured annually while the pollution instrument is daily. A standard way of proceeding is to estimate Wald 2SLS using annualized values (conditional on wind direction in the first stage). We show below that these estimates are inefficient. Instead, we employ M2SLS which provides estimates that are consistent and asymptotically normal (Dhrymes and Lleras-Muney, 2006) provided that the groupings are independent of the structural error as they are when the grouping is a primitive (in our case grouping daily

[20] For example, suppose a factory moved from a focal city to a nearby city mid-year. For the first half-year, the local pollution it produces would lower productivity but this would not affect our estimates since this pollution is uncorrelated with nearby-city pollution conditional on wind direction. In the second half-year, this would increase the pollution that drifts to the focal city from the nearby city. It would also decrease productivity in the focal city in the last half-year due to spillovers. Our estimates would capture this since we condition on wind direction.

[21] For M2SLS, these are controlled for by region-by-year-by-month fixed effects in the first stage. For Wald 2SLS the first stage is biased by these effects; however, the second stage remains unbiased because the predicted values from the first stage are uncorrelated with the common shocks to output that may be present in the second stage.

observations into years).[22] Theoretically, M2SLS can be more or less efficient but we show in our setting that it is more efficient.

The first-stage equation predicts air pollution for firm i located in focal city f of region r on day d in month m of year t conditional on the wind blowing from the nearby to the focal city. While the spillover equation in step one uses city data, this equation uses firm data to be consistent with the firm data used in the second stage:

$$P_{itd}^{f} = \gamma_1 abs\left[cos\left(\theta_{itd}^{fN^*}\right)\right] P_{itd}^{N^*} + \gamma_2 W_{itd}^{f} + \alpha_i + \kappa_{rtm} + \epsilon_{itd}^{f}, \quad (3)$$

where P_{itd}^{f} is the pollution in firm i's focal city f on day d of year t, $\theta_{itd}^{fN^*}$ is the wind direction relative to the ray from the nearest nearby city to firm i's focal city on day d of year t, and $P_{itd}^{N^*}$ is the pollution level on that same day in focal city f's nearest nearby city $N^* \in \mathcal{F}$ within a maximum radius distance. If no nearby city is available for a focal city it is dropped from the estimation. Every nearby city is also a focal city although it might be paired with a different nearby city that is closer. We test the sensitivity of the results to maximum distance cutoffs ranging from 150 to 300 kilometers.[23] W_{itd} is a vector of daily weather variables faced by firm i on day d of year t. We include linear and quadratic functions of daily relative humidity, wind speed, and cumulative precipitation. We allow for a flexible, nonlinear function of temperature following Deschênes and Greenstone (2011) and Zhang *et al.* (2018) since it has been found to affect productivity (Zhang *et al.*, 2018). We construct indicator variables for the daily average temperature below 0°, 5° intervals from 0 to 30°, and above 30° Celsius.

In defining whether the wind blows toward the focal city, we impose more stringent criteria than in the pollution decay function estimation to ensure a sufficient quantity of pollution is imported from the nearby city. This is necessary for the instrument to be powerful.[24] For the baseline estimates, we include a day if the wind passes within a 45° arc on either side of the ray connecting the two cities. We refer to this as the "middle" funnel. Figure 2 illustrates this for the example in which the focal city lies at an angle of 21° from the nearby city. In this case a day is included as long as $-24° < \theta_{td}^{fn} < 66°$ (the shaded region of the figure). We check the robustness of the results to arcs of $\pm 40°$ ("narrow" funnel) and $\pm 50°$ ("broad" funnel). As in the pollution decay function estimation, the nearby-city's pollution is weighted by the absolute value of the cosine of the angle.

---

[22] Lleras-Muney (2005) applies M2SLS to estimate the causal impact of education on health, Massa and Žaldokas (2014) to estimate international demand for US bonds, and Jordan (2016) to estimate local environmental preferences on mine closures.

[23] Distances below 150 kilometers yielded insufficient data and distances above 300 kilometers yielded a weak instrument as we demonstrate below.

[24] Footnote 17 provides evidence that nearby-city pollution is a stronger instrument when the wind blows more directly in the direction of the focal city.

[Insert Figure 2 here]

Firm fixed effects (α_i) capture time-persistent unobservables that affect firm i's pollution exposure. Since no firms switch focal cities or industries over the sample period, these also absorb city-specific and industry-specific time-invariant factors that affect local pollution. Region-by-year-by-month fixed effects (κ_{rtm}) control for any year-month specific unobservables affecting the pollution in a region. We cluster standard errors at the focal-city level to allow for spatial correlation for all firms within each focal city and serial correlation across days within a focal city over time.

This equation differs from the pollution decay function (Equation (2)) in two ways. First, in order to ensure the power of the instrument, Equation (3) restricts estimation to shorter distances (a maximum of 300 kilometers), it utilizes only the nearest nearby city, and includes only days when the wind direction is within a funnel rather than within a half-circle. This maximizes the potential for the nearby city's pollution to drift to and affect the focal city. The objective of Equation (2) is to estimate spatial decay and it therefore utilizes all of the nearby cities to a focal city, utilizes all days of wind direction within a half-circle, and extends the measurement of these spillovers to a much greater distance. Second, Equation (2) also allows for a much more flexible functional form for estimating the spillover decay function than the linear restriction that 2SLS imposes on Equation (3).

Using the results from estimating Equation (3), we compute predicted values $\hat{P}_{itd}^{f}$ for each day included in the estimation (wind blowing toward the focal city) and average them over days within each firm-year to obtain instrumented pollution for the second-stage: $\bar{\hat{P}}_{it}^{f}$. The second-stage equation is:

$$\ln\left(Y_{it}^{f}\right) = \beta_1 \bar{\hat{P}}_{it}^{f} + \gamma_2 \bar{W}_{it}^{f} + \alpha_i + \delta_{rt} + \eta_{it}^{f}, \quad (4)$$

where Y_{it}^{f} is value added per employee for firm i in the focal city f in year t and $\bar{W}_{it}^{f}$ contains the weather controls from the first stage averaged over all days within each firm-year (i.e., averages of the linear and quadratic functions of non-temperature variables and temperature bins containing the fraction of days in which the average temperature is below 0°, in 5° intervals from 0 to 30°, and above 30° Celsius).[25]

Firm fixed effects α_i capture time-persistent firm attributes that affect labor productivity. Region-by-year fixed effects (δ_{rt}) capture time-varying, regional shocks

[25] To ensure the exclusion restriction is met, the first-stage equation must include the non-averaged values of all the exogenous variables from the second stage. The weather controls in the second stage ($\bar{W}_{it}^{f}$) are yearly averages of the linear and quadratic terms of all non-temperature variables in the first stage. For the temperature variable, the bins in the second stage are annual averages of the daily indicator variables included in the first stage. The firm fixed effects remain the same as in the first stage. Finally, the region-by-year fixed effects included in the second stage are averages of the region-by-year-by-month fixed effects in the first stage.

to firm output. The error term (η_{it}) includes time-varying, firm-level shocks to productivity. We cluster standard errors at the focal-city level to allow for serial correlation within each firm over time and spatial correlation within each city. We adjust for the error introduced in the first-stage estimation using a block bootstrap as in Schlenker and Walker (2016) with 100 iterations.

4. Results

Before we show our results we establish that a straightforward reduced-form regression of focal-city productivity on nearby-city pollution produces inefficient estimates. To do so, we aggregate nearby-city pollution to the annual level conditional on wind direction, weighted by the cosine of the wind-direction angle, and include control variables corresponding to those used in M2SLS.[26] Appendix F graphs the results converting them to their monetary impact. It shows the effects of a one $\mu g/m^3$ annual decrease in nearby-city PM_{10} within a distance band (holding all others constant) on the average firm's annual productivity along with the 95% confidence interval in red, dashed lines. All the effects except for the 0-50 kilometer distance band are close to zero and almost all are insignificant. Given this lack of precision, we turn to our approach.

We report the first-step estimates (pollution decay function) followed by the second-step estimates (causal effects of focal-city air pollution on focal-city productivity) and then combine the results from these two steps to calculate the spillover effects of nearby-city pollution on focal-city productivity. After this, we demonstrate the advantages of the M2SLS procedure. In particular, we show that estimating causal effects using Wald 2SLS with annual data produces insignificant second-stage results and biased first-stage results. We offer supporting evidence that this is due to aggregating the high-frequency data to a lower frequency.

4.1 Pollution decay function

To estimate the pollution decay function we include all focal cities with at least one nearby city within 1,800 kilometers. This distance was chosen because it was far enough that the spillover effects were indistinguishable from zero.[27] We use all cities that have daily API and weather data available from 2001 to 2007. This yields 60 unique cities in a panel which is unbalanced because API data was not reported for some cities in the early years. There are some days with missing API or wind data but these are limited (all cities have at least 335 days of data in each year) and we believe are due to random non-reporting.

[26] An alternative reduced-form approach would be to regress annual productivity on daily nearby-city pollution but this would involve over two billion observations in our specification.

[27] Re-estimating with a maximum radius of 1,200 kilometers (just above the point at which the effects hit zero) yields almost identical coefficients and standard errors.

Table 1 shows the summary statistics for the pollution spillover data. There are 2,586 focal-nearby-city pairs (about 43 nearby cities for each focal city). If city B is a focal city for A then A is also a focal city for B. The focal cities' PM_{10} levels average 97.5 and exhibit significant variation. Wind blows toward the focal city on 52.1% of the days and PM_{10} is the dominant pollutant on 92% of the days for the focal cities. The mean distance between cities (1,004 kilometers) is about one-half the maximum allowed distance.

[Insert Table 1 here]

The solid, black line in Appendix G shows the λ_{2b} coefficients from estimating Equation (2) along with the 95% confidence interval in red, dashed lines. These are the effects of a one μg/m[3] increase in PM_{10} in nearby cities conditional on wind blowing directly toward the focal city $\left(\theta_{td}^{fn} = 0\right)$. The effect in each distance band is conditional on holding PM_{10} in other bands constant. Roughly 45% of pollution drifts from nearby cities that are within 50 kilometers and more than 18% at 400 kilometers.

The solid, black line in Figure 3 plots the effect of a one μg/m[3] *annual* increase in nearby-city PM_{10} along with the 95% confidence interval in red, dashed lines (for clarity we plot only to a distance of 1,200 kilometers). This adjusts the coefficients using the empirical distribution of θ_{td}^{fn}. That is, for the fact that the wind blows toward the average focal city on only 52.1% of days in a year and does not always blow directly towards the focal city. Again, this is the effect of increasing PM_{10} in the distance band conditional on holding pollution constant in all other bands.[28] The spillover effect within 50 kilometers is 0.355. That is, a one μg/m[3] annual increase in PM_{10} in all nearby cities within 50 kilometers, but not in any other distance band, increases annual focal city pollution by 0.355 μg/m[3]. Similarly, a one μg/m[3] annual increase in PM_{10} in all nearby cities within 50 to 100 kilometers, but not in any other band, increases annual focal city pollution by 0.185 μg/m[3]. A similar description applies to all further distance bands. These effects are for the average focal city with average weather. Spillovers drop somewhat quickly and smoothly from 0.355 at 50 kilometers to 0.062 at 600 kilometers after which they fall more slowly to zero at about 1,000 kilometers.

[Insert Figure 3 here]

[28] It would be useful to compare the local effect to spillovers from raising pollution in all nearby cities simultaneously. However, to do so requires including interaction effects between each distance band and all closer distance bands to estimate these "pass-through" effects. The number of independent variables required makes this infeasible with more than a few distance bands.

4.2 Randomness of daily wind data

Before estimating the causal effect of pollution on productivity, we check the randomness of wind direction with respect to pollution. To ensure that the instrument is exogenous we must exclude days in which the wind does not blow from the nearby to the focal city. If wind direction is not randomly distributed with respect to the distribution of nearby-city air quality, conditional on control variables, this may bias the coefficients.[29] Appendix H compares cumulative distribution functions (cdfs) of nearby-city air pollution conditional on the control variables used in the first stage of the M2SLS procedure for all days versus included days using the 150-, 200-, 250-, and 300-kilometer distance cutoffs in choosing nearby cities. The cdfs are very similar for all cutoffs.[30]

4.3 Effect of local air pollution on local labor productivity

In this subsection we estimate the causal effect of focal-city pollution on focal-city labor productivity using nearby-city pollution as an instrument. In choosing which nearby cities to include, we check robustness to maximum distances from the focal city of 150, 200, 250, and 300 kilometers. Below 150 kilometers there were insufficient data to identify effects and we show that beyond a distance of 300 kilometers the instrument is no longer powerful enough.

Table 2 shows summary statistics for the main variables for the 150- and 300-kilometer radiuses. The top panel summarizes the first-stage data which are at the firm-day level. The summary statistics are fairly similar across the two distance cutoffs. The PM_{10} levels are high enough to potentially affect productivity. The annual mean is 112 $\mu g/m^3$ compared to a World Health Organization (WHO) recommended guideline of 20 $\mu g/m^3$ annual average and many days exceed the WHO guideline of 25 $\mu g/m^3$ for a 24-hour average (World Health Organization, 2006). As the cutoff increases from 150 to 300 kilometers, the number of focal cities increases from 30 to 47. The average distance between nearby and focal cities does not increase much because we use the nearest nearby city for each focal city. The bottom panel summarizes the second-stage data which are at the firm-year level. The

[29] This highlights the importance of the control variables. For example, in northern regions of China air quality is worst in the winter. If wind directions are systematically different in winter than in other seasons this will introduce bias in the absence of control variables. In this example, the region-by-year-by-month fixed effects capture this region-specific seasonality.

[30] A two-sample Kolmogorov-Smirnov test rejects the null hypothesis of the equality of distributions for three of the radius cutoffs; however, the magnitude of the differences is very small. For the 200-kilometer radius the difference is significant at the 1.8% level but the maximum difference is only 0.016. For the 250-kilometer radius the difference is significant at the 3.0% level but the maximum difference is only 0.014 and for the 300-kilometer radius the difference is significant at the 3.9% level but the maximum difference is only 0.013. For a 150-kilometer radius the difference is not quite significant (10.8%) but the maximum difference is only 0.014. This is an example of Simpson's Paradox in which a large amount of data (for the 200-kilometer radius there are 55,088 observations) results in statistical significance for even small differences.

data exhibit significant variation in value-added per employee. Appendix I shows summary statistics for the 200- and 250-kilometer radiuses.

[Insert Table 2 here]

Panel A of Table 3 shows OLS results that do not address the endogeneity of air pollution. The firm-year data included here correspond to those included in the second stage of M2SLS estimation described below. For all four distance cutoffs, the coefficients on PM_{10} are insignificantly different from zero and for all but the 150-kilometer the point estimates themselves are close to zero.

[Insert Table 3 here]

We now turn to M2SLS estimates. Panel B shows the results of estimating the first-stage equation (Equation (3)) using PM_{10} of the focal city's nearest nearby city as an instrument conditional on wind blowing toward the focal city within the middle funnel. This estimation is at the firm-day level and the wind is within the middle funnel on about one-fourth of the days. The results reveal a strong instrument. A one μg/m^3 increase in a nearby city's PM_{10} increases the focal city's PM_{10} by between 0.70 and 0.72 with a high level of significance.[31] This is not too far from the theoretical upper bound of 1.0 because it uses only the nearest nearby city and pertains to days when the wind is blowing directly toward the focal city$\left(\theta_{td}^{fN*} = 0\right)$. The Kleibergen-Paap Wald rk (KP) *F*-statistic (Kleibergen and Paap, 2006) for weak identification significantly exceeds the Stock-Yogo critical value of 16.38 for all four cutoffs.[32]

Panel C shows the second-stage estimates of Equation (4) at the firm-year level using the average values of the predicted pollution from the first stage as an instrument and controlling for weather and region-by-year fixed effects. The estimated coefficients of PM_{10} are negative and significant for all but the 150-kilomoter cutoff. The estimates become more significant as the cutoff increases consistent with more data used in estimation. The coefficients imply that a one μg/m^3 annual increase in PM_{10} decreases productivity by 0.26% to 0.34%. Evaluated at the mean focal-city PM_{10} in each subsample, these estimates imply elasticities of labor productivity with respect to air pollution of -0.29 to -0.35.[33] These results are consistent with the instrument attenuating an upward endogeneity bias. The results also imply that improving air quality generates substantial productivity benefits.

[31] These coefficients exceed the estimates even at 50 kilometers in the spillover decay function (0.45 from Appendix G) because here we estimate using a funnel that is twice as narrow.

[32] Stock and Yogo (2005) critical values apply when model errors are independent and identically distributed. No critical values are available for the case when the model allows for standard errors that are robust to heteroskedasticity and clustering.

[33] Mean annual PM_{10} (unconditional on wind direction) in the second-stage data is 104.1 for 150-, 111.3 for 200-, 103.1 for 250-, and 104.1 for 300-kilometer radius.

The difference in second-stage coefficients across the different radiuses is primarily due to the influence of firms located in the Northeast region of China. As Columns 1 and 2 of Appendix J show, the point estimate for pollution's effect on productivity is much greater in the Northeast region than elsewhere.[34] Throughout the remainder of the paper we use the 300-kilometer estimate as our preferred since it includes the most comprehensive data and is not overly influenced by data from the Northeast region. Using the 300-kilometer cutoff, a 1% reduction in PM_{10} increases per-firm productivity for the average firm by CNY 47,700 (USD 6,276) annually. The implied elasticity (-0.31) is lower than the nationwide estimate of -0.44 for $PM_{2.5}$ in Fu *et al.* (2021).

Appendix J shows two other heterogeneity tests. Pollution has greater effects on the productivity of "dirty" firms (Column 3) than on "clean" firms (Column 4) and greater effects on the productivity of low- (Column 5) than on high-technology firms (Column 6).[35] This is consistent with clean and high-technology firms possibly investing more over time in defensive actions to avoid air pollution's effects.

Since our estimates capture pollution's effect on both per-hour productivity and working hours, it is useful to disentangle the two for comparisons to previous estimates of per-hour productivity effects. We borrow estimates from Aragón *et al.* (2017) which finds an elasticity of working hours with respect to $PM_{2.5}$ of -0.21 in Lima, Peru. Assuming $PM_{2.5}$'s effect on working hours is the same in China, our estimated elasticity of per-hour productivity with respect to pollution is -0.10.[36] This somewhat exceeds many previous estimates: elasticity of -0.052 with respect to $PM_{2.5}$ for garment factory workers in India (Adhvaryu *et al.*, 2019), elasticity of -0.062 for $PM_{2.5}$ for indoor pear packers in California (Chang *et al.*, 2016), and elasticity of -0.023 with respect to the API for services workers (Chang *et al.*, 2019). He *et al.* (2019) find much larger elasticities (ranging from -0.035 to -0.30) with respect to $PM_{2.5}$ for textile workers in two firms in two Chinese provinces if effects are accumulated over 25 to 30 days. Our estimates may differ from these previous estimates for two reasons. First, previous estimates apply only to particular worker types or small sets of firms. Second, previous studies measure daily or monthly effects while we capture cumulative annual effects.

[34] Only four percent of the firms used in estimating with the 150-kilometer radius are in the Northeast region and none of the firms added in moving to the 200-kilometer radius. However, 91% of the firms added in moving from the 200- to the 250-kilometer radius are located in the Northeast region and 33% in moving to the 300-kilometer radius. This explains why the coefficient for the 250-kilometer cutoff is the largest (in absolute value) followed by the 300- and then 200-kilometer cutoffs. The Northeast-region coefficient is not significant but there is little data. It is unclear why the Northeast region experiences greater effects. We investigated industry mix, pollution levels, firm density, timing of data, firm size, capital and labor intensity, and ownership type and none explained the difference.

[35] We define "dirty" and "clean" based on the 3-digit SIC codes in Mani and Wheeler (1998). We define high- and low-technology industries based on OECD (2011).

[36] This uses the fact that the elasticity of productivity equals the elasticity of productivity per hour (a) plus the elasticity of annual hours worked (H): $\frac{dln[productivity]}{\mathrm{dln[pollution]}} = \frac{dln[aH]}{\mathrm{dln[pollution]}} = \left[\frac{dln[a]}{\mathrm{dln[pollution]}} * H + a * \frac{dln[H]}{\mathrm{dln[pollution]}}\right]$.

4.4 Counterfactuals and robustness

Column 2 of Appendix K reports results of a counterfactual test of the instrument. It uses M2SLS with the middle funnel and a 300-kilometer radius but conditions on wind blowing away from the focal city in instrumenting for focal-city PM_{10}. The first-stage results (shown in Panel A) are nearly identical to those using the baseline model (reproduced in Column 1). This is not surprising: the two cities should have roughly the same effect on each other. The second-stage results (shown in Panel B) are very different than the baseline results. The coefficient is much lower in magnitude and insignificant, consistent with the instrument not addressing endogeneity bias. The estimates are similar to the OLS results in Panel A of Table 3.

Appendix K contains other robustness checks using the 300-kilometer cutoff. Column 3 uses a narrow funnel (an 80° arc). The point estimate is slightly smaller and is significant only at the 16% level due to the loss of data in the first stage. Employing a broad funnel (a 100° arc) with more data in Column 4 produces a somewhat more significant and larger effect than the baseline estimate. Dropping days with API below 50, for which the major pollutant is unknown, lowers the coefficient somewhat (Column 5). This is presumably due to years with more low-pollution days also having more high-productivity days.

Column 6 shows the importance of including weather controls. Without them, the second-stage coefficient is lower and no longer significant. This could be due either to correlations between these weather controls and wind direction or because weather directly affects productivity. No single weather control alone appears to be determinative.[37] Including the potentially manipulated range of API (Column 7) produces almost identical results to the baseline. Including year-by-month rather than region-by-year-by-month fixed effects in the first stage (Column 8) yields similar results to the baseline but even more significant while including region-by-year fixed effects in the first stage results in somewhat different and less significant estimates (Column 9).[38] Therefore, the estimates are sensitive to controlling for overall seasonality more so than region-specific effects.

Since daily nearby-city pollution is used as an instrument, the presence of autocorrelation in daily output could introduce bias in the second stage.[39] With autocorrelation, today's focal-city output is affected by today's nearby-city pollution directly rather than only through focal-city pollution. This is true regardless of yesterday's wind direction. Suppose that wind originates in the nearby city

[37] The second-stage coefficients from including each individual weather control alone are: temperature (-0.0024), humidity (-0.0023), wind speed (-0.0025), and precipitation (-0.0019).
[38] We experimented with using province-by-year-by-month fixed effects but the model was too saturated. There is an average of only 1.5 cities per province in the data.
[39] We thank the editor for pointing this issue out.

yesterday. Yesterday's output in the nearby city will be correlated with its output today due to autocorrelation. At the same time, the nearby city's pollution will drift to the focal city on the previous day and affect the focal city's previous-day output. This in turn will affect its current-day output due to autocorrelation. This means that today's nearby-city output (and therefore pollution) is correlated with today's focal-city output. An analogous argument follows if wind originates in the focal city on the previous day.[40]

To address this we would ideally include lagged focal-city output as a control variable in both stages of M2SLS since it is pre-determined with respect to current output. Conditioning on this would eliminate the correlation between today's focal- and nearby-city output variables and restore nearby-city pollution as a valid instrument. Controlling for the annual average of lagged output in the second stage is actually unnecessary because autocorrelation is negligible in annual measures.[41] Since daily output data are unavailable, we instead include lagged focal-city pollution as a control variable in M2SLS as a proxy for it. Lagged focal-city pollution is correlated with lagged focal-city output via the pollution production function and is pre-determined with respect to current output. As with using lagged focal-city output as a control, it is unnecessary to include it as a control variable in the second stage because autocorrelation in annual pollution measures is negligible. Appendix L provides a formal description of the autocorrelation and the validity of including lagged focal-city pollution as a control.

Table 4 compares the results including lagged focal-city pollution in the first stage of M2SLS (Column 2) to our original results (Column 1) using the 300-kilometer maximum radius and middle funnel. The first-stage results indicate that there is significant autocorrelation in focal-city output. However, the second-stage coefficients are of similar magnitude. Although contemporaneous focal-city pollution is significantly correlated with its lagged value, this does not greatly change the predicted values – the correlation between the second-stage fitted values used in Columns 1 and 2 is 0.968 with a significance level better than 0.01%. This suggests that the autocorrelation does not introduce significant endogeneity. The coefficient on a twice-lagged value of focal-city pollution is close to zero and insignificant (Column 3) indicating that including a single lag is sufficient.

[Insert Table 4 here]

[40] Specifically, yesterday's output in the focal city will be correlated with its output today due to autocorrelation. At the same time, the focal city's pollution will drift to the nearby city on the previous day and affect the nearby city's previous-day output. This in turn will affect its current-day output due to autocorrelation. This means that today's nearby-city output (and therefore pollution) is correlated with today's focal city output.

[41] The only influence of the daily autocorrelation would be on the first and last days of the year divided by the number of days in a year.

We use output per worker for our baseline results to be consistent with the environmental economics literature but the results are robust to using total factor productivity (TFP). In estimating TFP, we instrument for firms' endogenous choices of inputs using two different approaches: investment as an instrument (Olley and Pakes, 1996) and intermediate inputs as an instrument (Levinsohn and Petrin, 2003). Both yield a slightly lower elasticity -0.29 (Columns 4 and 5 of Table 4) than the baseline estimate of -0.31.

Appendix M provides supporting evidence for the choice of 300 kilometers as the maximum distance for the nearest nearby city to include as an instrument. Column 1 reproduces the baseline estimates. Column 2 estimates M2SLS using as an instrument pollution in the nearest nearby city for each focal city that is further than 300 but less than 350 kilometers away and using the middle funnel. Columns 3 through 5 expand the data by increasing the range of distances for the nearest nearby city. The average distance between the focal and nearby cities increases from 106.5 kilometers in the baseline estimates compared to more than 323.9 kilometers in the counterfactual estimates. The first-stage results in Panel A reflect the reduced power of the instrument compared to the baseline. The coefficient is about half that in the baseline estimates and the KP *F*-statistic is much lower. The second-stage coefficients (Panel B) are all insignificant consistent with a weak instrument.

4.4 Spillover effect of nearby-city pollution on focal-city labor productivity

As shown in Section 2.1, multiplying the first-step spillover decay function by the second-step causal effects yields the spillover effects of nearby-city pollution on focal-city productivity. To obtain appropriate standard errors clustered at the city level for these spillover effects we employ a block bootstrap with 100 iterations.[42] We estimate this using a 300-kilometer cutoff and middle funnel for the instrument.

Figure 4 summarizes the results converting them to the monetary impact for the average firm's annual productivity on an average weather day. The solid, black line shows the effect of a one μg/m^3 annual decrease in nearby-city PM_{10} in that distance band (holding pollution in all other bands constant) on focal-city productivity with 95% confidence intervals shown in dashed, red lines. This assumes a one μg/m^3 decrease in nearby-city PM_{10} for the entire year and adjusts for the empirical distribution of wind direction during the year. The gains are CNY 16,248 (USD 2,138) for nearby cities within 50 kilometers and decline fairly quickly and smoothly to CNY 2,847 (USD 375) for nearby cities at 550 to 600 kilometers. Beyond this, the spillovers decline slowly to approach zero at about 1,150 kilometers (for clarity we

[42] For each iteration we draw (with replacement) a block bootstrap by city. In the first step (spillover decay function) we use all days in all years for these cities. In the second step (causal effects) we use all firms and all days in all years for the sampled cities.

plot only to 1,200 kilometers). In comparison the effect of local sources of PM_{10} on productivity is CNY 45,809 (USD 6,028).

[Insert Figure 4 here]

While the spillover decay function estimates (Figure 3) alone tell us the relative tradeoff between local and extra-territorial effects, they do not tell us the absolute amounts at stake. This requires both steps of the procedure. For example, if PM_{10} increases by one μg/m^3 annually in both a focal city and a nearby city located 90 kilometers away, productivity falls by CNY 45,809 annually for the average firm due to local sources of pollution and another CNY 8,494 due to imported pollution. The latter is smaller because pollution dissipates as it drifts and because the wind blows directly toward the focal city only part of the time. These absolute costs can be used to determine the geographic scope of environmental regulation necessary to internalize externalities that are above a given cost.

These results can also be used to calculate Coasian prices. Consider Tianjin which is located 107 kilometers from Beijing and let θ_{td}^{BT} be the angle of the wind relative to the ray from Tianjin to Beijing. If each city were assigned rights to keep its city free of other cities' air pollution, Tianjin would have to compensate Beijing CNY 35*$abs[cos(\theta_{td}^{BT})]$ times the number of firms in Beijing on each day when $-90^{\circ} \leq \theta_{td}^{BT} \leq 90^{\circ}$, where 35 is the λ_{2b} coefficient from Equation (2) multiplied by the annual causal effect converted to a daily cost.[43] Similarly, on days when the wind blows toward Tianjin, Beijing would have to compensate Tianjin 35*$abs[cos(\theta_{td}^{TB})]$ times the number of firms in Tianjin for each μg/m^3 of PM_{10} that Beijing produces on a day when the wind blows between $-90^{\circ} \leq \theta_{td}^{TB} \leq 90^{\circ}$where θ_{td}^{TB} is the angle of the wind relative to the ray from Beijing to Tianjin. Some of the pollution blowing from Beijing to Tianjin may have originated in other cities before being passed on to Tianjin. These other cities would compensate Beijing using the same approach so that Beijing's net payment would correspond only to the pollution that it created.

4.5 Wald 2SLS estimates

An alternative to our procedure is to combine the first-step estimates of the pollution decay function using daily data with causal estimates based on Wald 2SLS, which requires aggregating the first-stage data to match the annual data used in the second stage. We aggregate the first-stage data by taking firm-year averages conditional on wind blowing toward the focal city (i.e., computing mean values of focal-city pollution and cosine-weighted nearby-city pollution using only days when the wind blows toward the focal city). We also include annual averages of weather controls,

[43] The λ_{2b} coefficient is 0.279 for nearby cities between 100 and 150 kilometers away. The annual causal effect is CNY 45,809 or CNY 125 daily. Multiplying these two numbers yields CNY 35.

firm and region-by-year fixed effects, and cluster standard errors by focal city to be consistent with the M2SLS estimates. Table 5 shows the results at the different distance cutoffs using the middle funnel.

The coefficients for the first-stage results (Panel A) are all significant but are opposite of the expected sign. This is because when there is intra-group variation, grouped estimation identifies different parameters than does ungrouped estimation (Angrist and Pischke, 2011: 314). Appendix N shows scatter plots that relate focal-city PM_{10} conditional on first-stage control variables and nearby-city PM_{10} for increasing levels of aggregation along with fitted regression lines. In all cases we condition on wind blowing toward the focal city. The daily plot shows a clear positive relationship. The primary effect of aggregation is a loss of precision in the relationship but the relationship also becomes negative with annual data. This is consistent with "clean" focal cities being located near "dirty" cities or vice versa and this effect overwhelming any upward bias due to monthly regional shocks to pollution (see Equations (E10) and (E11) of Appendix E for a formal exposition). Such juxtaposition would occur from some cities being manufacturing intensive and others not in the absence of systematic sorting. The results also suggest weak instruments with all of the KP *F*-statistics below the critical value of 16.38.[44]

[Insert Table 5 here]

As Appendix E shows, Wald 2SLS still produces unbiased second-stage estimates even with a biased first-stage coefficient as first-stage fitted values isolate focal-city pollution that is due to variation in nearby-city pollution. However, M2SLS may be more efficient for two reasons. The first reason is that finer control variables can be included in the first stage of M2SLS (in our case region-by-year-by month fixed effects) than in Wald 2SLS (in our case region-by-year fixed effects). There is no way to directly test whether this increases efficiency so we provide indirect evidence by comparing the goodness of fit of the M2SLS first stage using region-by-year-by-month fixed effects versus region-by-year fixed effects. Using the 300-kilometer radius and middle funnel, the null hypothesis that the restricted model using region-by-year fixed effects is a better fit is rejected at better than the 0.0000% level.[45] This suggests that region-by-year-by-month fixed effects are important for controlling seasonal effects that affect both nearby- and focal-city pollution and that M2SLS potentially provides more precise predicted values for the exogenous shock to pollution in the second stage than does Wald 2SLS. Consistent with this, the causal

[44] Consistent with a single instrument that is very significant, a standard Cragg-Donald (1993) test overwhelmingly rejects the null hypothesis of weak instruments (e.g., a test statistic of 64,400 for the estimates using a 300-kilometer radius). However, the KP tests which adjust for correlation in the errors result in much lower test statistics.

[45] The test statistic is 8,299 distributed $F_{336,19250825}$.

effects estimated by M2SLS using region-by-year fixed effects (Appendix J, Column 9) are somewhat lower and less significant than our baseline results using region-by-year-by-month fixed effects.

The second reason is identified in Dhrymes and Lleras-Muney (2006). Even with the same control variables in the first stage, M2SLS is more efficient because it uses disaggregated data in the first stage thereby utilizing more information; however, the grouping of the first-stage predicted values changes the nature of the first stage errors and their relationship to the second-stage errors which could decrease efficiency. We compare the efficiency of annual Wald 2SLS versus M2SLS by applying Theorem 4 of Dhrymes and Lleras-Muney (2006) to models that include region-by-year fixed effects in both stages of both estimation approaches.[46] We do this for increasing levels of aggregation (daily, weekly, monthly, quarterly, semiannual, and annual levels) in the first stage (second-stage estimates are all at the firm-year level) using the 300-kilometer radius and middle funnel. Table 6 shows the results. Fairly clear patterns emerge as the level of aggregation is increased. The first-stage coefficient declines in magnitude (and turns negative with annual aggregation) while the second-stage coefficients become less and less significant. The test statistics for efficiency are shown in the bottom row. The statistics are all well above the cutoff value of 1.64 for a 5% level of significance indicating M2SLS estimates are much more efficient than Wald 2SLS at all radiuses. The same gain in efficiency is likely to be achieved when applying M2SLS to other outcomes because of the much greater information in daily data than in annual. These results suggest that daily data are necessary to generate sufficient variation for precise estimates.

[Insert Table 6 here]

5. Conclusion

We provide a methodology for estimating the causal effect of air pollution spillovers on outcomes that are measured with lower frequency than pollution and weather data. Measuring air pollution spillovers requires high-frequency (such as daily) data to ensure that shifts in wind direction are properly captured, but outcome variables are often available on only an annual basis.

We proceed by estimating the pollution decay function at high frequency separately from the causal effects and estimating the causal effects using a mixed two-stage least squares (M2SLS) procedure using high-frequency changes in imported pollution from nearby cities as an instrument. The M2SLS procedure allows high-frequency data for the instrumenting in the first stage but low-frequency outcome data in the second stage. This estimation is a natural by-product of estimating the

[46] Appendix E shows how to calculate the inputs to this theorem.

spillover decay function since this also requires high-frequency wind and pollution data. We show that typical Wald 2SLS fails in estimating causal effects due to the aggregation of pollution data over a long period and the resulting loss of efficiency.

Use of high-frequency data also allows spillovers to be examined at relatively short distances while minimizing the chance of spurious correlation from regional and seasonal shocks to the outcome variable. This allows an examination of spillovers between cities that are geographically close but administratively distinct and therefore potentially suffer from a free-rider problem in pollution production.

Although previous papers document the presence of spillovers, our paper specifically quantifies how their intensity varies with distance— a necessary input for determining the scope of administrative control necessary to internalize externalities. PM_{10} spillovers in China are large and extend quite far suggesting the need to coordinate environmental policies at the supra-provincial level.

While we apply our procedure to quantify spillover effects of PM_{10} on productivity, our procedure can easily be adapted to estimate the spillover effects for other pollutants and on any outcome for which data are of a lower frequency than the pollution and weather data. For example, if only annual health measures are available the instrumenting technique works as long as daily pollution and weather data are available. It is also potentially applicable to estimating outcomes over periods longer than one year.

References

Adhvaryu, A., N. Kala, and A. Nyshadham. (2019). "Management and Shocks to Worker Productivity" NBER working paper #25865.

Altindag, D. T., D. Baek, and N. Mocan (2017). "Chinese Yellow Dust and Korean Infant Health," *Social Science & Medicine*, 186, 78 – 86.

Anderson, M. L. (2020). "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality," *Journal of the European Economic Association*, forthcoming.

Andrews, S. Q. (2008). "Inconsistencies in Air Quality Metrics: 'Blue Sky' Days and PM10 Concentrations in Beijing," *Environmental Research Letters*, 3, 1 – 14.

Angrist, J. D. and J. S. Pischke (2011). *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Angrist, J. D. and J. S. Pischke (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press.

Aragón, F., J. J. Miranda, and P. Oliva. (2017). "Particulate Matter and Labor Supply: The Role of Caregiving and Non-Linearities," *Journal of Environmental Economics and Management*, 86, 295 – 309.

Arzaghi, M. and J. V. Henderson (2008). "Networking off Madison Avenue," *Review of Economic Studies*, 75, 1011 – 1038.

Banzhaf, H. S. and B. A. Chupp (2012). "Fiscal Federalism and Interjurisdictional Externalities: New Results and an Application to US Air Pollution," *Journal of Public Economics*, 96, 449 – 464.

Becker, R. and J. Henderson. (2000). "Effects of Air Quality Regulations on Polluting Industries," *Journal of Political Economy*, 108, 379 – 421.

Bošković, B. (2015). "Air Pollution, Externalities, and Decentralized Environmental Regulation," working paper.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012). "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 97, 339 – 351.

Brunnermeier, S. and A. Levinson. (2004). "Examining the Evidence on Environmental Regulations and Industry Location," *Journal of Environment & Development*, 13, 6 – 41.

Cai, H., Y. Chen, and Q. Gong (2016). "Polluting thy Neighbor: Unintended Consequences of China' s Pollution Reduction Mandates," *Journal of Environmental Economics and Management*, 76, 86-104.

Cai, H. and Q. Liu. (2009). "Competition and Corporate Tax Avoidance: Evidence from Chinese Industrial Firms," *The Economic Journal*, 119, 764 – 795.

Chang, T., J. Graff Zivin, T. Gross, and M. Neidell (2016). "Particulate Pollution and the Productivity of Pear Packers," *American Economic Journal: Economic Policy, 8,* 141 – 69.

Chang, T., J. Graff Zivin, T. Gross, and M. Neidell (2019). "The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China," *American Economic Journal: Applied Economics,* 11, 151 – 172.

Chay, K. Y. and M. Greenstone. (2003). "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *Quarterly Journal of Economics*, 118, 1121 – 1167.

Chen, S., P. Oliva, and P. Zhang (2017). "The Effect of Air Pollution on Migration: Evidence from China," *NBER Working Paper 24036*.

Chen, S., P. Oliva, and P. Zhang. (2018). "The Effect of Air Pollution on Mental Health: Evidence from China," *NBER Working Paper 24686.*

Chen, Y., A. Ebenstein, M. Greenstone, and H. Lie. (2013). "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 of the United States of America*, 110, 12936 – 12941.

Coase, R. H. (1960). "The Problem of Social Costs," *Journal of Law and Economics*, 3, 1 – 44.

Cragg, J. G. and S. G. Donald (1993) "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222 – 240.

Dell, M., B. F. Jones, and B. A. Olken (2012). "Temperature Shocks and Economic Growth: Evidence from the Last Half Century," *American Economic Journal: Macroeconomics*, 4, 66 – 95.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, 109, 4178 – 4219.

Deschênes, O. and M. Greenstone (2011). "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US," *American Economic Journal: Applied Economics*, 3, 152 – 185.

Dhrymes, P. J. and A. Lleras-Muney (2006). "Estimation of Models with Grouped and Ungrouped Data by Means of '2SLS'," *Journal of Econometrics*, 133, 1 – 29.

Ebenstein, A., Fan, M., Greenstone, M., He, G., and Zhou, M. (2017). "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*, *114*(39), 10384 – 10389.

Eichner, T. and M. Runkel (2012). "Interjurisdictional Spillovers, Decentralized Policymaking, and the Elasticity of Capital Supply," *The American Economic Review*, 102, 2349 – 2357.

EPA (U.S. Environmental Protection Agency). (2004). "Air Quality Criteria for Particulate Matter."

Federal Register (2017). "Revisions to the Guideline on Air Quality Models: Enhancements to the AERMOD Dispersion Modeling System and Incorporation of Approaches to Address Ozone and Fine Particulate Matter," 40 CFR Part 51, 5182 – 5235.

Fell, H. and D. T. Kaffine (2014). "Can Decentralized Planning Really Achieve First-Best in the Presence of Environmental Spillovers?" *Journal of Environmental Economics and Management*, 68, 46 – 53.

Freeman, R., W. Liang, R. Song, and C. Timmins (2019). "Willingness to Pay for Clean Air in China," *Journal of Environmental Economics and Management*, 94, 188 – 216.

Fu, S. (2007). "Smart Café Cities: Testing Human Capital Externalities in the Boston Metropolitan Area," *Journal of Urban Economics,* 61, 86 – 111.

Fu, S., V. B. Viard, and P. Zhang (2021). "Air Pollution and Manufacturing Firm Productivity: Nationwide Estimates for China," *Economic Journal*, 131, 3241 - 3273.

Ghanem, D. and J. Zhang (2014). "'Effortless Perfection:' Do Chinese Cities Manipulate Air Pollution Data?" *Journal of Environmental Economics and Management*, 68, 203 – 225.

Goodkind *et al.* (2019). "Fine-Scale Damage Estimates of Particulate Matter Air Pollution Reveal Opportunities for Location-Specific Mitigation of Emissions," *PNAS*, 116, 8775 – 8780.

Graff Zivin, J. and M. Neidell. (2012). "The Impact of Pollution on Worker Productivity," *American Economic Review*, 102, 3652 – 3673.

Greenstone, M. (2002). "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and Census of Manufactures," *Journal of Political Economy*, 110, 1175 – 219.

Hanna, R. and P. Oliva. (2015). "The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico," *Journal of Public Economics*, 122, 68 – 79.

He, J., H. Liu, and A. Salvo (2019). "Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China," *American Economic Journal: Applied Economics*, 11, 173 – 201.

He, G., S. Wang, and B. Zhang (2020). "Watering down Environmental Regulation in China," *Quarterly Journal of Economics*, 135, 2135 – 2185.

Henderson, J. V. (2007). "Understanding Knowledge Spillovers," *Regional Science and Urban Economics*, 37, 497 – 508.

Herrnstadt, E., A. Heyes, E. Muehlegger, and S. Saberian (2019). "Air Pollution and Criminal Activity: Evidence from Chicago Microdata," working paper.

Hsiang, S. M. (2010). "Temperatures and Cyclones Strongly Associated with Economic Production in the Caribbean and Central America," *Proceedings of the National Academy of Sciences*, 107, 15367 – 15372.

Jia, R. (2017). "Pollution for Promotion," working paper.

Jia, R. and H. Ku (2019). "Is China's Pollution the Culprit for the Choking of South Korea? Evidence from the Asian Dust," *The Economic Journal*, 129, 3154 – 3188.

Jin, H., Y. Qian, and B. R. Weingast. (2005). "Regional Decentralization and Fiscal Incentives: Federalism, Chinese Style," *Journal of Public Economics*, 89, 1719 – 1742.

Jordan, B.W. (2016). "Can we Dig It? An Empirical Study of Environmental Attitudes and Social License to Operate in Mineral Extraction," working paper.

Kahn, M. E., P. Li, and D. Zhao (2015). "Water Pollution Progress at Borders: The Role of Changes in China's Political Promotion Incentives," *American Economic Journal: Economic Policy*, 7, 223 – 242.

Kleibergen, F. and Paap, R. (2006). "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97 – 126.

Lavy, V., A. Ebenstein, and S. Roth. (2014). "The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation," NBER Working Paper No. 20648.

Levinsohn, J. and A. Petrin. (2003). "Estimating Production Functions using Inputs to Control for Unobservables." *Review of Economic Studies*, 70, 317 – 342.

Levinson, A. (2012). "Valuing Public Goods using Happiness Data: The Case of Air Quality," *Journal of Public Economics*, 96, 869 – 880.

Li, H. and L. Zhou (2005). "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China," *Journal of Public Economics*, 89, 1743 – 1762.

Lin, Y. (2017). "The Long Shadow of Industrial Pollution: Environmental Amenities and the Distribution of Skills," working paper.

Lipscomb, M. and A. M. Mobarak (2017). "Decentralization and Pollution Spillovers: Evidence from the Re-drawing of County Borders in Brazil," *The Review of Economic Studies*, 84, 464 – 502.

Lleras-Muney, A. (2005). "The Relationship between Education and Adult Mortality in the U.S.," *Review of Economic Studies*, 72, 189 – 221.

Mani, M. and D. Wheeler. (1998). "In Search of Pollution Havens? Dirty Industry in the World Economy, 1960-1995," *Journal of Environment & Development*, 7, 215 – 247.

Massa, M. and A. Žaldokas (2014). "Investor Base and Corporate Borrowing: Evidence from International Bonds," *Journal of International Economics*, 92, 95 – 110.

Ministry of Civil Affairs (2017). "The Brochure of Administrative Divisions of Townships in China."

Moussiopoulos, N. *et al.* (2000). "Ambient Air Quality, Pollutant Dispersion, and Transport Models," *International Journal of Environment and Pollution*, European Topic Center on Air Quality Report #19/1996.

Neidell, M. J. (2004). "Air Pollution, Health, and Socio-Economic Status: The Effect of Outdoor Air Quality on Childhood Asthma," *Journal of Health Economics*, 23, 1209 – 1236.

Oates, W. E. (1972). *Fiscal Federalism*. New York: Harcourt, Brace Jovanovich.

Oates, W. E. and R. M. Schwab (1988). "Economic Competition among Jurisdictions: Efficiency Enhancing or Distortion Inducing?" *Journal of Public Economics*, 35, 333 – 354.

OECD Directorate for Science, Technology and Industry. (2011). "ISIC Rev. 3 Technology Intensity Definition, https://www.oecd.org/sti/ind/48350231.pdf.

Ogawa, H. and D. E. Wildasin (2009). "Think Locally, Act Locally: Spillovers, Spillbacks, and Efficient Decentralized Policymaking," *American Economic Review*, 99, 1206 – 1217.

Olley, G. S., and A. Pakes. (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263 – 1297.

Ostro, B. *et al*. (2018). "Assessing the Recent Estimates of the Global Burden of Disease for Ambient Air Pollution: Methodological Changes and Implications for Low- and Middle-Income Countries," *Environmental Research*, 166, 713 – 725.

Pun, V. C., J. Manjouride, and H. Suh. (2016). "Association of Ambient Air Pollution with Depressive and Anxiety Symptoms in Older Adults: Results from the NSHAP Study, Environmental Health Perspectives," working paper.

Rangel, M. A. and T. Vogl (2019). "Agricultural Fires and Health at Birth," *Review of Economics and Statistics*, 101, 616 – 630.

Rosenthal, S. S. and W. C. Strange (2003). "Geography, Industrial Organization, and Agglomeration," *The Review of Economics and Statistics*, 85, 377 – 393.

Rosenthal, S. S. and W. C. Strange (2008). "The Attenuation of Human Capital Spillovers," *Journal of Urban Economics*, 64, 373 – 389.

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

Seigneur, C. and R. Dennis (2011). "Atmospheric Modeling" in *Technical Challenges of Multipollutant Air Quality Management*, London: Springer Dordecht Heidelberg.

Seigneur, C. and M. Moran (2004). "Chemical Transport Models" in *Particulate Matter Science for Policy Makers*, Cambridge: Cambridge University Press.

Sheldon, T. L. and C. Sankaran (2017). "The Impact of Indonesian Forest Fires on Singaporean Pollution and Health," *American Economic Review: Papers & Proceedings*, 107, 526 – 529.

Sigman, H. (2002). "International Spillovers and Water Quality in Rivers: Do Countries Free Ride?" *American Economic Review*, 92, 1152 – 1159.

Sigman, H. (2005). “Transboundary Spillovers and Decentralization of Environmental Policies,” *Journal of Environmental Economics and Management*, 50, 82 – 101.

Stock, J. H. and M. Yogo. (2005). “Testing for Weak Instruments in Linear IV Regression,” in *Andrews DWK Identification and Inference for Econometric Models*. New York: Cambridge University Press, 80 – 108.

Syverson, C. (2011). “What Determines Productivity,” *Journal of Economic Literature*, 49, 326 – 365.

Wang, S. and Z. Wang (2021). “The Environmental and Economic Consequences of Internalizing Border Spillovers,” working paper.

Williams, R. C. (2012). “Growing State–Federal Conflicts in Environmental Policy: The Role of Market-Based Regulation,” *Journal of Public Economics*, 96, 1092 – 1099.

World Health Organization. (2006). *WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide: Global Update 2005*, Geneva, Switzerland.

Xu, C. (2011). “The Fundamental Institutions of China’s Reform and Development," *Journal of Economic Literature*, 49, 1076 – 1151.

Yu, M. (2014). “Processing Trade, Tariff Reductions and Firm Productivity: Evidence from Chinese Firms,” *The Economic Journal*, 125, 943 – 988.

Zhang, P., J. Zhang, O. Deschênes, and K. Meng (2018). “Temperature Effects on Productivity and Factor Reallocation: Evidence from a Half Million Chinese Manufacturing Plants,” *Journal of Environmental Economics and Management*, 88, 1 – 17.

Zheng, S., J. Cao, M. Kahn, and C. Sun (2014). “Real Estate Valuation and Cross-Boundary Air Pollution Externalities: Evidence from Chinese Cities,” *Journal of Real Estate Finance Economics*, 48, 398 – 414.

Figure 1: Example of wind directions between nearby and focal city included in pollution decay function estimation

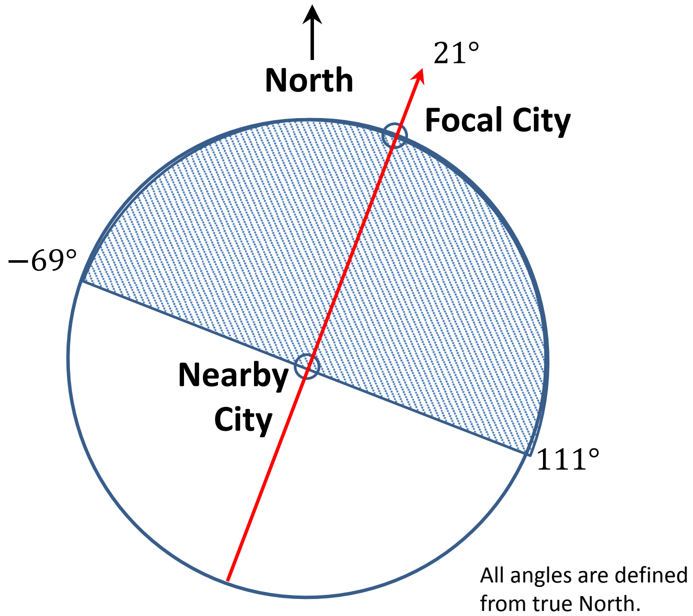


Figure 2: Example of wind directions included in estimating the causal effects of pollution on productivity (middle funnel)

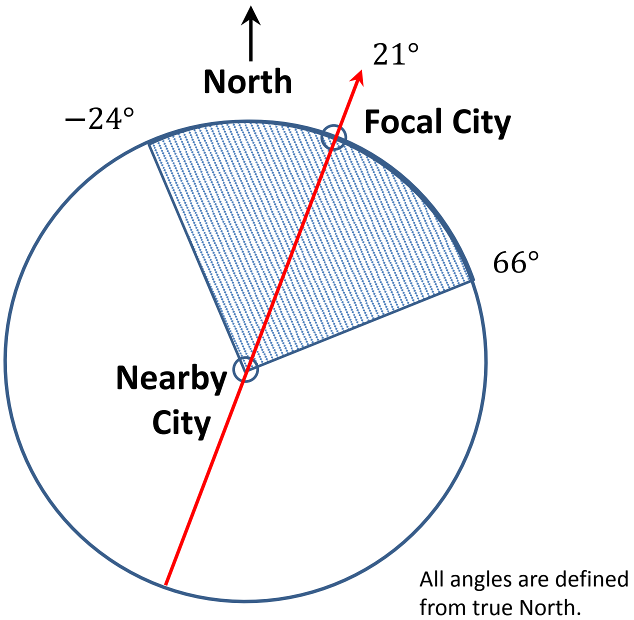
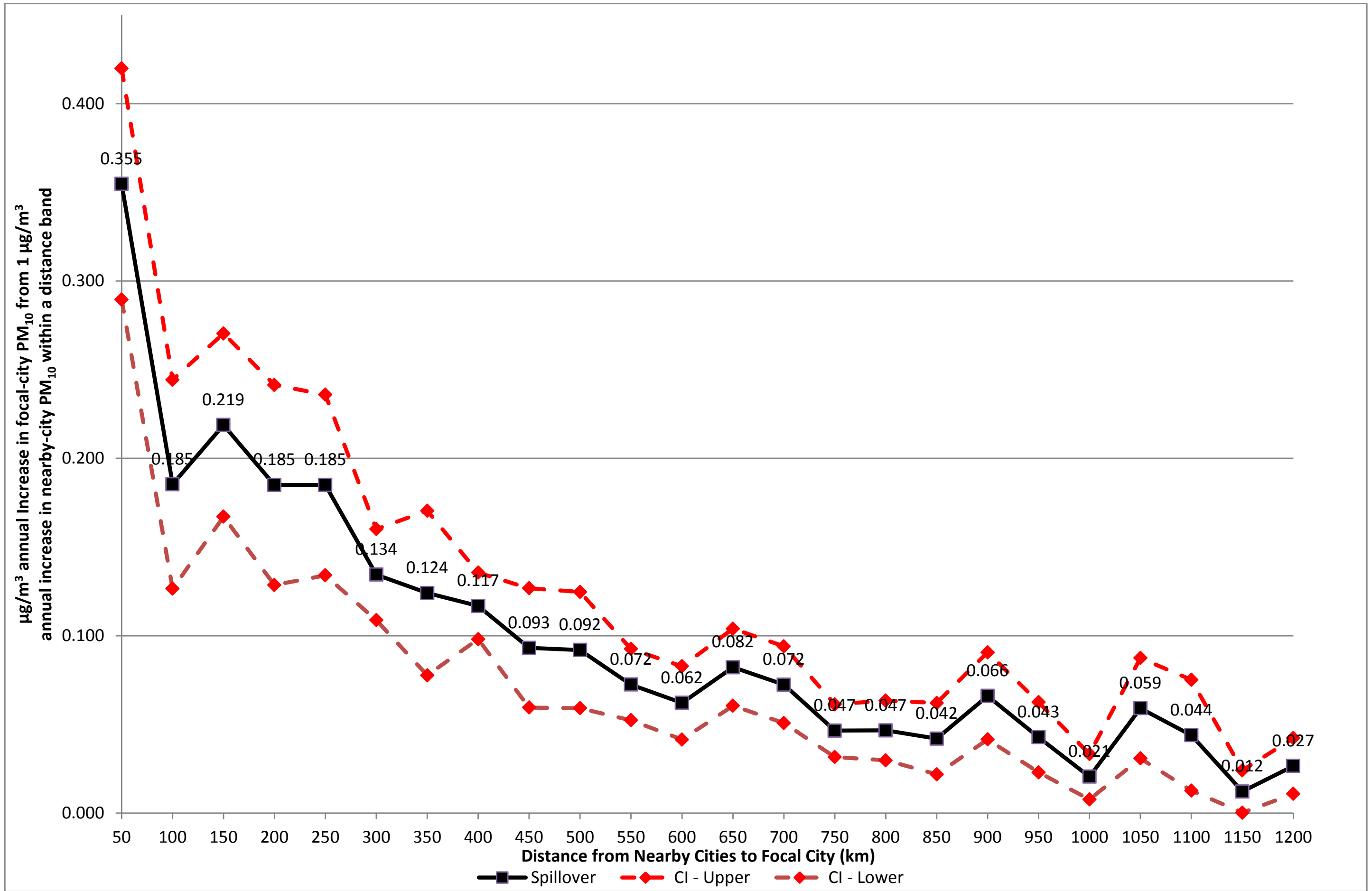
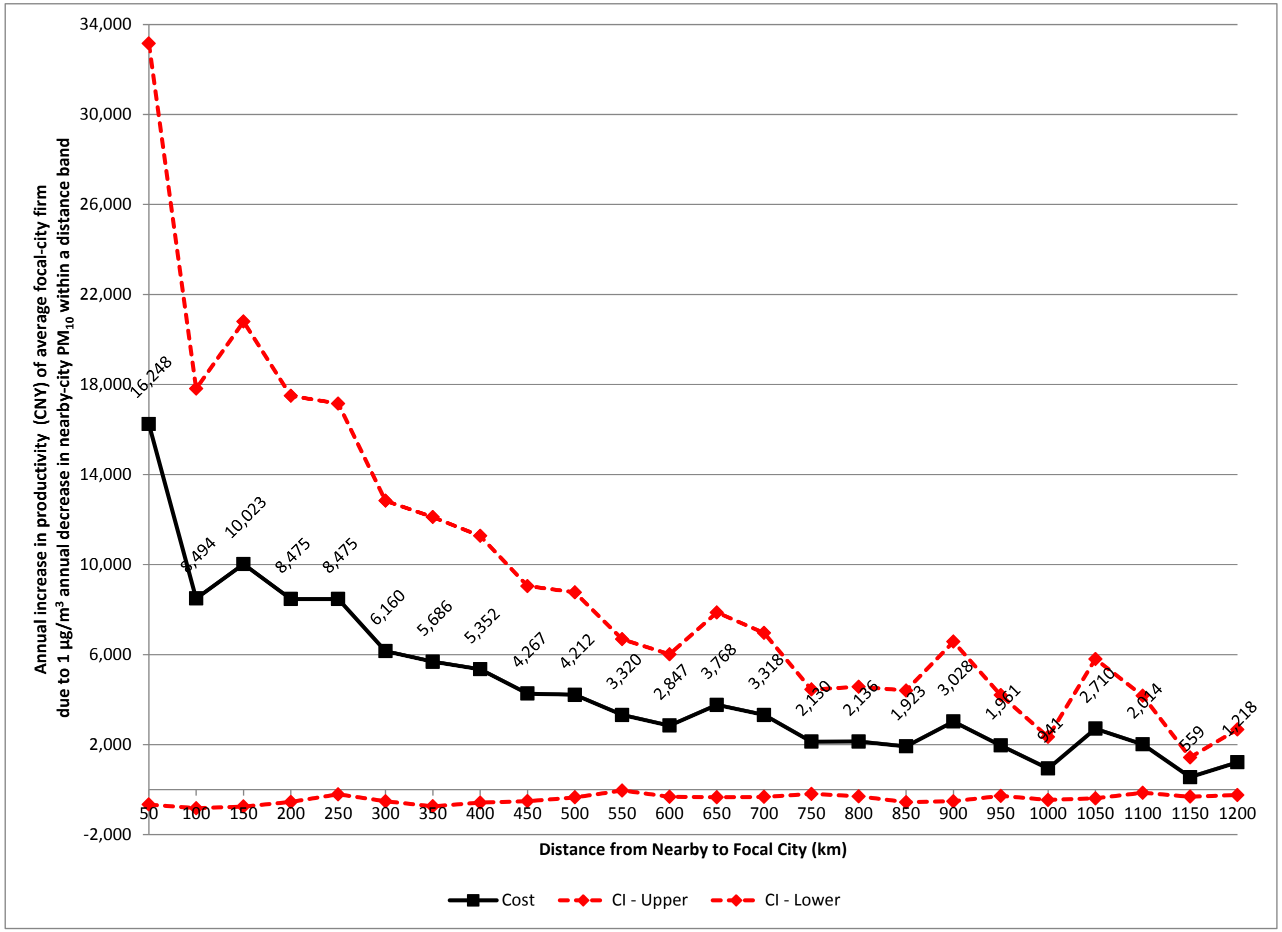


Figure 3: Pollution decay function: effect of one μg/m³ annual increase in nearby-city PM_{10} within a distance band on annual focal-city PM_{10} as a function of distance



Solid, black line shows effect of a one μg/m³ annual increase in nearby-city PM_{10} within a distance band (holding pollution in all other distance bands constant) on annual focal-city PM_{10} as a function of distance controlling for weather variables, focal-city fixed effects, and region-by-year-by-month fixed effects. Estimation allows for piecewise linear effects in increments of 50 kilometers. Effects are adjusted for the empirical distribution of wind directions during the year. Dashed, red lines show 95% confidence intervals estimated using 100 iterations of a block bootstrap by focal city.

Figure 4: Air pollution spillover effects from a one μg/m³ annual decrease in nearby-city PM_{10} within a distance band on average annual labor productivity of focal-city firms as a function of distance



Solid, black line shows effect of a one μg/m³ annual decrease in nearby-city PM_{10} within a distance band (holding pollution in all other distance bands constant) on average annual productivity of focal-city firms as a function of distance estimated by the two-step procedure described in the text. Estimation allows for piecewise linear effects in increments of 50 kilometers. Effects are adjusted for the empirical distribution of wind directions during the year. Dashed, red lines show 95% confidence intervals estimated using 100 iterations of a block bootstrap by focal city.

Table 1: Summary statistics for pollution decay function estimation 2001 to 2007 (N = 988,320)

	(1) Mean	(2) Std. dev.	(3) Min	(4) Max
Focal city PM_{10} (μg/m^3)	97.5	59.5	8.0	600.0
Distance between focal/nearby city (km)	1,003.9	444.0	44.0	1,799.2
Nearby cities per focal city	43.1	11.8	2.0	56.0
Fraction of days wind toward focal city		52.1%		
Fraction of days API = PM_{10}		91.9%		
# of focal/nearby cities		60		
# of focal-nearby city-year pairs		2,586		

Table 2: Summary statistics for M2SLS estimation 2001 to 2007 (150- and 300-kilometer maximum distances)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
	150 kilometers proximity				**300 kilometers proximity**			
First-stage sample (firm-day)								
	(N = 16,271,706)				**(N = 19,339,917)**			
Focal city PM_{10} ($\mu g/m^3$)	111.6	69.0	10.0	600.0	110.5	67.8	10.0	600.0
Nearby city PM_{10} ($\mu g/m^3$)	97.5	65.2	11.0	600.0	97.2	63.2	11.0	600.0
Distance between focal/nearby city (km)	89.2	28.5	44.0	143.8	106.5	50.8	44.0	291.8
# of city-years	103				166			
# of focal cities	30				47			
Second-stage sample (firm-year)								
	(N = 243,368)				**(N = 291,339)**			
Value added (CNY1,000)	15,181.5	27,121.6	105.7	357,934.3	15,269.8	27,296.6	101.3	366,425.6
Total workers	166.9	244.7	10.0	3,012.0	171.6	252.9	10.0	3,012.0
Value added per worker (CNY1,000)	119.7	216.2	0.5	16,247.6	118.9	219.9	0.1	16,247.6
# of firms	75,390				88,716			

Summary statistics for data used in M2SLS estimation of causal effect of local air pollution on local firms' labor productivity. First-stage data is conditional on wind blowing toward the focal city.

Table 3: Effect of local PM_{10} on local labor productivity – OLS and M2SLS estimates using nearest-nearby city pollution within middle funnel and different maximum distances as an instrument

	(1)	(2)	(3)	(4)
	Maximum distance cutoff			
	150 km	**200 km**	**250 km**	**300 km**
Panel A: OLS (firm-year sample)				
Dependent variable:	**ln(value added/worker)**			
Mean annual focal city PM_{10}	-0.0015	-0.0003	-0.0005	-0.0005
	(0.0014)	(0.0014)	(0.0013)	(0.0013)
R^2	0.0738	0.0777	0.0740	0.0839
Sample size	243,368	264,746	276,528	291,339
Panel B: M2SLS first stage (firm-day sample)				
Dependent variable:	**Daily focal city PM_{10}**			
Daily nearby city PM_{10}	0.7172***	0.7025***	0.7004***	0.6959***
	(0.0756)	(0.0708)	(0.0687)	(0.0669)
Fraction of days wind toward focal city	0.246	0.248	0.250	0.246
KP *F*-statistic	90.0	98.4	104.0	108.1
# cities	30	40	44	47
Sample size	16,271,706	17,858,505	18,758,702	19,339,917
Panel C: M2SLS second stage (firm-year sample)				
Dependent variable:	**ln(value added/worker)**			
Mean annual predicted focal city PM_{10}	-0.0019	-0.0026*	-0.0034**	-0.0030**
	(0.0015)	(0.0014)	(0.0015)	(0.0014)
Implied elasticity	-0.198	-0.289	-0.351	-0.312
# firms	75,390	82,714	86,941	88,716
Sample size	243,368	264,746	276,528	291,339

Data included in Panel A corresponds to firm-year data included in Panel C. First stage models include firm and region-by-year-by-month fixed effects; linear and quadratic terms of daily humidity and wind speed; and categorial variables for temperature bins as described in the text. The OLS and second-stage models include firm and region-by-year fixed effects; annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorial variables for temperature (i.e., number of days in each temperature bin). OLS R^2 is the "within" R^2 from the fixed effects regression. Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in Panel C are also adjusted for two-stage estimation using 100 block-bootstrap iterations.

Table 4: Robustness of causal effect of local PM_{10} on productivity – M2SLS estimates allowing for autocorrelation in focal-city productivity and using TFP as productivity measure

	(1)	(2)	(3)	(4)	(5)
	Middle funnel, 300-kilometer maximum distance cutoff				
Panel A: M2SLS first stage (firm-day sample)					
Dependent variable:	**Daily focal city PM_{10}**				
Daily nearby city PM_{10}	0.6959***	0.5804***	0.5549***	0.6959***	0.6959***
	(0.0669)	(0.0487)	(0.0462)	(0.0669)	(0.0669)
Lagged, daily focal city PM_{10}		0.3283***	0.3273***		
		(0.0287)	(0.0315)		
Second lagged, daily focal city PM_{10}			0.0024		
			(0.0152)		
Fraction of days wind toward focal city	0.246	0.246	0.246	0.246	0.246
KP F-statistic	108.1	465.7	372.24	108.1	108.1
# cities	47	47	47	47	47
Sample size	19,339,917	16,958,644	15,137,125	19,339,917	19,339,917
Panel B: M2SLS second stage (firm-year sample)					
Dependent variable:	**ln(value added/worker)**			**TFP (OP)**	**TFP (LP)**
Mean annual predicted focal city PM_{10}	-0.0030**	-0.0032*	-0.0034***	-0.0028**	-0.0028**
	(0.0014)	(0.0017)	(0.0013)	(0.0014)	(0.0014)
Implied elasticity	-0.312	-0.333	-0.354	-0.291	-0.291
# firms	88,716	88,716	88,716	88,716	88,716
Sample size	291,339	291,339	291,339	291,339	291,339

All columns use a 300-kilometer maximum radius when choosing nearest nearby city and apply the middle funnel in choosing days when wind blows toward focal city. All first-stage models include firm fixed effects, region-by-year-by-month fixed effects and linear and quadratic terms of daily humidity and wind speed; and categorial variables for temperature bins as described in the text in the first stage. All second-stage models include region-by-year fixed effects and annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorial variables for temperature (i.e., number of days in each temperature bin). Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Second-stage standard errors are also adjusted for two-stage estimation using 100 block-bootstrap iterations.

Table 5: Wald 2SLS estimates of causal effect of local PM_{10} on local labor productivity using pollution of nearest-nearby city within middle funnel and different maximum distances as an instrument

	(1)	(2)	(3)	(4)
	Maximum distance cutoffs			
	150 km	**200 km**	**250 km**	**300 km**
Panel A: 2SLS first stage (firm-year sample)				
Dependent variable:	**Mean annual focal city PM_{10}**			
Mean annual nearby city PM_{10} (conditional on wind blowing toward focal city)	-0.2339**	-0.2680***	-0.2435***	-0.2562***
	(0.1005)	(0.0787)	(0.0786)	(0.0637)
KP *F*-statistic	5.4	11.6	9.6	16.2
# cities	30	40	44	47
Sample size	243,368	264,746	276,528	291,339
Panel B: 2SLS second stage (firm-year sample)				
Dependent variable:	**Focal city ln(value added/worker)**			
Mean annual predicted focal city PM_{10}	0.0026	0.0052	0.0080	0.0065
	(0.0038)	(0.0032)	(0.0055)	(0.0040)
# firms	75,390	82,714	86,941	88,716
Sample size	243,368	264,746	276,528	291,339

All models include include firm and region-by-year fixed effects; annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorial variables for temperature (i.e., number of days in each temperature bin) in both stages. Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in Panel B are also adjusted for two-stage estimation.

Table 6: Estimates of causal effect of local PM_{10} on local labor productivity at different levels of aggregation in the first stage

	(1)	(2)	(3)	(4)	(5)	(6)
	Middle funnel, 300-kilometer maximum distance cutoff					
Panel A: First stage:	**Firm-Day M2SLS**	**Firm-Week M2SLS**	**Firm-Month M2SLS**	**Firm-Quarter M2SLS**	**Firm-Semi-Annual M2SLS**	**Firm-Year 2SLS**
Dependent variable:	**Focal city PM_{10}**					
Nearby city PM_{10}	0.7572***	0.7262***	0.7255***	0.5954***	0.5160***	-0.2562***
	(0.0827)	(0.0846)	(0.1133)	(0.0722)	(0.0633)	(0.0637)
Fraction of days wind toward focal city	0.246	0.246	0.246	0.246	0.246	0.246
KP F-statistic	83.9	73.6	41.0	68.0	66.5	16.2
# cities	47	47	47	47	47	47
Sample size	19,339,917	9,190,704	3,182,582	1,162,124	582,678	291,339
Panel B: second stage (firm-year sample)						
Dependent variable:	**ln(value added/worker)**					
Mean annual predicted focal city PM_{10}	-0.0021*	-0.0024	-0.0017	-0.0020	-0.0024	0.0065
	(0.0012)	(0.0015)	(0.0015)	(0.0017)	(0.0019)	(0.0040)
# firms	88,716	88,716	88,716	88,716	88,716	88,716
Sample size	291,339	291,339	291,339	291,339	291,339	291,339
M2SLS efficiency (test statistic)	1,735.1	915.5	319.3	115.6	72.8	

All columns use a 300-kilometer radius and the middle funnel in choosing days when wind blows toward focal city. Columns 1 through 5 use M2SLS to estimate at different levels of aggregation in the first stage: daily in Column 1, weekly in Column 2, monthly in Column 3, quarterly in Column 4, and semi-annually in Column 5 - and data at the annual level in the second stage. Column 6 estimates using Wald 2SLS with data at the annual level in both stages. First-stage models include firm and region-by-year fixed effects; linear and quadratic terms of daily humidity and wind speed; and categorial variables for temperature bins as described in the text aggregated to the corresponding level. Second-stage models include firm and region-by-year fixed effects; annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorial variables for temperature (i.e., number of days in each temperature bin). Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Second-stage standard errors are also adjusted for two-stage estimation. In Columns 1 through 5 this is done using 100 block-bootstrap iterations. M2SLS efficiency test statistic based on Dhrymes and Lleras-Muney (2006) Theorem 4.

Online Appendix A: Cities included in analysis

Beihai	Nanchong
Beijing	Nanjing
Changchun	Nanning
Changde	Qingdao
Changsha	Qinhuangdao
Chengdu	Qiqihaer
Chifeng	Quanzhou
Chongqing	Qujing
Dalian	Rizhao
Datong	Shanghai
Fushun	Shantou
Fuzhou	Shaoguan
Guangzhou	Shaoxing
Guilin	Shenyang
Guiyang	Shijiazhuang
Haerbin	Taian
Haikou	Taiyuan
Hangzhou	Tianjin
Hefei	Weifang
Huhehaote	Weinan
Jinan	Wuhan
Jining	Wuhu
Jiujiang	Wulumuqi
Kelamayi	Xiamen
Kunming	Xian
Lanzhou	Xining
Lhasa	Yinchuan
Lianyungang	Yuxi
Mianyang	Zhanjiang
Mudanjiang	Zhengzhou

All cities used in spillover decay function estimation.

Online Appendix B: Conversion from API to PM_{10}

API	PM_{10}	Conversion formula
0 – 50	0 – 50	$API = PM_{10}$
50 – 200	50 – 350	$API = (1/2)*PM_{10} + 25$
200 – 300	350 – 420	$API = (10/7)*PM_{10} – 300$
300 – 400	420 – 500	$API = (5/4)*PM_{10} – 225$
400 – 500	500 – 600	$API = PM_{10} – 100$

Based on Andrews (2008).

Online Appendix C: Distribution of day-to-day wind direction changes for all cities included in estimating the spillover decay function 2001 – 2007 (60 focal cities, N = 52,940)

Difference in wind direction day-to-day (degrees)	Percentage of days	Cumulative percentage
10	18.0%	18.0%
20	13.6%	31.6%
30	10.6%	42.1%
40	8.6%	50.7%
50	6.6%	57.4%
60	5.5%	62.9%
70	4.7%	67.6%
80	4.1%	71.7%
90	3.6%	75.3%
100	3.3%	78.5%
110	3.2%	81.7%
120	2.9%	84.6%
130	2.7%	87.3%
140	2.6%	89.9%
150	2.5%	92.5%
160	2.5%	94.9%
170	2.6%	97.5%
180	2.5%	100.0%

Percentage of days that wind direction changes day-to-day in ten-degree brackets. Includes all cities and days used in estimating the spillover decay function.

Online Appendix D: Definition of regions

Geographic regions	Provinces
North	Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia
Northeast	Liaoning, Jilin, Heilongjiang
East	Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong
Central	Henan, Hubei, Hunan
South	Guangdong, Guangxi, Hainan
Southwest	Chongqing, Sichuan, Guizhou, Yunan, Tibet
Northwest	Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang
Based on Zhang *et al* . (2018).	

Online Appendix E: A comparison of OLS, Wald 2SLS, and M2SLS estimators in the presence of monthly regional shocks to focal- and nearby-city pollution

This appendix examines three different estimators in the presence of a monthly regional shock to focal- and nearby-city pollution. Ordinary least squares (OLS) results in biased estimates of pollution's effect on the outcome due to endogeneity bias regardless of the common shock. Using daily nearby-city pollution, conditional on wind direction, as an instrument eliminates the endogeneity bias using either Wald 2SLS or M2SLS; however, Wald 2SLS will result in biased first-stage coefficients. Although the model assumes monthly shocks, the results extend to shocks of any duration longer than a day.

Relative to the model in the paper, we generalize by allowing for any annual outcome. We also simplify in a number of ways for transparency. The model: a) abstracts from firms to focus on the identification issues which occur at the city level, b) does not include weather control variables but they can be added without altering the results, c) and does not weight wind direction by the cosine of the angle. We abstract from constants in all equations to simplify the exposition.

Model

Daily outcome in focal-city f on day d of year t is given by:

$$Y_{td}^{f} = \theta P_{td}^{f} + \epsilon_{td}^{f}, \text{ (E1)}$$

where θ is the effect of daily pollution $\left(P_{td}^{f}\right)$ on the outcome. ϵ_{td}^{f} reflects daily unobservables affecting the outcome assumed to be independent of all other variables and identically distributed within and across cities.

Daily focal- and nearby-city pollution levels on day d in month m of year t are given by:

$$P_{td}^{f} = \mu Y_{td}^{f} + \rho \mathrm{I}_{td}^{fn} P_{td}^{n} + \xi_{tm}^{r} + \varepsilon_{td}^{f} \text{ (E2a)}$$
$$P_{td}^{n} = \mu Y_{td}^{n} + \rho\left(1 - \mathrm{I}_{td}^{fn}\right) P_{td}^{f} + \xi_{tm}^{r} + \varepsilon_{td}^{n}, \text{ (E2b)}$$

where I_{td}^{fn} is an indicator variable equal to one if the wind is blowing toward the focal city and zero when it is not. $\mu > 0$ captures the effect of output on pollution (pollution production) and $\rho > 0$ captures the inter-city transport between the two cities assumed to be symmetric for simplicity. ξ^r_{tm} is a monthly regional shock to pollution for all cities in the region, uncorrelated with all other variables and unobserved by the econometrician. For simplicity we assume that all focal-nearby city pairs are contained within a single region and each pair is self-contained (the other firm is its nearest nearby city). ε^f_{td} and ε^n_{td} reflect daily unobservables affecting pollution assumed to be independent of each other and all other variables and identically distributed within and across cities.

***Assumption 1*:**

Daily wind direction $\left(\mathrm{I}^{fn}_{td}\right)$ is uncorrelated with the common shock to pollution (ξ^r_{tm}). This also implies it is uncorrelated with the outcome $\left(Y^f_{td}\right)$ and the pollution levels (P^f_{td} and P^n_{td}).

OLS Estimator

OLS estimates yield biased estimates of pollution's effect on the outcome due to simultaneity bias.

An OLS regression using annual data with region-by-year fixed effects yields the following demeaned regression (net of region-by-year fixed effects) based on Equation (E1):

$$\ddot{Y}^f_t = \theta_{OLS}\ddot{P}^f_t + \ddot{\epsilon}^f_t, \text{ (E3)}$$

where umlauts indicate deviations from annual regional values for a variable x: $\ddot{x}^f_t = \frac{1}{D}\sum_d x^f_{td} - \frac{1}{FD}\sum_{f\in r}\sum_d x^f_{td}$. D is the number of days in a year and F is the number of focal firms in the region.

Average annual focal-city pollution (based on Equation (E2a)) net of region-by-year fixed effects is:

$$\ddot{P}^f_t = \mu\ddot{Y}^f_t + \rho\ddot{P}^n_t + \ddot{\varepsilon}^f_t, \text{ (E4)}$$

$\ddot{\xi}^r_t$ is differenced out by the region-by-year fixed effects.

The coefficient on pollution (θ_{OLS}) is biased because the error term in Equation (E3) is correlated with focal-city pollution (i.e., $Cov\left(\ddot{P}^f_t, \ddot{\epsilon}^f_t\right) \neq 0$) due to the simultaneity of the outcome and pollution. To see this, substitute $\ddot{Y}^f_t$ from Equation (E3) into Equation (E4).

Wald 2SLS Estimator

We now show that annual Wald 2SLS using nearby-city pollution (conditional on wind direction) as an instrument yields an unbiased estimate of pollution's effect on the outcome although the first-stage coefficient is biased. To illustrate the properties of the Wald 2SLS estimates we simplify by considering two cities (A and B) in a region.

The first-stage of Wald 2SLS regresses annual focal- on annual nearby-city pollution where averages are computed conditional on wind blowing toward the focal city ($\mathrm{I}^{fn}_{td} = 1$) and net of region-by-year fixed effects:

$$\breve{P}^f_t = \rho_{2SLS}\breve{P}^n_t + \breve{\xi}^r_t + \varpi^f_t, \text{ (E5)}$$

where upside-down umlauts indicate deviations from annual regional values conditional on wind blowing toward the focal city for a variable x: $\breve{x} = \frac{1}{D}\sum_d \mathrm{I}_{td}^{kl} x - \frac{1}{FD}\sum_{k \in r}\sum_d \mathrm{I}_{td}^{kl} x\,;\, k \neq l$. ϖ_t^f reflects measurement error assumed to be independent from all other variables and identically distributed within and across cities. The average annual shock in the region is:

$$\frac{1}{2D}\sum_d \left(\mathrm{I}_{td}^{AB}\xi_{tm}^r + \mathrm{I}_{td}^{BA}\xi_{tm}^r\right). \text{ (E6)}$$

The first term in brackets is the contribution of days when wind blows toward city A as focal city and the second when it blows toward city B as the focal city. Since wind direction is uncorrelated with the monthly regional shock (Assumption 1) this simplifies to:

$$\frac{1}{2}\sum_d \xi_{tm}^r. \text{ (E6')}$$

$\breve{\xi}_t^r$ contains the focal city's unobservable less the average unobservable in the region given by Equation (E6):

$$\frac{1}{D}\sum_d \mathrm{I}_{td}^{ij}\xi_{tm}^r - \frac{1}{2D}\sum_d \left[\mathrm{I}_{td}^{ij}\xi_{tm}^r + \mathrm{I}_{td}^{ji}\xi_{tm}^r\right] = \frac{1}{2D}\sum_d \left[\mathrm{I}_{td}^{ij}\xi_{tm}^r - \mathrm{I}_{td}^{ji}\xi_{tm}^r\right];\ i,j \in \{A,B\}, i \neq j. \text{ (E7)}$$

If there is no seasonality in wind direction this simplifies to:

$$\breve{\xi}_t^r = \frac{1}{2D}\sum_d \xi_{tm}^r\left(\bar{\mathrm{I}}_t^{ij} - \bar{\mathrm{I}}_t^{ji}\right);\ i,j \in \{A,B\}, i \neq j, \text{ (E7')}$$

where $\bar{\mathrm{I}}_t^{kl} = \sum_d \mathrm{I}_{td}^{kl}$.

The first stage coefficient in Wald 2SLS is given by:

$$\rho_{2SLS} = \frac{Cov\left(\breve{P}_t^f, \breve{P}_t^n\right)}{Var\left(\breve{P}_t^n\right)} = \frac{Cov\left(\frac{1}{D}\sum_d \mathrm{I}_{td}^{fn}P_{td}^f, \frac{1}{D}\sum_d \mathrm{I}_{td}^{fn}P_{td}^n\right)}{Var\left(\frac{1}{D}\sum_d \mathrm{I}_{td}^{fn}P_{td}^n\right)}. \text{ (E8)}$$

Separating focal- and nearby-city pollution on the same days versus different days this equals:

$$\frac{1}{D}\sum_d \left[\frac{Cov\left(\mathrm{I}_{td}^{fn}P_{td}^f, \mathrm{I}_{td}^{fn}P_{td}^n\right)}{Var\left(\mathrm{I}_{td}^{fn}P_{td}^n\right)} + \sum_{k \neq d}\frac{Cov\left(\mathrm{I}_{tk}^{fn}P_{tk}^f, \mathrm{I}_{td}^{fn}P_{td}^n\right)}{Var\left(\mathrm{I}_{td}^{fn}P_{td}^n\right)}\right]. \text{ (E9)}$$

Using Equation (E5), the first term in brackets equals:

$$\frac{Cov\left(\mathrm{I}_{td}^{fn}\left(\rho P_{td}^n + \breve{\xi}_t^r + \varpi_t^f\right), \mathrm{I}_{td}^{fn}P_{td}^n\right)}{Var\left(\mathrm{I}_{td}^{fn}P_{td}^n\right)} = \rho + \frac{Var\left(\breve{\xi}_t^r\right)}{Var\left(\mathrm{I}_{td}^{fn}P_{td}^n\right)}. \text{ (E10)}$$

The presence of the monthly regional shock in both focal- and nearby-city pollution biases ρ_{2SLS} upward unless there is no seasonality in the shocks or annual wind directions are perfectly balanced $\left(\bar{\mathrm{I}}_t^{kl} \neq \bar{\mathrm{I}}_t^{lk}\right)$ for all city pairs kl (see Equation (E7')). This occurs because the region-by-year fixed effects do not perfectly control for the monthly regional shocks to focal- and nearby-city pollution. Substituting Equation (E10) into (E9):

$$\rho_{2SLS} = \rho + \frac{1}{D}\sum_d \left[\frac{Var\left(\breve{\xi}_t^r\right)}{Var\left(\mathrm{I}_{td}^{fn}P_{td}^n\right)} + \sum_{k \neq d}\frac{Cov\left(\mathrm{I}_{tk}^{fn}P_{tk}^f, \mathrm{I}_{td}^{fn}P_{td}^n\right)}{Var\left(\mathrm{I}_{td}^{fn}P_{td}^n\right)}\right]. \text{ (E11)}$$

The last term in brackets is the correlation of focal- and nearby-city pollution (conditional on wind direction) on different days during the year. This could bias the coefficient either upward (if "clean" focal cities are located near "clean" nearby cities) or downward (if "clean" focal cities are located near "dirty" nearby cities or vice versa). This is an example of the ecological fallacy in which the relationship at the aggregate level is misleading about the relationship at the individual level.

The first-stage coefficient is unbiased only if both terms in brackets in Equation (E11) equal zero: an absence of regional monthly shocks to pollution and no systematic covariance in annual-level pollution across city pairs. Appendix N shows that in our data, "clean" focal cities tend to be located near "dirty" nearby cities or vice versa and this effect overwhelms the upward bias due to the monthly regional shock. As a result, the first-stage Wald coefficient is negative.

Regardless of whether the first-stage coefficient is biased or not, Wald 2SLS controls for endogeneity and leads to unbiased estimates in the second stage. The fitted values from Equation (E5) $\left(\widehat{\breve{P}_t^f}\right)$ are used in the second-stage regression to predict the annual focal-city outcome. Including region-by-year fixed effects yields the following demeaned second-stage regression based on Equation (E1):

$$\ddot{Y}_t^f = \theta_{2SLS}\widehat{\breve{P}_t^f} + \ddot{\epsilon}_t^f. \text{ (E12)}$$

This is unbiased since the fitted values $\left(\widehat{\breve{P}_t^f}\right)$ are purged of $\ddot{Y}_t^f$ eliminating the simultaneity bias (i.e., $Cov\left(\widehat{\breve{P}_t^f}, \ddot{\epsilon}_t^f\right) = 0$). This is true regardless of the whether the first-stage coefficient is biased or not. Therefore, the exclusion restriction is met. The inclusion restriction is also met since $Cov\left(\breve{P}_t^f, \breve{P}_{td}^n\right) > 0$.

M2SLS Estimator

We now show that M2SLS using nearby-city pollution (conditional on wind direction) as an instrument yields an unbiased estimate of pollution on the outcome. It also yields an unbiased coefficient in the first-stage.

The first-stage of M2SLS regresses daily focal- on daily nearby-city pollution conditional on wind blowing toward the focal city $\left(\mathrm{I}_{td}^{fn} = 1\right)$. Because daily data is used, region-by-year-by-month fixed effects can be included and the equation net of these is:

$$\tilde{P}_{td}^f = \rho_{M2SLS}\tilde{P}_{td}^n + \iota_{td}^f. \text{ (E13)}$$

where tildes indicate deviations from monthly regional values conditional on wind blowing toward the focal city: $\tilde{P}_{td}^k = \frac{1}{M}\sum_{d\in m}\mathrm{I}_{td}^{kl}P_{td}^j - \frac{1}{FM}\sum_{k\in r}\sum_{d\in m}\mathrm{I}_{td}^{kl}P_{td}^k$; $k, l \in \{f, n\}$; $k \neq l$ and M is the number of days in a month. ι_{td}^f reflects measurement error assumed to be independent from all other variables and identically distributed within and across cities.

The common shock to pollution is absorbed by the region-by-year-by-month fixed effects since $\mathrm{I}_{td}^{fn} = 1$ and therefore $\xi_{tm}^r - \frac{1}{FM}\sum_{f\in r}\mathrm{I}_{td}^{fn}\xi_{tm}^r = 0$. This is the key difference from the first stage of Wald 2SLS.

The fitted values from this regression averaged over the year for each city $\left(\overline{\widehat{\tilde{P}_{td}^f}}\right)$ are used in the second-stage regression to predict the annual focal-city outcome:

$$\ddot{Y}_t^f = \theta_{M2SLS}\overline{\widehat{\tilde{P}}_{td}^f} + \ddot{\epsilon}_t^f. \text{ (E14)}$$

This produces an unbiased estimate of θ for the same reason that Wald 2SLS does.

***Efficiency of M2SLS versus Wald 2SLS Estimators with the Same Covariates*[1]**

Dhrymes and Lleras-Muney (2006) develop a test statistic for whether M2SLS is more efficient than Wald 2SLS given the same covariates. Their Theorem 4 defines a test statistic and critical values:

$$\hat{t}_n = \frac{\hat{\tau}}{\hat{\sigma}_\tau}, \text{ (E15)}$$

where all unknown parameters have been replaced by their consistent estimators in:

$$\tau = d_1\sqrt{n}\theta_{M2SLS} + d_2\sqrt{n}\sigma_\varepsilon^2 + d_3\sqrt{G}\sigma_{\bar{\varepsilon}v} - \theta_{M2SLS}^2\sigma_\varepsilon^2, \text{ (E16a)}$$
$$\sigma_\tau = \sqrt{d_1^2\sigma_{\theta_{M2SLS}}^2 + d_2^2\phi_{22} + d_3^2\phi_{12}}. \text{ (E16b)}$$

And:

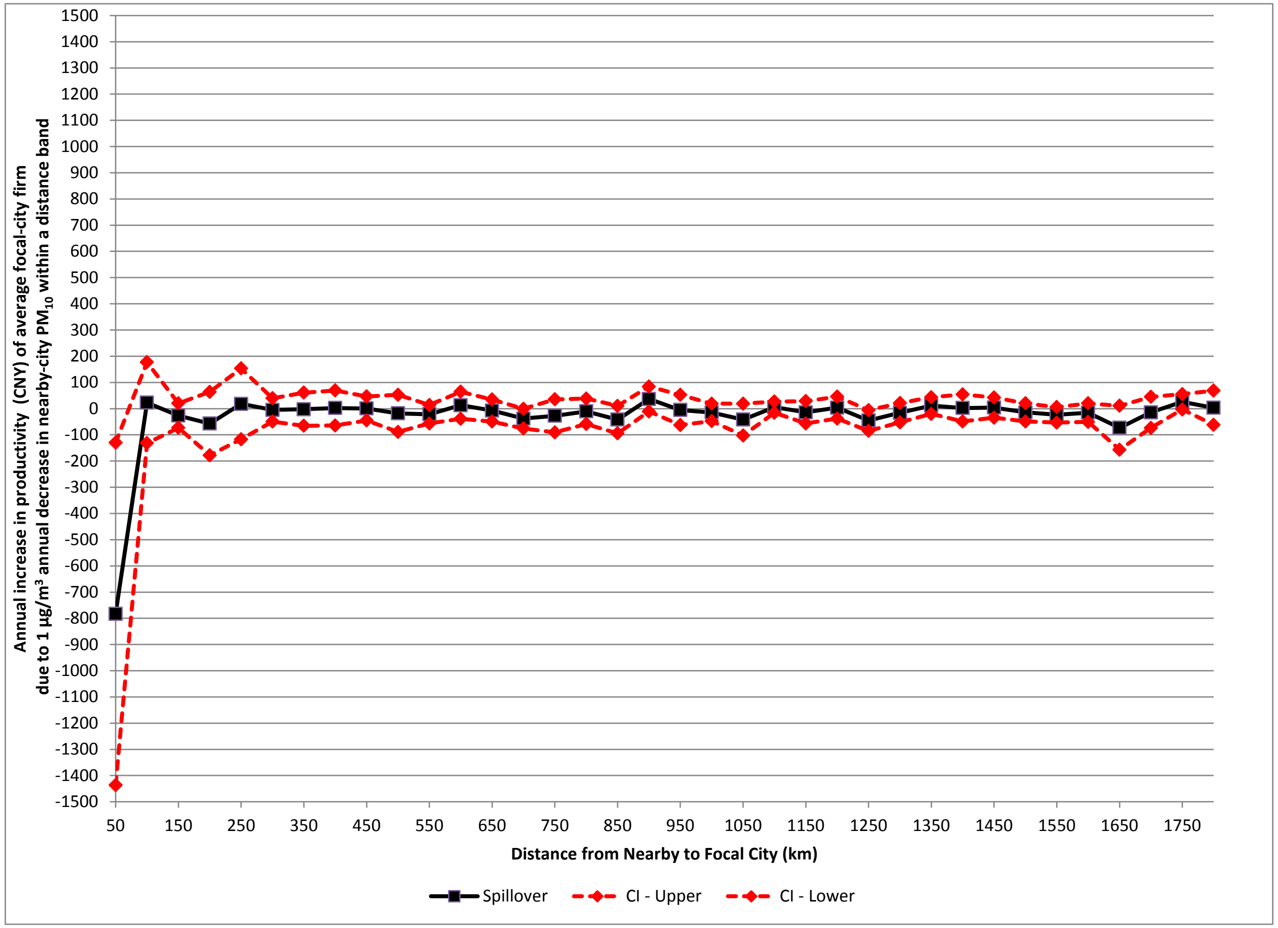
$$d_1 = 2\,(\sigma_{\bar{\varepsilon}v} + \theta_{M2SLS}\sigma_\varepsilon^2), d_2 = \theta_{M2SLS}^2, d_3 = 2\theta_{M2SLS}, \text{ (E17a)}$$
$$\phi_{22} = \mu_4 - (\sigma_\varepsilon^2)^2, \phi_{12} = (\sigma_{\bar{\varepsilon}v})^2 + \sigma_\varepsilon^2\sigma_v^2. \text{ (E17b)}$$

σ_ε^2 is the variance of ε_{td}^f, σ_v^2 is the variance of v_t^f, $\sigma_{\bar{\varepsilon}v}$ is the covariance of v_t^f and the annual averaged values of ε_{td}^f, $\sigma_{\theta_{M2SLS}}^2$ is the variance of the estimate of θ_{M2SLS}, μ_4 is the fourth moment of ε_{td}^f, n is the number of observations in the first stage of M2SLS, and G is the number of observations (groups) in the second stage of M2SLS.

M2SLS is more efficient than Wald 2SLS when $\hat{t}_n \geq 1.64$ with a 5% level of significance and $\hat{t}_n \geq 1.28$ with a 10% level of significance.

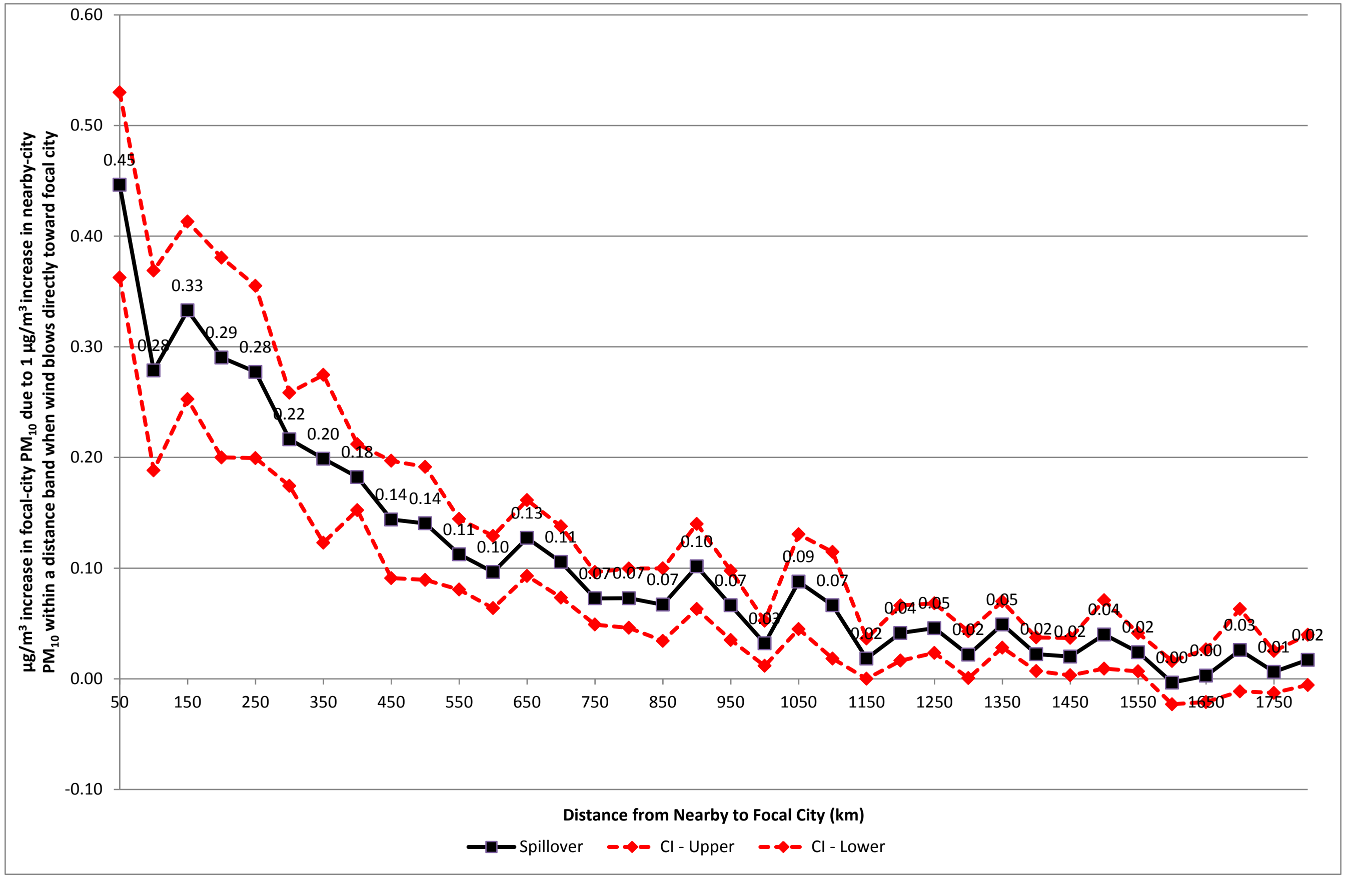
[1] As we describe in the paper, M2SLS might also be more efficient than Wald 2SLS because finer controls can be included in the first stage of M2SLS.

Online Appendix F: Reduced-form estimates of air pollution spillover effects from a one μg/m³ annual decrease in nearby-city PM_{10} within a distance band on average annual labor productivity of focal-city firms as a function of distance (N = 11,559,133)



Solid, black line shows effect of a one μg/m³ annual decrease in nearby-city PM_{10} within a distance band (holding pollution in all other distance bands constant) on average annual productivity of focal-city firms as a function of distance estimated using a reduced-form regression with annual data (conditioning pollution data on wind direction) and controlling for weather variables, firm fixed effects, and region-by-year fixed effects. Estimation allows for piecewise linear effects in increments of 50 kilometers. Dashed, red lines show 95% confidence intervals based on standard errors clustered at the focal-city level.

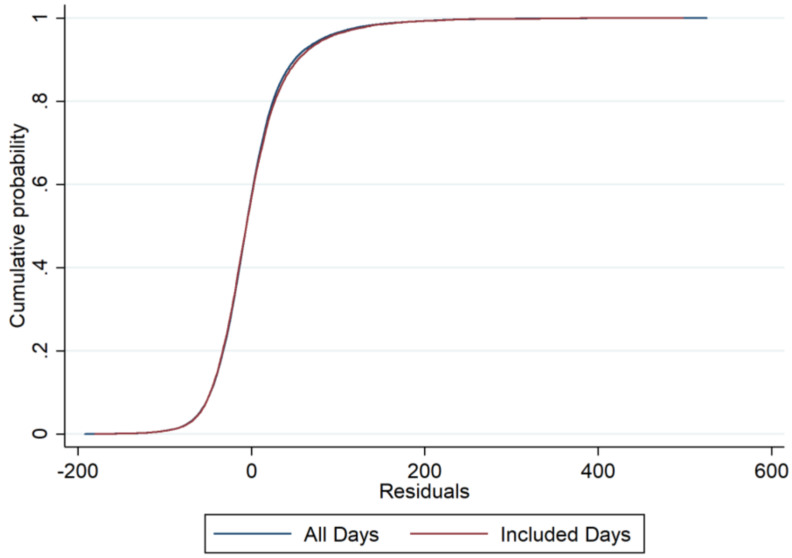
Online Appendix G: Coefficients and confidence intervals for estimate of spillover decay function (effect of one µg/m³ increase in nearby-city PM_{10} within a distance band on focal-city PM_{10} when wind blows directly toward the focal city) as a function of distance (N = 988,230)



Solid, black line shows effect of a one µg/m³ increase in nearby-city PM_{10} within a distance band (holding pollution in all other distance bands constant) on focal-city PM_{10} when the wind is blowing directly toward the focal city as a function of distance controlling for weather variables, focal-city fixed effects, and region-by-year-by-month fixed effects. Estimation allows for piecewise linear effects in increments of 50 kilometers. Dashed, red lines show 95% confidence intervals estimated using 100 iterations of a block bootstrap by focal city.

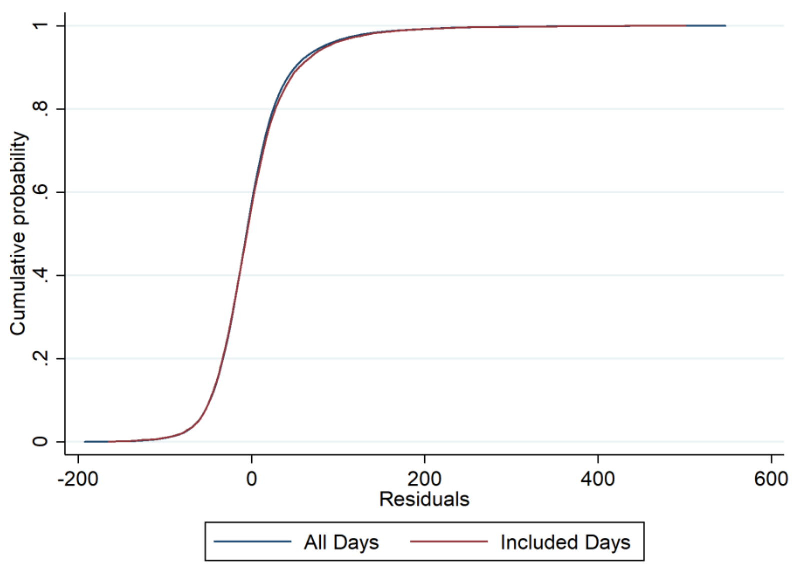
Online Appendix H: Cumulative distribution functions of all days versus included days of nearby-city air pollution conditional on control variables at different maximum distance cutoffs

150-kilometer radius



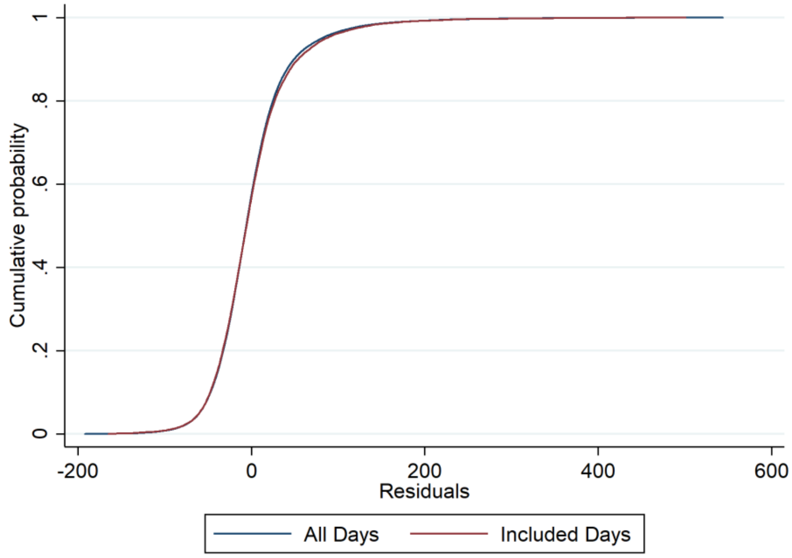
Cumulative distribution functions of residuals from regressing nearby-city PM_{10} on the control variables used in the first stage of the M2SLS procedure (daily weather controls, city fixed effects, and region-by-year-by-month fixed effects) separately for all days and included days using a 150-kilometer cutoff in choosing nearby cities.

200-kilometer radius



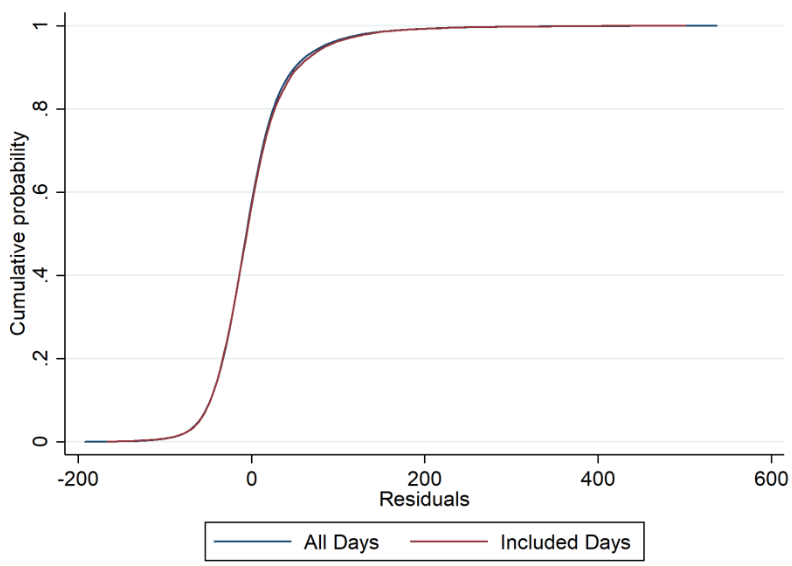
Cumulative distribution functions of residuals from regressing nearby-city PM_{10} on the control variables used in the first stage of the M2SLS procedure (daily weather controls, city fixed effects, and region-by-year-by-month fixed effects) separately for all days and included days using a 200-kilometer cutoff in choosing nearby cities.

250-kilometer radius



Cumulative distribution functions of residuals from regressing nearby-city PM_{10} on the control variables used in the first stage of the M2SLS procedure (daily weather controls, city fixed effects, and region-by-year-by-month fixed effects) separately for all days and included days using a 250-kilometer cutoff in choosing nearby cities.

300-kilometer radius



Cumulative distribution functions of residuals from regressing nearby-city PM_{10} on the control variables used in the first stage of the M2SLS procedure (daily weather controls, city fixed effects, and region-by-year-by-month fixed effects) separately for all days and included days using a 300-kilometer cutoff in choosing nearby cities.

Online Appendix I: Summary statistics for M2SLS estimation 2001 to 2007 (200- and 250-kilometer distances)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
	200 kilometers proximity				**250 kilometers proximity**			
First-stage sample (firm-day)								
	(N = 17,858,505)				**(N = 18,758,702)**			
Focal city PM_{10} ($\mu g/m^3$)	111.1	68.5	10.0	600.0	110.2	67.9	10.0	600.0
Nearby city PM_{10} ($\mu g/m^3$)	97.4	64.2	11.0	600.0	96.9	63.3	11.0	600.0
Distance between focal/nearby city (km)	96.0	35.3	44.0	193.0	102.0	43.5	44.0	234.7
# of city-years	135				149			
# of focal cities	40				44			
Second-stage sample (firm-year)								
	(N = 264,746)				**(N = 276,528)**			
Value added (CNY1,000)	15,317.6	27,318.0	105.7	366,425.6	15,349.9	27,353.5	105.7	366,425.6
Total workers	168.7	247.9	10.0	3,012.0	169.2	248.8	10.0	3,012.0
Value added per worker (CNY1,000)	120.3	221.4	0.5	16,247.6	120.6	222.5	0.5	16,247.6
# of firms	82,714				86,941			

Summary statistics for data used in M2SLS estimation of causal effect of local air pollution on local firms' labor productivity. First-stage data is conditional on wind blowing toward the focal city.

Online Appendix J: Heterogeneity – M2SLS estimates of causal effect of local PM_{10} on local labor productivity using as an instrument pollution of nearest-nearby city within 300 kilometers

	(1)	(2)	(3)	(4)	(5)	(6)
	Middle funnel, 300-kilometer maximum distance cutoff					
Panel A: First stage:	**Northeast**	**Non-Northeast**	**"Dirty" Firms**	**"Clean" Firms**	**High-Technology**	**Low-Technology**
Dependent variable:	**Focal city PM_{10}**					
Nearby city PM_{10}	0.6481***	0.6972***	0.6857***	0.7008***	0.7010***	0.6919***
	(0.0376)	(0.0718)	(0.0614)	(0.0695)	(0.0743)	(0.0620)
Fraction of days wind toward focal city	0.296	0.242	0.253	0.243	0.244	0.248
KP *F*-statistic	297.3	94.3	124.6	101.6	89.0	124.5
# cities	7	40	47	47	47	47
Sample size	1,917,981	17,421,936	6,278,115	13,061,802	7,743,619	11,4596,298
Panel B: second stage (firm-year sample)						
Dependent variable:	**ln(value added/worker)**					
Mean annual predicted focal city PM_{10}	-0.0143	-0.0026*	-0.0041**	-0.0024**	-0.0024*	-0.0033**
	(0.0179)	(0.0014)	(0.0017)	0.0013	0.0013	0.0013
Implied elasticity	-1.470	-0.271	-0.429	-0.249	-0.256	-0.338
# firms	8,468	80,248	28,292	60,424	35,545	53,171
Sample size	23,963	267,376	92,747	198,592	119,407	171,932

Columns 1 and 2 divide firms into those located in the Northeast region (defined in Appendix D) and outside. Columns 3 and 4 divide firms into "clean" and "dirty" based on the 3-digit SIC codes in Mani and Wheeler (1998). Columns 5 and 6 divide firms into high- and low-technology based on OECD (2011). All columns use the middle funnel in choosing days when wind blows toward focal city and 300-kilometer radius and exclude days when API is between 95 and 105. First-stage models include firm and region-by-year-by-month fixed effects; linear and quadratic terms of daily humidity and wind speed; and categorial variables for temperature bins as described in the text aggregated to the corresponding level. Second-stage models include firm and region-by-year fixed effects; annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorial variables for temperature (i.e., number of days in each temperature bin). Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Second-stage standard errors are also adjusted for two-stage estimation using 100 block-bootstrap iterations.

Online Appendix K: Counterfactual and robustness checks – M2SLS estimates of causal effect of local PM_{10} on local labor productivity using as an instrument pollution of nearest-nearby city within 300 kilometers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Region-by-Year-by-Month Fixed Effects							**"Middle Funnel"**	
	Baseline	**Counter-factual**	**"Narrow" Funnel**	**"Broad" Funnel**	**Drop API Below 50 "Middle" Funnel**	**No Weather Controls**	**Include 95 < API < 105**	**Year-by-Month Fixed Effects**	**Region-by-Year Fixed Effects**
Panel A: M2SLS first stage (firm-day sample)									
Dependent variable:	**Daily focal city PM_{10}**								
Daily nearby city PM_{10}	0.6959***	0.7097***	0.6946***	0.6921***	0.6684***	0.7361***	0.6533***	0.7058***	0.7572***
	(0.0669)	(0.0660)	(0.0684)	(0.0629)	(0.0679)	(0.0709)	(0.0741)	(0.0794)	(0.0827)
Fraction of days wind toward focal city	0.246	0.246	0.216	0.273	0.249	0.246	0.247	0.246	0.246
KP *F*-statistic	108.1	115.47	103.0	120.9	97.0	107.8	77.7	79.0	83.9
# cities	47	47	47	47	47	47	47	47	47
Sample size	19,339,917	20,366,530	16,998,909	21,464,701	15,015,417	19,339,917	22,337,870	19,339,917	19,339,917
Panel B: M2SLS second stage (firm-year sample)									
Dependent variable:	**ln(value added/worker)**								
Mean annual predicted focal city PM_{10}	-0.0030**	-0.0008	-0.0025	-0.0038**	-0.0023*	-0.0018	-0.0030**	-0.0034***	-0.0021*
	(0.0014)	(0.0016)	(0.0018)	(0.0018)	(0.0012)	(0.0015)	(0.0015)	(0.0008)	(0.0012)
Implied elasticity	-0.312		-0.260	-0.396	-0.239	-0.187	-0.326	-0.354	-0.219
# firms	88,716	88,716	88,716	88,716	88,716	88,716	88,716	88,716	88,716
Sample size	291,339	291,339	291,339	291,339	291,339	291,339	291,339	291,339	291,339

All columns use a 300-kilometer maximum radius when choosing nearest nearby city. Columns 1, 2 and 5 through 9 apply the middle funnel in choosing days when wind blows toward focal city; Column 3 uses the narrow funnel; and Column 4 the broad funnel. All columns include firm fixed effects and linear and quadratic terms of daily humidity and wind speed; and categorial variables for temperature bins as described in the text in the first stage. Columns 1 through 7 also include region-by-year-by-month fixed effects; Column 8 also includes year-by-month fixed effects; and Column 9 also includes region-by-year fixed effects. Second stage models include firm and region-by-year fixed effects in Columns 1 through 7 and 9 and firm and year fixed effects in Column 8. First-stage model in Column 2 conditions on wind blowing away from the focal city while all other columns condition on wind blowing toward the focal city. All second-stage models include annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorical variables for temperature (i.e., number of days in each temperature bin). Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in Panel B are also adjusted for two-stage estimation using 100 block-bootstrap iterations.

Online Appendix L: Effect of daily autocorrelation in output

Problem

If there is daily autocorrelation in output, the true model for focal- ($k = f$) and nearby-city ($k = n$) output on day d of year t is:

$$Y_{td}^{k} = \alpha Y_{td-1}^{k} + \theta P_{td}^{k} + \epsilon_{td}^{k}. \text{ (L1)}$$

Specify the pollution production function as:

$$P_{td}^{k} = \mu Y_{td}^{k} + \eta_{td}^{k}, \text{ (L2)}$$

where η_{td}^{k} reflects daily unobservables affecting pollution production assumed to be independent of all other variables and identically distributed within and across cities. The pollution transport equation (depending on wind direction) is:

$$P_{td}^{k} = \rho P_{td}^{l} + \varepsilon_{td}^{k}, k \neq l. \text{ (L3)}$$

This creates a problem for annual Wald 2SLS and M2SLS because nearby-city pollution is correlated directly with focal-city output rather than only through focal-city pollution.[2] This is the case regardless of which direction the wind blows on the previous day.

Current Day

The first-stage model for annual Wald 2SLS or M2SLS conditions on wind blowing toward the focal city. Suppose that occurs on day d so that Equation (L3) is:

$$P_{td}^{f} = \rho P_{td}^{n} + \varepsilon_{td}^{f}. \text{ (L4)}$$

Now by Equation (L2):

$$P_{td}^{n} = \mu Y_{td}^{n} + \eta_{td}^{n}, \text{ (L5)}$$

and by Equation (L1):

$$Y_{td}^{n} = \alpha Y_{td-1}^{n} + \theta P_{td}^{n} + \epsilon_{td}^{n} = \alpha^{2} Y_{td-2}^{n} + \alpha\theta P_{td-1}^{n} + \theta P_{td}^{n} + \alpha\epsilon_{td-1}^{n} + \epsilon_{td}^{n}. \text{ (L6)}$$

Wind Blows toward Nearby City on Previous Day

By Equation (L3):

$$P_{td-1}^{n} = \rho P_{td-1}^{f} + \varepsilon_{td-1}^{n}. \text{ (L7)}$$

And by Equation (L2):

$$P_{td-1}^{f} = \mu Y_{td-1}^{f} + \eta_{td-1}^{f}. \text{ (L8)}$$

Using Equations (L5) through (L8) this means:

$$P_{td}^{n} \propto \alpha\mu^{2}\theta\rho Y_{td-1}^{f}. \text{ (L9)}$$

[2] Other papers that use daily wind direction as an instrument have this problem if there is autocorrelation in the endogenous variable of interest.

Finally, using Equation (L1):

$$P_{td}^{n} \propto \mu^{2}\theta\rho Y_{td}^{f}. \text{ (L10)}$$

Wind Blows toward Focal City on Previous Day

By Equation (L3):

$$P_{td-1}^{f} = \rho P_{td-1}^{n} + \varepsilon_{td-1}^{f}. \text{ (L11)}$$

Now by Equation (L1):

$$Y_{td-1}^{f} = \alpha Y_{td-2}^{f} + \theta P_{td-1}^{f} + \epsilon_{td-1}^{f}. \text{ (L12)}$$

Using Equations (L5), (L6), (L11), and (L12):

$$P_{td}^{n} \propto \alpha\mu\, Y_{td-1}^{f}/\rho. \text{ (L13)}$$

Finally, using Equation (L1):

$$P_{td}^{n} \propto \mu\, Y_{td}^{f}/\rho. \text{ (L14)}$$

Therefore, under either annual Wald 2SLS and M2SLS without controlling for lagged focal-city pollution in the first stage, nearby-city pollution is correlated with focal-city output directly rather than only through focal-city pollution. Thus, the exclusion restriction fails under annual Wald 2SLS and M2SLS.

Solution

Ideally we would include lagged, focal-city output as a control variable in both stages of 2SLS since it is pre-determined (exogenous) with respect to current output. Conditioning on this would eliminate the correlation between current focal- and nearby-city output and would restore the exogeneity of nearby-city pollution as an instrument. Controlling for lagged output (its annual average) is actually unnecessary in the second stage because there is negligible autocorrelation in annual measures.[3] However, lagged focal-city output is unavailable since the data are annual. We instead include lagged, focal-city pollution as a control variable. Lagged, focal-city pollution is correlated with lagged focal-city output via the pollution production function (Equation (L2)) and is pre-determined (exogenous) with respect to current output. As with using lagged focal-city output, it is unnecessary to include it as a control in the second stage because autocorrelation in annual pollution measures is negligible.

Formally, we change the first-stage, equation to include lagged, focal-city pollution:

$$P_{td}^{f} = \tilde{\rho}\, P_{td}^{n} + \tilde{\gamma}\, P_{td-1}^{f} + \tilde{\varepsilon}_{td}^{f}. \text{ (L15)}$$

Substituting a lagged version of Equation (L2):

$$P_{td-1}^{f} = \mu Y_{td-1}^{f} + \eta_{td-1}^{f}, \text{ (L2')}$$

into Equation (L15) the first-stage equation now controls for lagged, focal-city output:

[3] The only influence of the daily autocorrelation is the difference between the first day of last year and the first day of the current year divided by the number of days in a year.

$$P_{td}^{f} = \ddot{\rho} P_{td}^{n} + \gamma Y_{td-1}^{f} + \iota_{td}^{f}. \text{ (L16)}$$

where $\gamma = \ddot{\tilde{\gamma}}\, \mu$ and $\iota_{td}^{f} = \ddot{\tilde{\gamma}}\, \eta_{td-1}^{f} + \ddot{\tilde{\varepsilon}}_{td}^{f}$.

After converting to average annual values, the fitted values from Equation (L16) are proper instrumented values. They are correlated with focal-city pollution (as before) so that the inclusion restriction is met. They are also exogenous with respect to current focal-city output so that the exclusion restriction is met. Thus, this is a proper instrument under annual Wald 2SLS (which would use the annual averages conditional on wind direction) or M2SLS.

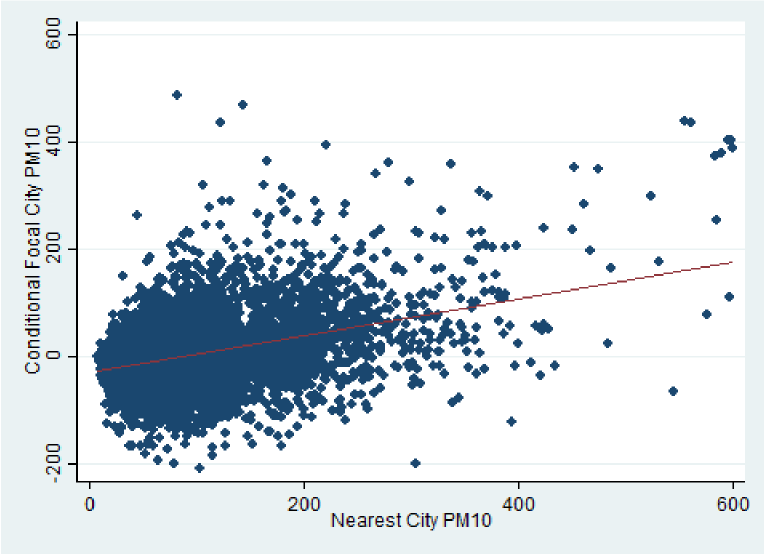
Online Appendix M: Counterfactual estimates – M2SLS estimates of causal effect of local PM_{10} on local labor productivity using as an instrument pollution of nearest nearby city at distances beyond 300 kilometers

	(1)	(2)	(3)	(4)	(5)
	Distance cutoffs				
	Baseline < 300 km	**300 - 350 km**	**300 - 400 km**	**300 - 450 km**	**300 - 500 km**
Panel A: M2SLS first stage (firm-day sample)					
Dependent variable:	**Daily focal city PM_{10}**				
Daily nearby city PM_{10}	0.6959***	0.4067***	0.3665***	0.3952***	0.3935***
	(0.0669)	(0.0805)	(0.0703)	(0.0571)	(0.0541)
Average distance focal to nearby cities	106.5	323.9	338.6	353.9	357.8
Fraction of days wind toward focal city	0.246	0.210	0.228	0.200	0.201
KP *F*-statistic	108.1	22.9	27.2	47.9	52.9
# cities	47	31	40	45	49
Sample size	19,339,917	7,494,962	11,184,946	14,358,319	14,862,542
Panel B: M2SLS second stage (firm-year sample)					
Dependent variable:	**ln(value added/worker)**				
Mean annual predicted focal city PM_{10}	-0.0030**	-0.0001	0.0001	0.0032	0.0034
	(0.0014)	(0.0035)	(0.0021)	(0.0024)	(0.0022)
# firms	88,716	47,564	60,945	81,302	83,921
Sample size	291,339	142,774	192,403	272,753	281,037

Column 1 uses as an instrument the nearest nearby city within 300 kilometers and the middle funnel. Column 2 uses as an instrument the nearest nearby city beyond 300 kilometers but below 350 kilometers and the middle funnel. Columns 3 through 5 increase the maximum distances to 400, 450, and 500 kilometers respectively. First-stage models include firm and region-by-year-by-month fixed effects; linear and quadratic terms of daily humidity and wind speed; and categorial variables for temperature bins as described in the text. The second-stage models include firm and region-by-year fixed effects; annual averages of linear and quadratic terms of daily humidity and windspeed; and annual counts of the daily categorial variables for temperature (i.e., number of days in each temperature bin). Standard errors are clustered at the focal-city level in all models and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors in Panel B are also adjusted for two-stage estimating using 100 block-bootstrap iterations.

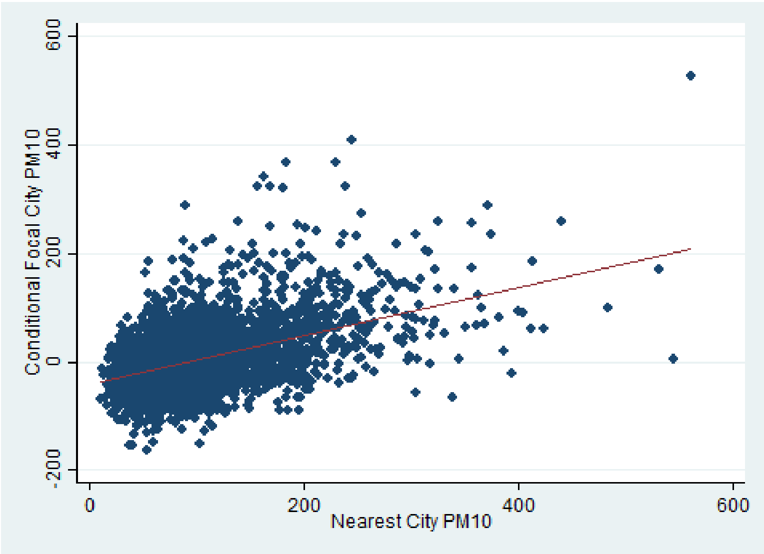
Online Appendix N: Relationship between focal-city pollution (conditional on first-stage control variables) and nearby city pollution using the middle funnel and 300-kilometer radius – at different levels of aggregation

Daily data



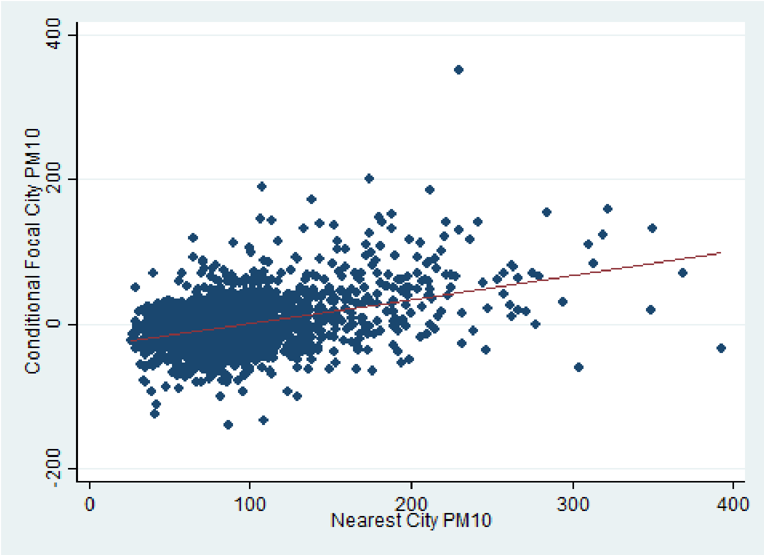
Scatter plot of daily values of focal-city pollution (conditioning on weather variables and region-by-year-by month fixed effects) versus nearby-city pollution using the middle funnel and a 300-kilometer radius. Data includes all cities, years, and days used in the first-stage of M2SLS estimation conditional on wind blowing toward the focal city on that day. The line of best fit shown is $P_t^f\,residuals = \underset{(0.79)}{-30.2} + \underset{(0.0076)}{0.34} P_t^{N^*}$, $N = 12{,}534$.

Weekly averages



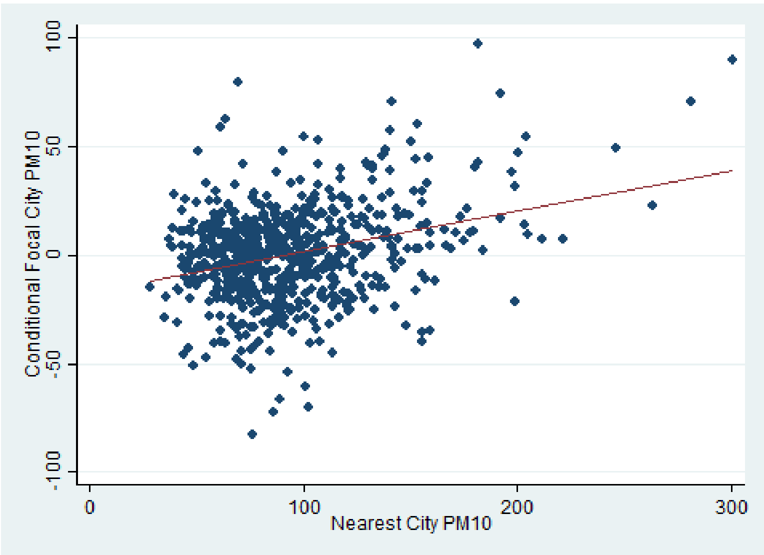
Scatter plot of weekly averages of daily values of focal-city pollution (conditioning on weather variables and region-by-year fixed effects) versus nearby-city pollution using the middle funnel and a 300-kilometer radius. Data includes all cities, years, and days used in the first-stage of Wald 2SLS estimation and conditional on wind blowing toward the focal city before taking weekly averages. The line of best fit shown is $P_t^f\,residuals = \underset{(1.19)}{-40.42} + \underset{(0.011)}{0.442} P_t^{N^*}$, $N = 5{,}546$.

Monthly averages



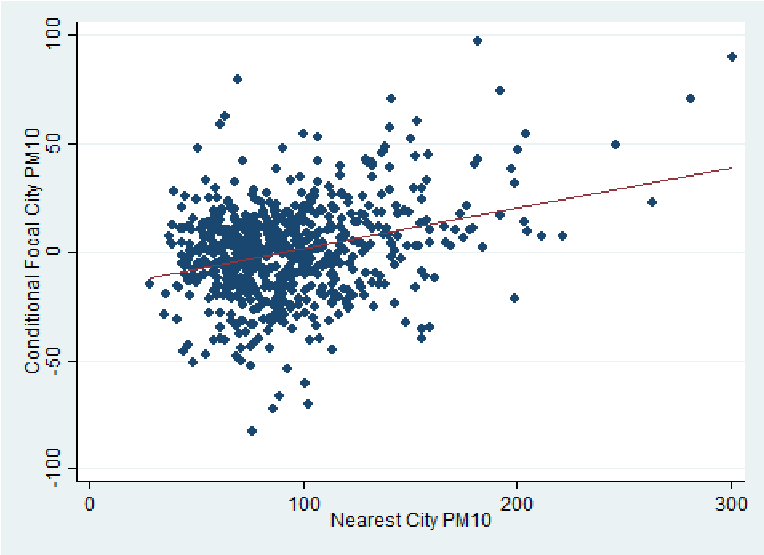
Scatter plot of monthly averages of daily values of focal-city pollution (conditioning on weather variables and region-by-year fixed effects) versus nearby-city pollution using the middle funnel and a 300-kilometer radius. Data includes all cities, years, and days used in the first-stage of Wald 2SLS estimation and conditional on wind blowing toward the focal city before taking monthly averages. The line of best fit shown is $P_t^f\, residuals = \underset{(1.85)}{-30.94} + \underset{(0.018)}{0.330} P_t^{N^*}$, $N = 1{,}824$.

Quarterly averages



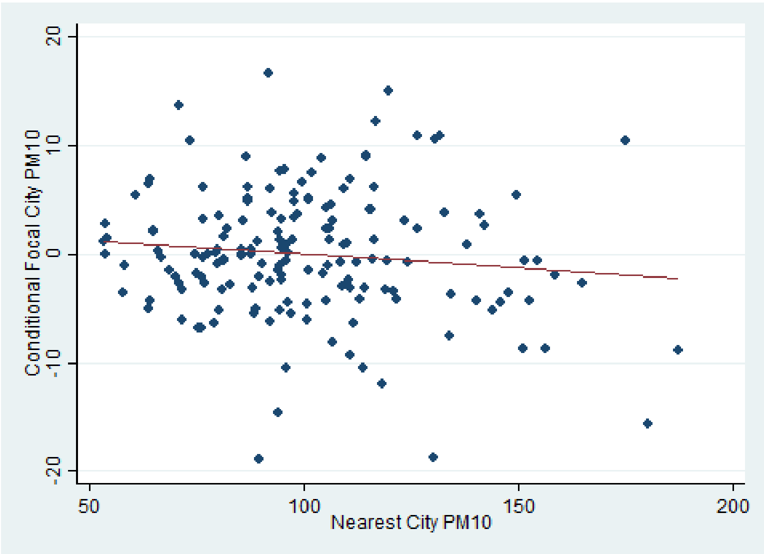
Scatter plot of quarterly averages of daily values of focal-city pollution (conditioning on weather variables and region-by-year fixed effects) versus nearby-city pollution using the middle funnel and a 300-kilometer radius. Data includes all cities, years, and days used in the first-stage of Wald 2SLS estimation and conditional on wind blowing toward the focal city before taking quarterly averages. The line of best fit shown is $P_t^f\, residuals = \underset{(2.26)}{-17.48} + \underset{(0.023)}{0.188} P_t^{N^*}$, $N = 659$.

Semi-annual averages



Scatter plot of semi-annual averages of daily values of focal-city pollution (conditioning on weather variables and region-by-year fixed effects) versus nearby-city pollution using the middle funnel and a 300-kilometer radius. Data includes all cities, years, and days used in the first-stage of Wald 2SLS estimation and conditional on wind blowing toward the focal city before taking semi-annual averages. The line of best fit shown is $P_t^f\,residuals = \underset{(2.53)}{-7.41} + \underset{(0.027)}{0.082} P_t^{N^*}$, $N = 332$.

Annual averages



Scatter plot of annual averages of daily values of focal-city pollution (conditioning on weather variables and region-by-year fixed effects) versus nearby-city pollution using the middle funnel and a 300-kilometer radius. Data includes all cities, years, and days used in the first-stage of Wald 2SLS estimation and conditional on wind blowing toward the focal city before taking annual averages. The line of best fit shown is $P_t^f\,residuals = \underset{(1.78)}{2.49} - \underset{(0.017)}{0.025} P_t^{N^*}$, $N = 166$.