HKUST IEMS
INSTITUTE FOR EMERGING MARKET STUDIES

Air Pollution and COVID-19 Transmission in China

Guojun HE, Yuhang PAN and Takanao TANAKA

HKUST IEMS Working Paper No. 2021-80

Updated February 2021

HKUST Institute for Emering Market Studies
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
T: +852 3469 2215 E: iems@ust.hk W: iems.ust.hk

Air Pollution and COIVD-19 Transmission in China

GUOJUN HE*, YUHANG PAN†, AND TAKANAO TANAKA‡

Accurately estimating the effect of air pollution on COVID-19 transmission requires researchers to account for the epidemiological characteristics, deal with endogeneity, and capture the dynamic impact of air pollution. To do so, we propose a new econometric framework by combining the Susceptible-Infectious-Recovered-Deceased model, the Instrument Variable model, and the Flexible-Distributed-Lag model. Using data covering all Chinese cities, we find that a 10-point increase in the Air Quality Index would lead to a 2.80 percentage point increase in the daily COVID-19 growth rate with 2 to 13 days of delay, implying that improving air quality can help slow the COVID-19 spread.

Keyword: COVID-19, Air Pollution, Communicable Diseases

JEL classification: Q56, Q53, I10

* Division of Social Science, Division of Environment and Sustainability, and Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: gjhe@me.com).

† Division of Environment and Sustainability, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: yhyhpan@gmail.com).

‡ Tanaka: Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: takanao_tanaka@hotmail.co.jp).

.

I. Introduction

The rapid spread of COVID-19 has led to a global public health and economic crisis, bringing tens of millions of infections and massive layoffs. To design effective responses to this unprecedented pandemic, the government needs to identify different factors affecting virus transmission. Existing studies have documented that virus transmission and its severity depends on age, gender, and comorbidities of the infected, as well as climatic conditions.[1] However, less is known about how ambient air pollution, which causes severe damage to various health outcomes (e.g., Graff Zivin and Neidell 2013), can affect the transmission of COIVD-19.

Scientists have been speculating that ambient air pollution could affect the spread of infections, as it might increase people's the susceptibility and exposure to the virus. On the one hand, a rich link of literature has documented that air pollution can cause a persistent inflammatory response and impair the respiratory and immune systems (Ciencewicki and Jaspers, 2007; Diamond et al., 2000), making it more challenging for an individual to resist infection. On the other hand, recent studies suggest that aerosols in the air may maintain the viability and transmissibility of the virus (Liu et al., 2020; van Doremalen et al., 2020). Therefore, degraded air quality, dominated by particulate matter or aerosols, may extend the survival of the virus in the air, which increases individuals' exposure to the virus. A thin but growing literature has documented that air pollution is correlated with COVID-19 incidence in the U.S. (Persico and Johnson, 2020; Wu et al., 2020), Germany (Isphording and Pestel, 2020), the Netherlands (Cole et al., 2020), Italy (Ogen, 2020), the U.K. (Travaglio et al., 2021), and China (Zhu et al., 2020).

However, at least two limitations have plagued the existing studies linking ambient air pollution to the COVID-19 outbreak. First, unlike other health outcomes, the transmission of the

[1] See, for example, Guan et al. (2020) and Williamson et al. (2020) for the relationships between age, gender, and comorbidities and COVID-19 infections, and see Carleton et al. (2020) for the relationship between climate and the COVID-19 transmission.

virus grows exponentially and the changes in the number of daily new cases will depend on the size of the susceptible and infected population. As will be discussed in the Empirical Strategy section, failure to account for such non-linearity in the number of new cases, which is common in the existing literature, will lead to mis-specified econometric models and generate potentially biased estimates.

Second, isolating the impact of air pollution from potential confounders is challenging. It is well documented in the economics literature that air pollution levels are endogenous.[2] Such concerns are exacerbated in the case of COVID-19, because economic activities (e.g., opening industries and schools), health interventions (e.g., social distancing and business closure), and human mobility (e.g., use of transportation), not only changed the transmission of COVID-19, but also affected the air pollution levels (Almond, Du, and Zhang 2020; He, Pan, and Tanaka 2020). Existing evidence on the relationship between air pollution and COVID-19 relies mostly on associational approaches to quantify the impact and does not explicitly address the endogeneity of air pollution. Hence, it remains unclear to policymakers, healthcare professionals, and researchers whether air pollution can causally affect the transmission of COVID-19.

In this study, we propose a new empirical framework to tackle these empirical challenges and estimate the plausibly causal impact of air pollution on COVID transmission. Specifically, we derive our econometric specification from the Susceptible-Infectious-Recovered-Deceased (SIRD) model, which is widely used by epidemiologists to characterize the transmission of infectious disease. Based on this model, we show that the use of daily confirmed cases (or deaths) as the outcome variable in the regressions, as commonly employed by previous studies, could be problematic. Instead, the daily growth rate of the confirmed active infections should be used as

[2] See, for example, Arceo, Hanna, and Oliva (2016); Chay and Greenstone (2003); Cheung, He, and Pan (2020); Deryugina et al. (2019); Deschênes, Greenstone, and Shapiro (2017); He, Liu, and Zhou (2020); Heft-Neal et al. (2020); Knittel, Miller, and Sanders (2016).

the outcome variable, which allows us to account for the exponential epidemic growth. Then, to deal with the endogeneity issue, we use thermal inversions as the instrumental variable. Thermal inversion is a natural phenomenon that involves changes in the normal tendency of the air to cool down with altitude and has been used as the instrumental variable for air quality in multiple previous studies (e.g., Arceo, Hanna, and Oliva 2016; He, Liu, and Salvo 2019). When a thermal inversion occurs, a layer of warmer air overlays a layer of cooler air in the atmosphere. Because warmer air has a lower density, the air pollutants emitted from the ground surface are "trapped," which eventually leads to higher levels of local air pollution. Arguably, thermal inversions induce exogenous changes in air pollution concentrations, which allows us the isolate the impacts of air pollution on COVID-19 transmission from confounders. Further, to account for the potential lags between infections and case confirmations due to the period of incubation, testing, and reporting (X. He et al., 2020; Q. Li et al., 2020), we incorporate the Flexible Distributed Lag (FDL) models into our statistical framework, which uses finite-order B-spline functions to approximate the delayed effects (e.g., Barwick et al. 2018).

Our analyses use data from China, in which we can observe the entire circle of the COVID-19 transmission and has high levels of air pollution. We collect comprehensive data at the day-by-city level from all Chinese prefectural cities and find that air pollution indeed facilitated the COVID-19 transmission: a 10-point (14.3%) increase in the Air Quality Index (AQI) would lead to a 2.80 percentage point increase in the daily COVID-19 growth rate with a delay of 2 to 13 days (0.14 ~ 0.16 increase in the reproduction number (R_0)).[3] A back-of-the-envelope calculation suggests that if China were able to bring down the daily AQI to below 100 in every city-by-day during our study period (mean AQI would have been reduced by 13.2%), the total confirmed

[3] Air Quality Index is a comprehensive measure of air pollution. The index is constructed using $PM_{2.5}$, PM_{10}, SO_2, CO, O_3, and NO_2 concentrations. See Data (Section III), Appendix Figure 1, and Appendix Table 1

cases would have been decreased by 25.7% (30,376 to 22,578). In other words, improving environmental quality can be a powerful tool to slow the virus transmission, particularly in high-pollution countries.

Our findings contribute to the literature in two important ways. First and foremost, in terms of methodology, we advance existing studies on air pollution and COVID-19 relationship by proposing a new econometric framework to estimate the causal impact. We show that estimating this effect requires researchers to account for the exponentiality in virus growth, the endogeneity of air pollution, as well as the dynamic impact. Failure to account for any of the above features will result in potentially biased estimates. Our empirical framework can be easily applied to estimate the impacts of air pollution on other transmittable diseases, as well as to test the COVID-pollution relationship using data from other countries.

Besides, our work also adds to the growing literature on the effects of air pollution on various health outcomes in China. Existing studies have investigated how air pollution affects Chinese people's life-expectancy (e.g., Ebenstein et al. 2017), mortality (e.g., He, Fan, and Zhou 2016; He, Liu, and Zhou 2020), morbidity (e.g., Barwick et al. 2018), cognition (e.g., Graff Zivin et al. 2020; Zhang, Chen, and Zhang 2018), labor productivity (e.g., He, Liu, and Salvo 2019), and defensive expenditure (e.g., Ito and Zhang 2020). Our findings imply that the true costs of air pollution are greater and the government should continue its efforts to fight against pollution.

II. Scientific Background

Ambient air pollution can affect the spread of infections through increasing individuals' susceptibility and exposure to the virus. First, short-term exposure to severe air pollution can weaken the immune system's response to the virus and thereby increase the chance of infection (Ciencewicki and Jaspers, 2007; Diamond et al., 2000). In physiological studies, inhaled

particulates and other air pollutants can interact with immune cells within the airways. It is found that pollutants can trigger cellular signaling pathways, leading to multicellular immune responses and perturbation, and eventually cause disease or fail to prevent disease (Ghio et al., 2012). In epidemiological studies, air pollutants such as PMs, NO_2, SO_2, and O_3 are also found to be associated with inflammatory and immune responses (e.g., Tellier et al. 2019; Vawda et al. 2014).

Second, aerosols in air pollutants may affect virus activity and facilitate its transmission. Aerosols are suspensions of solid or liquid particles in the air. These particles are small and have a low settling velocity; therefore, they can remain airborne for prolonged periods. For example, coughing and sneezing can generate a substantial quantity of particles. Because of evaporation, a large number of these particles shrink and then behave as aerosols. Aerosols due to coughing, sneezing, and breathing can result in virus transmission, in which the virus from an infected person can be carried over a fairly long distance, compared to the transmission through large droplets and direct contact. Both laboratory and epidemiological studies have shown that aerosol transmission can be an important mode of transmissions of SARS (Yu et al., 2014), influenza (Clay et al., 2018; Singer et al., 2020), tuberculosis (Escombe et al., 2007), chickenpox (Leclair et al., 1980), and SARS-CoV-2 (i.e., COVID-19).

III. Data

1. COVID-19 data

The COVID-19 data are retrieved from the China National Health Commission (CNHC).[4] The data comprise newly infected, recovered, and death cases from 20 January to 1 April in 2020 in 330 prefectural cities in China (See Figure 1 and Appendix Table 2 for summary statistics). During

[4] China National Health Commission (accessed 1 July 2020); http://en.nhc.gov.cn/index.html

this period, there were 49,982 cases and 2,553 deaths in Wuhan and 30,441 cases and 727 deaths in other cities. By 1 April 2020, COVID-19 was largely controlled in the country and more than 95% of confirmed cases recovered (tested as negative), so we observed the entire cycle of the virus transmission (See Figure 2 and Appendix Figure 2). In our baseline analyses, we exclude Wuhan city due to concerns about the city's data quality (Q. Li et al., 2020).[5]

2. Air quality data

The air quality data are obtained from 1,605 air quality monitoring stations covering all of the prefectural cities in China. These data are collected from China's National Urban Air Quality Real-Time Publishing Platform.[6] We use the Air Quality Index, which is a comprehensive measure of air pollution: the index is constructed using $PM_{2.5}$, PM_{10}, SO_2, CO, O_3, and NO_2 concentrations, with a lower AQI meaning better air quality. In China, the AQI is determined by the maximum concentration of different air pollutants. We describe the relationship between the AQI and each pollutant in Appendix Figure 1 and Appendix Table 1. To create the city-level air quality data, we first calculated the distance from a city's population center to all monitoring stations within the corresponding city. We then aggregated station-level air pollution data to city-level data using inverse distance weights. For this process, stations closer to the population center are given higher weights so that city-level air pollution data can better represent the population of each city. The weights are inversely proportional to square distance.

3. Thermal inversion data

[5] Wuhan was the epicenter of COVID-19 in China. During the first few weeks after the COVID-19 outbreak, the city faced severe medical resource shortages, and many patients could not get immediate diagnosis and treatment. On 17 April, the Chinese government added another 1,290 COVID-19 deaths in Wuhan (without reporting the timing of the fatalities). COVID-19 data outside Wuhan do not suffer from similar problems because there were far fewer COVID-19 cases in these cities, and COVID-19 tests became widely available soon after scientists learned about the situation in Wuhan. However, our results remain robust when we include data from Wuhan.

[6] Data can be assessed through the following website: http://106.37.208.233:20035. A recent study suggests that air quality data's accuracy improved substantially when automatic air pollution monitoring was introduced from 2013 to 2015 (Greenstone et al., 2020)

The thermal inversion data are obtained from the MERRA-2.[7] The data include the temperature in 42 atmospheric layers (110m to 36,000m). Following Arceo, Hanna, and Oliva (2016) and He, Liu, and Salvo (2019), we use the difference in temperature between the first layer (110m) and the second layer (320m) because this is expected to be associated with the air pollution at ground level. The raw data include the information for each 50*60 km grid. We aggregate the grid-level data to the city level using the same methodology as the air pollution data.

4. Weather data

Weather data include temperature, precipitation, and snow depth. These data were obtained from the Global Historical Climatology Network (GHCN) from the U.S. National Oceanic and Atmospheric Administration (NOAA).[8] We collapse these data to the city by day level using the same method as the air quality data.

IV. Empirical Strategy

1. The SIRD Model

Our empirical analyses are based on the Susceptible-Infectious-Recovered-Deceased Model (SIRD Model), in which the individuals are classified into either Susceptible (S_t), Infected (I_t), Recovered (R_t), or Deceased (D_t). The size of each group evolves as follows:

$$S_t' = -\beta_t S_t I_t \tag{1}$$

$$I_t' = (\beta_t S_t - \gamma - \rho) I_t \tag{2}$$

$$R_t' = \gamma I_t \tag{3}$$

$$D_t' = \rho I_t \tag{4}$$

[7] MERRA-2 (National Aeronautics and Space Administration, Goddard Space Flight Center, accessed 1 July 2020); https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/

[8] Global Historical Climatology Network (National Oceanic and Atmospheric Administration, accessed 1 July 2020); https://go.nature.com/2Vy9fez

where β_t is the transmission rate, γ is the recovered rate, and ρ is the death rate. Our empirics aim to recover how air pollution affects the transmission rate (β_t), which is a widely used parameter to measure the spread of the epidemic because it deterministically affects disease development. Here, the reproduction number (R_{0t}) is proportional to the transmission rate ($R_{0t} = \beta_t/(\gamma + \rho)$). From Equation (2), we can write the active epidemic growth as:

$$I_t' = g_t I_t = (\beta_t S_t - \gamma - \rho) I_t =_{S_t \to 1} (\beta_t - \gamma - \rho) I_t \quad (5)$$

In our analyses, we assume the proportion of the susceptible population is close to 1 ($S_t \to 1$), i.e., almost the entire population can be thought of as susceptible.

The solution of equation (5) can be described by the following exponential function:

$$\frac{I_t}{I_{t-1}} = e^{\beta_t - \gamma - \rho} = e^{g_t} \quad (6)$$

where I_t is active infections at time t, and the infection growth rate (g_t) is proportional to the transmission rate (β_t). Taking the natural logarithms of this equation, we get

$$\log(I_{t+1}) - \log(I_t) = g_t = g_0 + \alpha * f(AP_t) + \varepsilon_t \quad (7)$$

$$where, g_t = \beta_t - \gamma - \rho$$

We expect ambient air pollution to alter the COVID-19 growth rate (g_t), through changes to the virus transmission rate (β_t).[9] Thus, we model the growth rate as a function of air pollution exposure $f(AP_t)$ and its average treatment effect (α), in addition to the baseline growth rate (g_0), which measures the growth rate without any exposure to air pollution, and a mean-zero error term (ε_{it}).

We use the growth rate as the outcome variable because it helps us understand how pollution

[9] While we are also interested in investigating the effects on death rates, we do not have sufficient statistical power to investigate whether air pollution affects COVID-19 deaths. This is because there have been only a few COVID-19 deaths in most Chinese cities. Outside Wuhan, more than 90% of Chinese cities have recorded only 0 or 1 death. While more than 3,000 people died from COVID-19 in Wuhan city, the data were not accurate at the early stage of the outbreak in the city. Therefore, we refrain from discussing the relationship between air pollution and the COVID-19 death rate. See Appendix A for the detail.

changes the transmission rate. In contrast, using new confirmed cases (or deaths) as an outcome variable, as commonly employed by previous studies, could produce estimates that are difficult to interpret:

$$new\ confirmed\ cases_t = C_t - C_{t-1} = (\beta_t S_{t-1}) I_{t-1} \quad (8)$$

Or,

$$\log(new\ confirmed\ cases_t) = log(\beta_t S_{t-1}) + \log(I_{t-1}) \quad (9)$$

The number of new confirmed cases depends on both virus transmission rate ($\beta_t S_{t-1}$) and the current number of active infections (I_{t-1}). Therefore, if we model the new cases as a function of air pollution exposure ($\log(new\ confirmed\ cases_t) = \vartheta_0' + \alpha' * f(AP_t) + \varepsilon_t$), the coefficient of air pollution (α') will reflect not only its effect on virus transmission but also on the current level of disease prevalence. Because air pollution is often associated with factors affecting the ongoing disease outbreak (such as the timing of the virus arrival, government interventions, and local economic activities), it is thus likely the estimated coefficient of the pollution variable captures the impact of these confounding factors.

2. The Flexible Distributed Lag Model

To examine the dynamic impacts of air pollution on COVID-19 transmission, we can use a more flexible model that includes lags of the air pollution levels in the regression. Specifically, we use a Flexible Distributed Lag (FDL) model to capture this relationship, which can be described as:

$$g_{it} = \log(I_{it}) - \log(I_{i,t-1}) = \sum\nolimits_k AQI_{i,t-k} * f(\alpha_{t-k}^z) + \sum\nolimits_k X_{i,t-k} * f(\sigma_{t-k}^z) + \theta_i + \pi_t + \varepsilon_{i,t} \quad (10)$$

where g_{it} is the daily growth rate of active infections in city i at day t. $AQI_{i,t-k}$ denotes Air Quality Index (AQI) in city i at day $t-k$, and $X_{i,t-k}$ is a set of control variables in city i at day $t-k$. We include temperature, precipitation, and snow depth as control variables because these

climatic conditions could affect virus transmission (Carleton et al., 2021). We include the AQI and these control variables during the period between $k = 0$, and $k = 21$ (previous 3 weeks). Existing studies suggest that the virus incubation period is often around 5-6 days, with 2 days being the lower bound (X. He et al., 2020; Q. Li et al., 2020). Therefore, we are particularly interested in the effect starting from $k = 2$ to $k = 13$.

α_{t-k} estimates the effect of air pollution in day *t-k* on the virus growth rate in day *t*. To investigate the dynamic impact of air pollution, we use the FDL model, which approximates the set of coefficients α_{t-k} as a cubic B-spline function with z segments, as denoted by $f(\alpha_{t-k}^{z})$. Because daily variation in air pollution is often highly correlated, this smoothing process helps reduce artificial oscillations in the parameter estimates (Barwick et al. 2018).[10]

θ_i and π_t denote city and date fixed effects, which are a set of city-specific and time-specific dummy variables. The inclusion of these sets of fixed effects helps isolate variation in air pollution exposure from time-invariant, time-specific, or seasonal confounders, which could be correlated with virus transmission. Specifically, city fixed effects (θ_i) account for time-invariant confounders specific to each city (e.g., the city's income level, natural endowments, and short-term industrial and economic structure), while time fixed effects (π_t) account for shocks that are common to all cities on a given day (e.g., national virus containment policies, macroeconomic conditions, and the national air pollution time trend). Conditional on these fixed effects and the time-varying control variables, we assess how air pollution affects virus growth rate. If air pollution accelerates virus transmission, α_{t-k} will be positive.

Note that, as is common in any infectious disease literature, the confirmed active cases could be substantially lower than the true cases, and the use of the confirmed cases might not be precise (R. Li et al., 2020). In our study, this concern can be partially alleviated for at least three reasons.

[10] In our primary model, we adopt z = 3 segments. For the robustness check, we also choose z = 2 and z = 4.

First, we use the growth rate of confirmed active infections as our outcome variable, so our findings will not be affected as long as the under-reporting does not change much (for example, 50% of infections are always confirmed) within a city. Second, our regression includes date fixed effects, and this can absorb the national-level testing policies or revision of the disease classification (for example, the Chinese government changed the disease classification on 18 February). Finally, we exclude Wuhan from the baseline analysis. Wuhan was the epi-center in China during the COVID-19 outbreak and most cases were identified in this city. During the early period of the outbreak, data reporting and case confirmation were likely inaccurate in the city, so we conduct our main analyses using data outside Wuhan.

3. The Instrumental Variable Approach

To estimate the causal relationship between air pollution and COVID-19 transmission, we need to isolate the effect of air pollution from other confounding factors that could also affect the virus spread. Our solution is to adopt the Instrumental Variable (IV) approach. We use thermal inversion to instrument air pollution (Arceo et al., 2016; He et al., 2019). Thermal inversion is a natural phenomenon in which a layer of cooler air is overlaid by a layer of warmer air in the atmosphere. When a thermal inversion occurs, the air pollutants emitted from the ground surface will be trapped, which raises the air pollution concentration. We use a 2SLS procedure to estimate the IV model. The first stage in 2SLS can be described as

$$\widehat{AQI}_{i,t-k} = \sum_j TInversion_{i,t-j} * f(\delta^z_{t-j}) + \sum_j X_{i,t-j} * f(\sigma^z_{t-j}) + \tau_i + \varphi_t + e_{i,t} \qquad (11)$$

where $TInversion_{i,t-j}$, denotes the difference in temperature between surface ground and upper layer in city i in time *t-j* ($0 \le j \le 21$), with the larger number representing the warmer temperature in the upper layer. (Therefore, we expect δ^z_{t-j} to be positive.) Because the occurrence of the thermal inversion can be affected by weather conditions, we control for

temperature, precipitation, and depth of snow ($X_{i,t-k}$). We also include city fixed effects (τ_i), and time fixed effects (φ_t).

In the second stage, we regress the daily virus growth rate on this predicted pollution concentration rather than on observed pollution concentration using the following equation:

$$g_t = \sum_k \widehat{AQI}_{i,t-k} * f(\hat{\alpha}^z_{t-k}) + \sum_k X_{i,t-k} * f(\sigma^z_{t-k}) + \theta_i + \pi_t + \varepsilon_{i,t} \quad (12)$$

where $\widehat{AQI}_{i,t-k}$ is predicted air pollution from the first stage. Other variables are analogous to equation (10). $\hat{\alpha}_k$ measures the plausibly causal impact of air quality on virus transmission. In other words, the thermal inversion-driven air pollution has to be uncorrelated with the error term $\varepsilon_{i,t}$, conditional on the set of fixed effects and time-varying control variables, which is not directly testable. We cluster the standard errors at the city level.

4. Placebo Test

To examine the relationship between thermal inversion and air pollution, we regress daily air pollution levels on the occurrence of thermal inversions, conditional on the set of control variables and fixed effects used in equation (10) - (12). We conduct a "placebo" test to rule out the possibility that the strong inversion-pollution relationship is driven by the local trends/seasonality, the spatial distribution of frequency of the thermal inversion, or other factors. Specifically, we randomly shuffle the observed thermal inversions (1) within the same location or (2) within the same day by 1000 times and re-estimate the relationship between pollution and placebo inversions. We plot the distribution of the estimated coefficients and find that their average effects are close to zero.

5. Back-of-the-envelope calculation

We estimate the excess COVID-19 cases attributable to poor air quality. Developing countries, including China, have suffered greatly from poor environmental quality as the cost of rapid

economic growth (Greenstone et al., 2020). In China, AQI lower than 100 is regarded as "moderate" or "good" air quality (See Appendix Table 1) and is recognized as the "blue sky". However, during our study period, 18.6% of the samples (city-by-day) did not meet this standard, and 30% of the cities explains 72.9% of the share with poor air quality. To estimate the excess COVID-19 cases, we first estimate their daily growth rate in two scenarios: observed air quality and the blue sky scenario (AQI is always lower than 100) by calculating the following equations.

$$\hat{g}_{it}^{b} = g_{0it} - \sum\nolimits_{k} (AQI_{i,t-k} - AQI_{i,t-k}^{b}) * f(\hat{\alpha}_{t-k}^{z}) \qquad (A2)$$

where g_{0it} denotes the observed daily growth rate in city *i* at time *t*, and g_{it}^{b} denotes the growth rate if the AQI were always lower than 100. $AQI_{i,t-k}$ is observed AQI and $AQI_{i,t-k}^{b}$ is the hypothetical AQI, in that AQI higher than 100 is replaced with 100. $f(\hat{\alpha}_{t-k}^{z})$ represents the estimated coefficients from our main estimates (equation (12)). As discussed in the main analyses, we only consider impacts between k = 2 and k = 13.

Using the predicted growth rate, the active infections in the blue sky scenario would have been:

$$\widehat{active}_{it}^{b} = \widehat{active}_{i,t-1}^{b} * \exp(\hat{g}_{it}^{b}) \qquad (A3)$$

The predicted number of active infections in city *i* at time *t* builds on that number in the previous period. To predict these numbers, we use the initial active cases in each city when it exceeds 10. We then use the observed removal rate in equations (3) and (4) to estimate the cumulative confirmed cases.

For the removal rate, we use the following equation.

$$removal\ rate_{t}\ (\gamma_{t} + \rho_{t}) = \frac{recovered\ individual + dead\ individual\ at\ time\ t}{infections\ at\ t-1} \qquad (A4)$$

For this projection, we allow the removal rate to vary over time because it should vary depending on the stage of the outbreak. Note that the individuals are classified as recovered when they are

tested as negative in our dataset. Therefore, its definition is slightly different from that in the classical SIRD model.

V. Results

1. First-stage: Thermal inversion and air quality

For the 2SLS procedure to be valid, the first assumption is that the instrument needs to be a strong predictor of the air pollution concentration. Figure 3A shows that thermal inversion is indeed highly correlated with local air pollution levels, even after we include all the time-varying control variables and a set of fixed effects. Specifically, we find that a 1℃ increase in the temperature inversion is associated with a 3.03-point increase in the Air Quality Index (AQI), as shown in Figure 3B. The results also hold when we further control for the city lockdown indicator [11] and the days since the disease outbreak, suggesting thermal inversion is not correlated with city lockdown status or the development of the COVID-19 pandemic. We also include data from Wuhan in the regression and find similar findings. These results are summarized in Appendix Figure 3.

To confirm that this strong relationship is not driven by the local time trends or the spatial distribution of frequency of thermal inversions, we conduct two sets of placebo tests. Specifically, we randomly shuffle the observed thermal inversions (1) within the same location and (2) within the same day 1000 times. We find that the average estimates using the placebo sample in each case are close to zero, as shown in Figure 3B.

[11] We collected local governments' lockdown information city by city from various news media and government announcements. In this paper, we designated a city as locked down when the following three measures were all enforced: (1) prohibition of unnecessary commercial activities in people's daily lives, (2) prohibition of any type of gathering by residents, (3) restrictions on private vehicles and public transportation. Following our definition, 95 out of 324 cities were locked down in our study period. For the detail see (G. He et al., 2020b; Qi et al., 2020) and Appendix Table 3.

2. Second-stage: Dynamic relationship between air quality and COVID-19 growth rate

In the second stage, we use predicted air pollution from the first-stage regression to estimate the pollution-transmission relationship. We include up to 22 days of lags in the regressions (current + previous 21 days) to capture its dynamic impacts.

Our IV estimates are summarized in Figure 4A (Appendix Table 4 for full results). We find positive impacts of air pollution on COVID-19 growth rates with 2 to 13 days of delay. This delay is consistent with epidemiological observations that the disease is usually confirmed after incubation, testing, and reporting (X. He et al., 2020; Q. Li et al., 2020). Specifically, a 10 point (14.3%) increase in AQI during this time window raises the growth rate by 2.80 percentage points. In contrast, outside the time window, air pollution does not have a meaningful effect.

These results are robust to a number of different model specifications. We add the city's lockdown status (Appendix Figure 4A) and days since the outbreak as control variables (4B), include Wuhan in our regression (4C), and use different FDL model segments (4D, 4E). All results remain similar. In addition, we add three days of future air pollution as another placebo test and find that, as expected, its effects are close to zero and statistically insignificant (4F). When we change the lengths of the lags of air pollution from 16 days to 24 days, we also observe similar patterns (Appendix Figure 5).

We observe that the average daily growth rate in the first week of the epidemic outbreak is 24.9% across cities, with the doubling time of infections at 2.78 days. If AQI increases 10 points, the doubling time would be shortened to 2.50 days. Alternatively, assuming the removal rate (total of recovered rate (γ) and death rate (ρ) in our SIRD model) is 17%~20%, which means the patients recover or die in 5.8~7.5 days,[12] a 10-point increase in AQI would lead to 0.14~0.16 higher

[12] In our dataset, the individuals with infections are classified as recovered when they test negative, rather than when they recover from symptoms. Therefore, to estimate the reproduction number from our main results, we adopt the removal rate which is reported by existing studies. They suggest that an average interval is 5.8 days (X. He et al. 2020)

reproduction number ($R_0 = \beta/(\gamma + \rho)$).

Existing studies linking air pollution to various health outcomes often find that ordinary least squares (OLS) estimates understate the impact of air pollution (Appendix Figure 6). We also find that the OLS estimates are substantially smaller than the IV estimates (even though the dynamic patterns are similar). OLS regression shows that a 10-point increase in AQI between 2 days and 13 days before the case is reported is associated with a 0.77 percentage point increase in the growth rate. This is about one-fourth of the IV estimate, suggesting that the OLS estimate can be biased downward. Appendix Figures 5 and 7 provide more robustness checks using different model specifications.

3. Results by different air pollutants

We also provide the results using specific air pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3). Generally, we observe that higher pollution levels increase the growth rate of COVID-19 with 2-13 days of delays (Appendix Figure 8A). For example, a 10% increase in $PM_{2.5}$ (A1), SO_2 (A3), and CO (A5) statistically significantly leads to a 1.4 ppt, 1.1 ppt, and 3.0 ppt rise in the virus growth rate. The only exception is ozone, for which we find a higher concentration of O_3 (A6) slightly decreases the disease transmission, even though the relationship is not statistically significant. This might be because (1) the concentration of ozone is often negatively correlated with other pollutants (Sillman, 1999), and (2) ambient ozone can inactivate the virus by disrupting the virus structure (Tseng and Li, 2006). Moreover, across most pollutants, OLS estimates (Appendix Figure 8B) are substantially smaller than the IV estimates, consistent with our baseline findings.

4. Back-of-the-envelope calculations for the "blue sky" scenarios

We estimate the excess COVID-19 cases attributable to poor air quality. In China, when daily AQI is below 100, it is regarded as "good" or "moderate" air quality (also called a "blue sky" day).

or 7.5 days (Q. Li et al. 2020), making the removal rate 17.4%-19.6%.

During our study period, 18.6% of the city-by-day observations do not meet this standard, and most of the "above-standard" (worse air quality) readings are obtained from northern Chinese cities. Here, we ask what would happen if we were able to bring the air quality index in all Chinese cities to below 100, holding other things constant.

Our estimates imply that the daily virus growth rate would have been slowed by 2.11% on average if all the cities had met the AQI=100 standard (Figure 5A). Consequently, the number of active infections would have been reduced substantially. For instance, the number of active infections would have dropped from 21,855 to 16,714 (23.5%) on 14 February, when we recorded the highest active infection number in China (Figure 5B). Applying our estimates to the observed removal rate[13], we expect that the cumulative confirmed cases would have been reduced by 25.7% (30,376 to 22,578) during our study period (Figure 5C).

The simulation shows that air pollution reshaped the exponential growth of the virus. We observe that the difference in the confirmed cases is small in the initial outbreak but gradually becomes larger as the outbreak progresses. If we do not model the virus transmission process using the SIRD model, we will not be able to capture the dynamic, cumulative, and non-linear impacts of air pollution on COVID-19 cases.

VI. Conclusion

Accurately estimating the effect of air pollution on COVID-19 transmission requires researchers to account for the exponentiality in virus growth and to introduce exogenous shocks to local air quality. Based on the SIRD model, we show that a non-structural framework that directly links the number of new cases to air pollution can be problematic. We also documented

[13] For the back-of-the-envelope calculation, we use the observed removal rate to consistently project the epidemic growth in different policy scenarios.

that associational analysis based on simple OLS regression models will understate the true impact of air pollution on COVID-19 transmission, which is consistent with the previous environmental economics literature.

We find that a 10-point (14.3%) increase in AQI leads to a 2.80 percentage point increase in the daily growth rate over 2 to 13 days after the pollution exposure. The effect is statistically significant and economically meaningful. For example, holding all else constant, if we were able to bring all the air quality levels to the country's standard (AQI = 100, mean AQI would have been reduced by 13.2%), a back-of-envelope calculation reveals that the cumulative cases of COVID-19 would have been reduced by 25.7%

Knowing that air pollution can increase the transmission of COVID-19 is particularly relevant to developing countries that rely heavily on manufacturing and coal (such as India, Indonesia, and Pakistan). These high-pollution countries face significant challenges in controlling COVID-19, in part because of the faster transmission rate caused by high levels of air pollution. Policymakers in these countries should thus consider adopting more stringent pollution control policies in their war to combat COVID-19.

References

Almond, D., Du, X., Zhang, S., 2020. Ambiguous Pollution Response to COVID-19 in China. NBER Working Paper.

Arceo, E., Hanna, R., Oliva, P., 2016. Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City. Econ. J. 126, 257–280. https://doi.org/10.1111/ecoj.12273

Barwick, P.J., Li, S., Rao, D., Zahur, N., 2018. Air Pollution, Health Spending and Willingness to Pay for Clean Air in China. NBER Working Paper. https://doi.org/10.2139/ssrn.2999068

Carleton, T., Cornetet, J., Huybers, P., Meng, K.C., Proctor, J., 2021. Global evidence for ultraviolet

radiation decreasing COVID-19 growth rates. Proc. Natl. Acad. Sci. 118, e2012370118. https://doi.org/10.1073/pnas.2012370118

Chay, K.Y., Greenstone, M., 2003. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. Q. J. Econ. https://doi.org/10.1162/00335530360698513

Cheung, C.W., He, G., Pan, Y., 2020. Mitigating the air pollution effect? The remarkable decline in the pollution-mortality relationship in Hong Kong. J. Environ. Econ. Manage. 101, 102316. https://doi.org/10.1016/j.jeem.2020.102316

Ciencewicki, J., Jaspers, I., 2007. Air pollution and respiratory viral infection. Inhal. Toxicol. https://doi.org/10.1080/08958370701665434

Clay, K., Lewis, J., Severnini, E., 2018. Pollution, infectious disease, and mortality: Evidence from the 1918 Spanish influenza pandemic. J. Econ. Hist. 78, 1179–1209. https://doi.org/10.1017/S002205071800058X

Cole, M.A., Ozgen, C., Strobl, E., 2020. Air Pollution Exposure and Covid-19 in Dutch Municipalities. Environ. Resour. Econ. 76, 581–610. https://doi.org/10.1007/s10640-020-00491-4

Deryugina, T., Heutel, G., Miller, N.H., Molitor, D., Reif, J., 2019. The mortality and medical costs of air pollution: Evidence from changes in wind direction. Am. Econ. Rev. 109, 4178–4219. https://doi.org/10.1257/aer.20180279

Deschênes, O., Greenstone, M., Shapiro, J.S., 2017. Defensive investments and the demand for air quality: Evidence from the NOx budget program. Am. Econ. Rev. 107, 2958–2989. https://doi.org/10.1257/aer.20131002

Diamond, G., Legarda, D., Ryan, L.K., 2000. The innate immune response of the respiratory epithelium. Immunol. Rev. https://doi.org/10.1034/j.1600-065X.2000.917304.x

Ebenstein, A., Fan, M., Greenstone, M., He, G., Zhou, M., 2017. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proc. Natl. Acad. Sci. U. S. A. 114, 10384–10389. https://doi.org/10.1073/pnas.1616784114

Escombe, A.R., Oeser, C., Gilman, R.H., Navincopa, M., Ticona, E., Martínez, C., Caviedes, L., Sheen, P., Gonzalez, A., Noakes, C., Moore, D.A.J., Friedland, J.S., Evans, C.A., 2007. The detection of airborne transmission of tuberculosis from HIV-infected patients, using an in vivo air sampling model. Clin. Infect. Dis. 44, 1349–1357. https://doi.org/10.1086/515397

Fan, M., He, G., Zhou, M., 2020. The winter choke: Coal-Fired heating, air pollution, and mortality in China. J. Health Econ. 71, 102316. https://doi.org/10.1016/j.jhealeco.2020.102316

Ghio, A.J., Carraway, M.S., Madden, M.C., 2012. Composition of air pollution particles and oxidative stress in cells, tissues, and living systems. J. Toxicol. Environ. Heal. - Part B Crit. Rev. https://doi.org/10.1080/10937404.2012.632359

Graff Zivin, J., Liu, T., Song, Y., Tang, Q., Zhang, P., 2020. The unintended impacts of agricultural fires: Human capital in China. J. Dev. Econ. 147, 102560. https://doi.org/10.1016/j.jdeveco.2020.102560

Graff Zivin, J., Neidell, M., 2013. Environment, health, and human capital. J. Econ. Lit. https://doi.org/10.1257/jel.51.3.689

Greenstone, M., He, G., Jia, R., Liu, T., 2020. Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution. Am. Econ. Rev. Insights Forthcoming. https://doi.org/10.2139/ssrn.3638591

Guan, W., Ni, Z., Hu, Yu, Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D.S.C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, Jin-lin, Liang, Z., Peng, Y., Wei, L., Liu, Y., Hu, Ya-hua, Peng, P., Wang, Jian-ming, Liu, J., Chen, Z., Li, G., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S., Zhong, N., 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. N. Engl. J. Med. 382, 1708–1720. https://doi.org/10.1056/nejmoa2002032

He, G., Fan, M., Zhou, M., 2016. The effect of air pollution on mortality in China: Evidence from the 2008 Beijing Olympic Games. J. Environ. Econ. Manage. 79, 18–39. https://doi.org/10.1016/j.jeem.2016.04.004

He, G., Liu, T., Zhou, M., 2020a. Straw burning, PM2.5, and death: Evidence from China. J. Dev. Econ. 145, 102468. https://doi.org/10.1016/j.jdeveco.2020.102468

He, G., Pan, Y., Tanaka, T., 2020b. The short-term impacts of COVID-19 lockdown on urban air pollution in China. Nat. Sustain. 3, 1005–1011. https://doi.org/10.1038/s41893-020-0581-y

He, J., Liu, H., Salvo, A., 2019. Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China. Am. Econ. J. Appl. Econ. 11, 173–201. https://doi.org/10.1257/app.20170286

He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B.J., Li, F., Leung, G.M., 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. 26, 672–675. https://doi.org/10.1038/s41591-020-0869-5

Heft-Neal, S., Burney, J., Bendavid, E., Voss, K.K., Burke, M., 2020. Dust pollution from the Sahara and African infant mortality. Nat. Sustain. 3, 863–871. https://doi.org/10.1038/s41893-020-0562-1

Isphording, I.E., Pestel, N., 2020. Pandemic Meets Pollution: Poor Air Quality Increases Deaths by COVID-19. Work. Pap.

Ito, K., Zhang, S., 2020. Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. J. Polit. Econ. 128, 1627–1672. https://doi.org/10.1086/705554

Knittel, C.R., Miller, D.L., Sanders, N.J., 2016. Caution, drivers! children present: Traffic, pollution, and infant health. Rev. Econ. Stat. 98, 350–366. https://doi.org/10.1162/REST_a_00548

Leclair, J.M., Zaia, J.A., Levin, M.J., Congdon, R.G., Goldmann, D.A., 1980. Airborne Transmission of Chickenpox in a Hospital. N. Engl. J. Med. 302, 450–453. https://doi.org/10.1056/nejm198002213020807

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N. Engl. J. Med. 382, 1199–1207. https://doi.org/10.1056/nejmoa2001316

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science (80-.). 368, 489–493. https://doi.org/10.1126/science.abb3221

Liu, T., He, G., Lau, A., 2018. Avoidance behavior against air pollution: evidence from online search indices for anti-PM2.5 masks and air filters in Chinese cities. Environ. Econ. Policy Stud. 20, 325–363. https://doi.org/10.1007/s10018-017-0196-3

Liu, Yuan, Ning, Z., Chen, Y., Guo, M., Liu, Yingle, Gali, N.K., Sun, L., Duan, Y., Cai, J., Westerdahl, D., Liu, X., Xu, K., Ho, K. fai, Kan, H., Fu, Q., Lan, K., 2020. Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. Nature 582, 557–560. https://doi.org/10.1038/s41586-020-2271-3

Ogen, Y., 2020. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. Sci. Total Environ. 726, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605

Persico, C., Johnson, K.R., 2020. The Effects of Increased Pollution on COVID-19 Cases and Deaths. Work. Pap. https://doi.org/10.2139/ssrn.3633446

Qi, Jinlei, Zhang, Dandan, Zhang, X., Yin, Peng, Liu, J., Pan, Y., Takana, T., Xie, P., Wang, Z., Liu, S., Gao, G.F., He, G., Zhou, Maigeng, Qi, J, Yin, P, Liu MPH, J., Zhou, M, Zhang, D, Zhang, X.B., Xie, P.M., Wang, Z.M., Kong, H., 2020. Do Lockdowns Bring about Additional Mortality Benefits or Costs? Evidence based on Death Records from 300 Million Chinese People. Work. Pap.

Sillman, S., 1999. The relation between ozone, NO(x) and hydrocarbons in urban and polluted rural environments. Atmos. Environ. https://doi.org/10.1016/S1352-2310(98)00345-8

Singer, G., Graff Zivin, J., Neidell, M., Sanders, N., 2020. Air Pollution Increases Influenza Hospitalizations. Work. Pap.

Tellier, R., Li, Y., Cowling, B.J., Tang, J.W., 2019. Recognition of aerosol transmission of infectious agents: a commentary. BMC Infect. Dis. 19, 101. https://doi.org/10.1186/s12879-019-3707-y

Travaglio, M., Yu, Y., Popovic, R., Selley, L., Leal, N.S., Martins, L.M., 2021. Links between air pollution and COVID-19 in England. Environ. Pollut. 268, 115859. https://doi.org/10.1016/j.envpol.2020.115859

Tseng, C.C., Li, C.S., 2006. Ozone for inactivation of aerosolized bacteriophages. Aerosol Sci. Technol. 40, 683–689. https://doi.org/10.1080/02786820600796590

van Doremalen, N., Bushmaker, T., Morris, D.H., Holbrook, M.G., Gamble, A., Williamson, B.N., Tamin, A., Harcourt, J.L., Thornburg, N.J., Gerber, S.I., Lloyd-Smith, J.O., de Wit, E., Munster, V.J., 2020. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. N. Engl. J. Med. 382, 1564–1567. https://doi.org/10.1056/nejmc2004973

Vawda, S., Mansour, R., Takeda, A., Funnell, P., Kerry, S., Mudway, I., Jamaludin, J., Shaheen, S., Griffiths, C., Walton, R., 2014. Associations between inflammatory and immune response genes and adverse respiratory outcomes following exposure to outdoor air pollution: A huge systematic review. Am. J. Epidemiol. https://doi.org/10.1093/aje/kwt269

Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H.I., MacKenna, B., Tomlinson, L., Douglas, I.J., Rentsch, C.T., Mathur, R., Wong, A.Y.S., Grieve, R., Harrison, D., Forbes, H., Schultze, A., Croker, R., Parry, J., Hester, F., Harper, S., Perera, R., Evans, S.J.W., Smeeth, L., Goldacre, B., 2020. Factors associated with COVID-19-related death using OpenSAFELY. Nature 584, 430–436. https://doi.org/10.1038/s41586-020-2521-4

Wu, X., Nethery, R.C., Sabath, M.B., Braun, D., Dominici, F., 2020. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. Sci. Adv. 6, eabd4049. https://doi.org/10.1126/sciadv.abd4049

Yu, I.T.S., Qiu, H., Tse, L.A., Wong, T.W., 2014. Severe acute respiratory syndrome beyond amoy gardens: Completing the incomplete legacy. Clin. Infect. Dis. 58, 683–686. https://doi.org/10.1093/cid/cit797

Zhang, Xin, Chen, X., Zhang, Xiaobo, 2018. The impact of exposure to air pollution on cognitive performance. Proc. Natl. Acad. Sci. U. S. A. 115, 9193–9197. https://doi.org/10.1073/pnas.1809474115

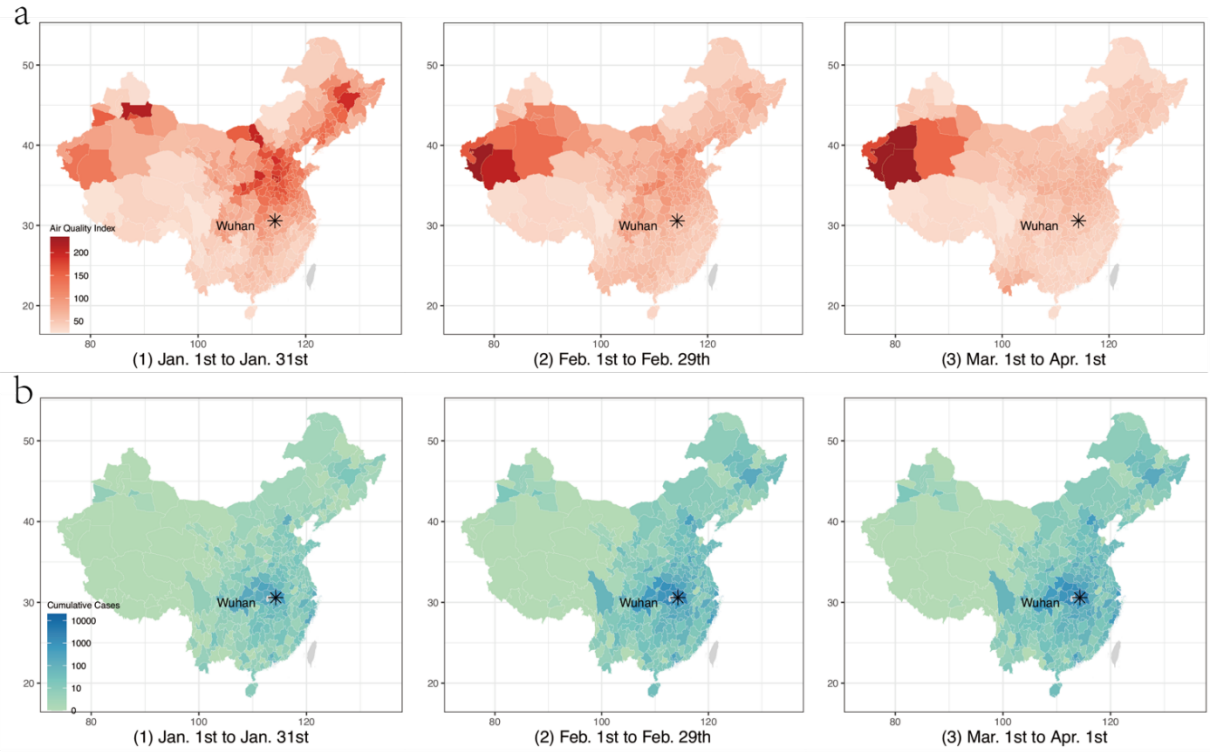
Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Sci. Total Environ. 727, 138704. https://doi.org/10.1016/j.scitotenv.2020.138704

Acknowledgement

We thank Wenbo Gong, Yatang Lin, Ryo Okui, Albert Park, Deyu Rao, and seminar and workshop participants at HKUST, Renmin University, and the 15th Applied Econometrics Conference for their valuable comments and suggestions. This work was supported by the Center for Industrial Development and Environmental Governance (CIDEG) of Tsinghua University, the Anti-epidemic Fund 2.0 from the HK Government, and the Bai Xian Asia Institute.

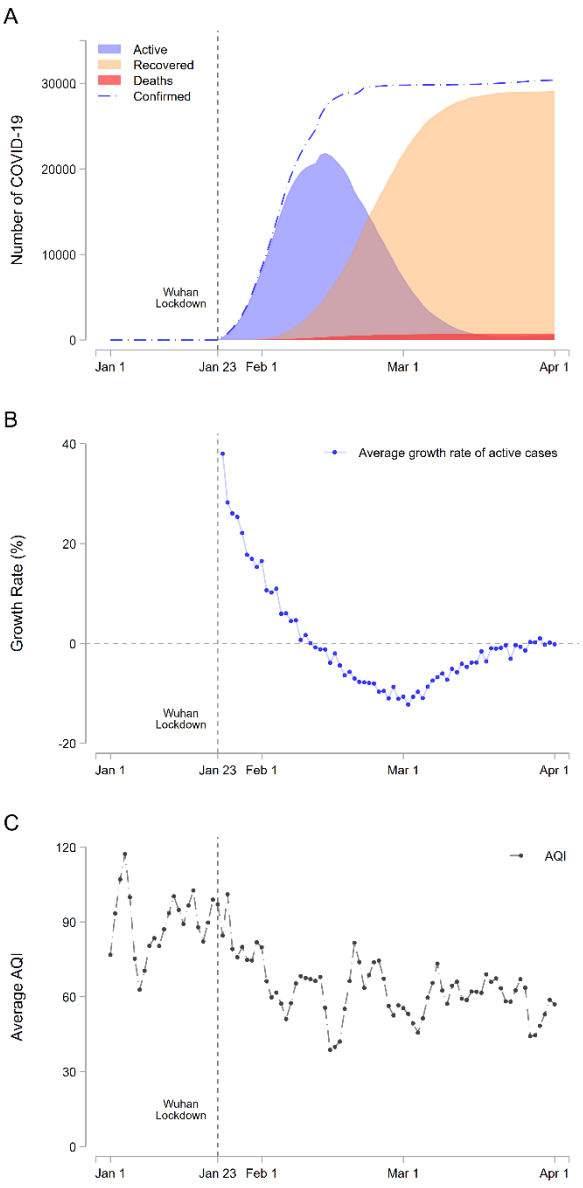
Figures

Figure 1. Data on COVID-19 infections and air pollution in China.



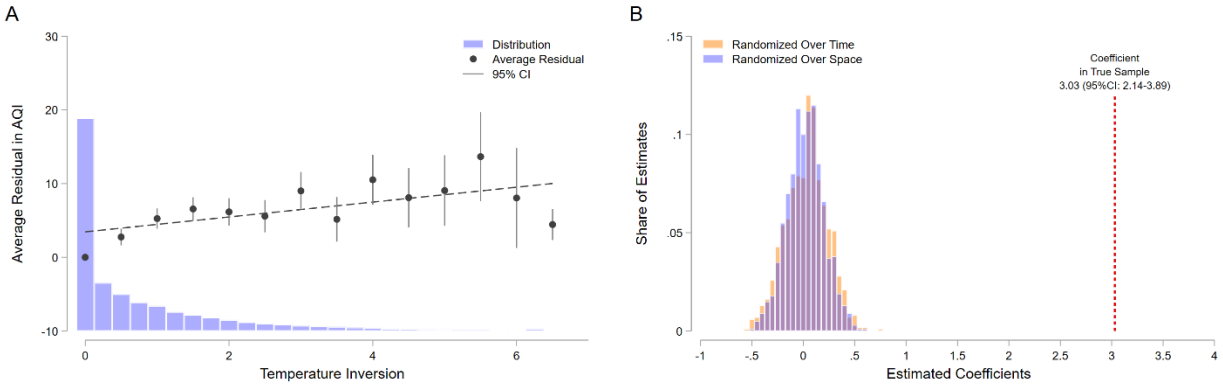
Notes: Panel A shows the trend of the Air Quality Index (AQI). Higher AQI means worse air pollution. AQI is a comprehensive measure of air pollution: the index is constructed using $PM_{2.5}$, PM_{10}, SO_2, CO, O_3, and NO_2 concentrations (See Appendix Table 1). AQI is high in the Gobi Desert (North West part). Panel B shows the distribution of confirmed COVID-19 cases. The gray color denotes the no-data areas.

Figure 2. Data on COVID-19 infections and air pollution.



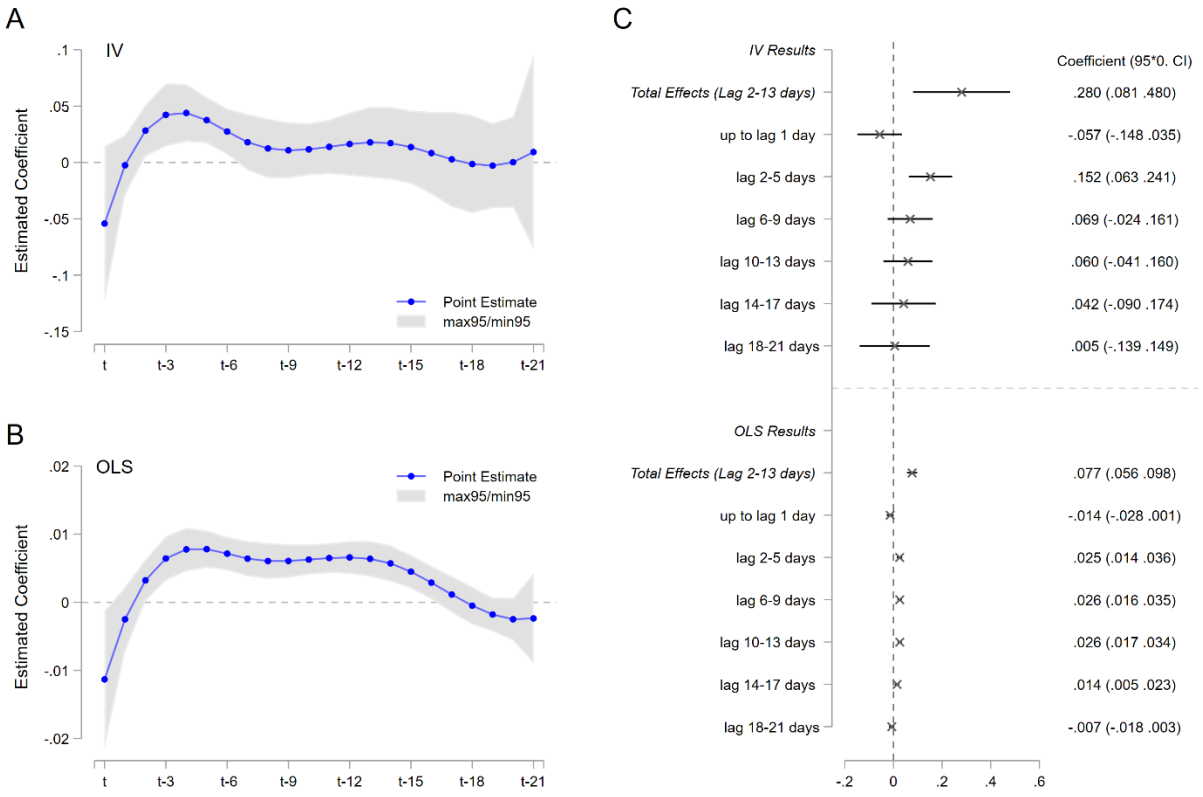
Notes: In Panel A, we plot daily confirmed cases (blue dash line), active infections (blue area), recovered (orange area), and deaths (red area) from 1 January to 1 April. The vertical line is the date of the Wuhan lockdown (23 January). Panel B plots the average growth rate of active infections. If it is larger than 0, the active infections increase. Panel C plots the trend of the Air Quality Index (AQI). Higher AQI means worse air pollution.

Figure 3. Variation in thermal inversion is strongly correlated with the Air Quality Index



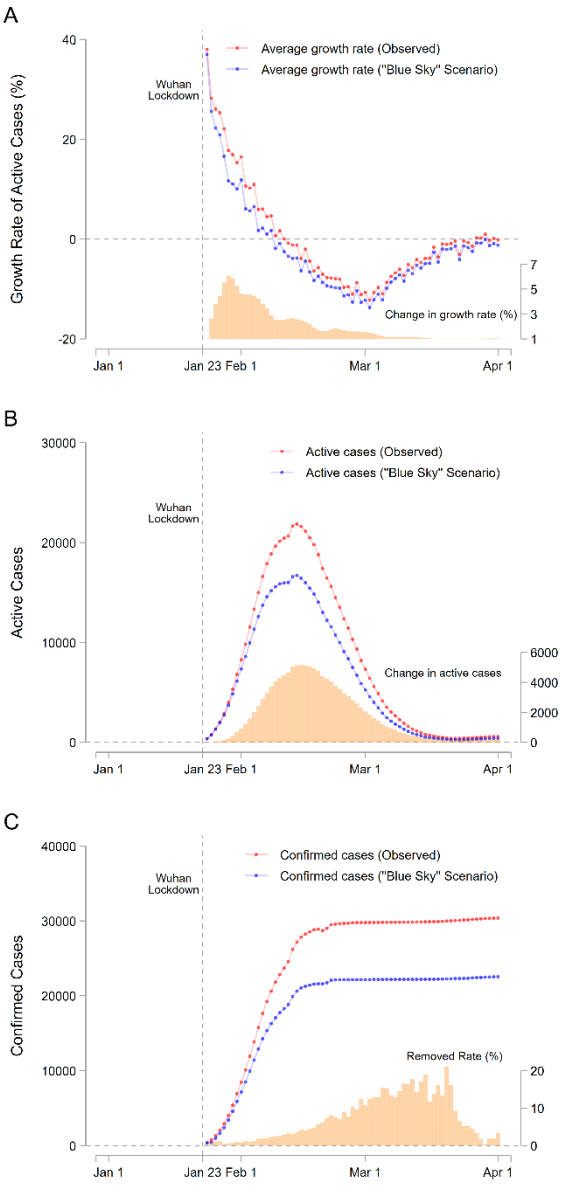
Notes: Panel A plots the distribution of thermal inversions, defined as the difference between the temperature in the second layer (320m) and that in the first layer (110m). The Y-axis represents the residual in variation in the AQI after controlling for temperature, precipitation, snow depth, city fixed effects, and date fixed effects. A higher temperature in the second layer is strongly associated with a larger residual in the AQI (worse air quality). In Panel B, we show that 1℃ increase in temperature inversion is associated with a 3.03 increase in the AQI. We find no statistically significant association between placebo thermal inversions and the AQI. In the placebo sample, we randomized the thermal inversions across different days within the city, or across different cities within a specific day.

Figure 4. Severe air pollution amplifies COVID-19 transmissivity with 2-13 days of delay



Notes: Panel A plots the IV estimates. Weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects are included in both the first and second stage regression. The blue dots represent the point estimates and the gray areas denote the 95% confidence intervals. Panel B reports the OLS estimates. The regression includes the same set of controls as in Panel A. Panel C on the right compares the magnitude of the effects between IV results and OLS results. The gray crosses represent the point estimates, and the gray lines show the 95% confidence intervals. In all regression, standard errors are clustered at the city level.

Figure 5. Simulated impacts of air pollution on the daily growth rate, active infections, and confirmed cases of COVID-19



Notes: Panel A represents the observed daily growth rate of COVID-19 and the hypothetical rate in the "blue sky" scenario (an AQI exceeding 100 is replaced by 100). The orange bar represents the change in the growth rate (right axis). Panel B shows the active cases in the two scenarios and the orange bar shows the difference. Panel C predicts the estimated cumulative confirmed cases over time. To compute these numbers, we use the observed removal rate, which is shown by the orange bars.

Appendix for

Air Pollution and COVID-19 Transmission in China

GUOJUN HE*, YUHANG PAN†, AND TAKANAO TANAKA‡

Contents:

* Division of Social Science, Division of Environment and Sustainability, and Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: gjhe@me.com).

† Division of Environment and Sustainability, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: yhyhpan@gmail.com).

‡ Tanaka: Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: takanao_tanaka@hotmail.co.jp).

Appendix A: The Effect of Air Pollution on the COVID-19 Death Rate

Existing studies suggest that air pollution increases deaths from COVID-19, likely because air pollution could impair the physical capability to recover from the infection. This is policy-relevant, but our dataset does not have sufficient statistical power to investigate the relationship. Outside Wuhan, there were only around 700 deaths from the pandemic in China, with about 90% of cities having no death or just one death. While more than 3,000 deaths are recorded in Wuhan, the data may not be accurate. For example, on 17 April 2020, the official COVID-19 death record in Wuhan was revised, with 1,290 deaths being added but without telling us the timing of the deaths.

Nevertheless, here we examine the relationship between air pollution and death rate by aggregating the daily level data to the weekly level so that we can have a wider variation in the outcome. Our outcome variable is not the growth rate of deaths because it should be proportional to the growth rate of the infections. Instead, we use the death rate, which is defined as the probability of death given infection. Using this outcome variable, we fit the following model, analogous to equation (7),

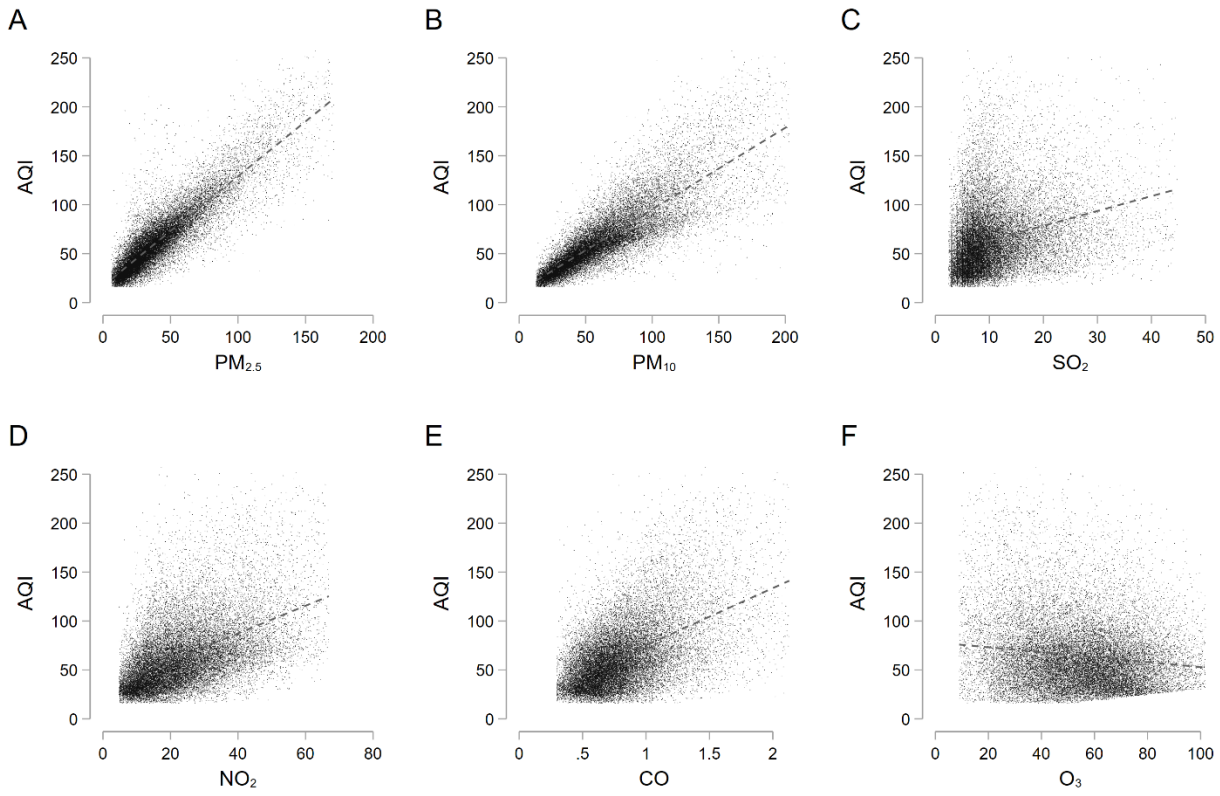
$$\rho_t = \frac{D_t - D_{t-1}}{I_{t-1}} = \frac{new\ deaths_t}{active\ infection_{t-1}} = \rho_{t0} + f(AP_{it}) + \varepsilon_{it} \qquad (A1)$$

where the death rate is a function of baseline death rate, air pollution, and mean zero error term. We use the same control variables and a set of fixed effects as in equations (9) – (11).

We do not find any suggestive evidence that air pollution increases the death rate of the patients (Appendix Figure 9). All coefficients seem to be very small, and the sign is not consistent. As expected, our data may not have sufficient statistical power to investigate the relationship accurately.

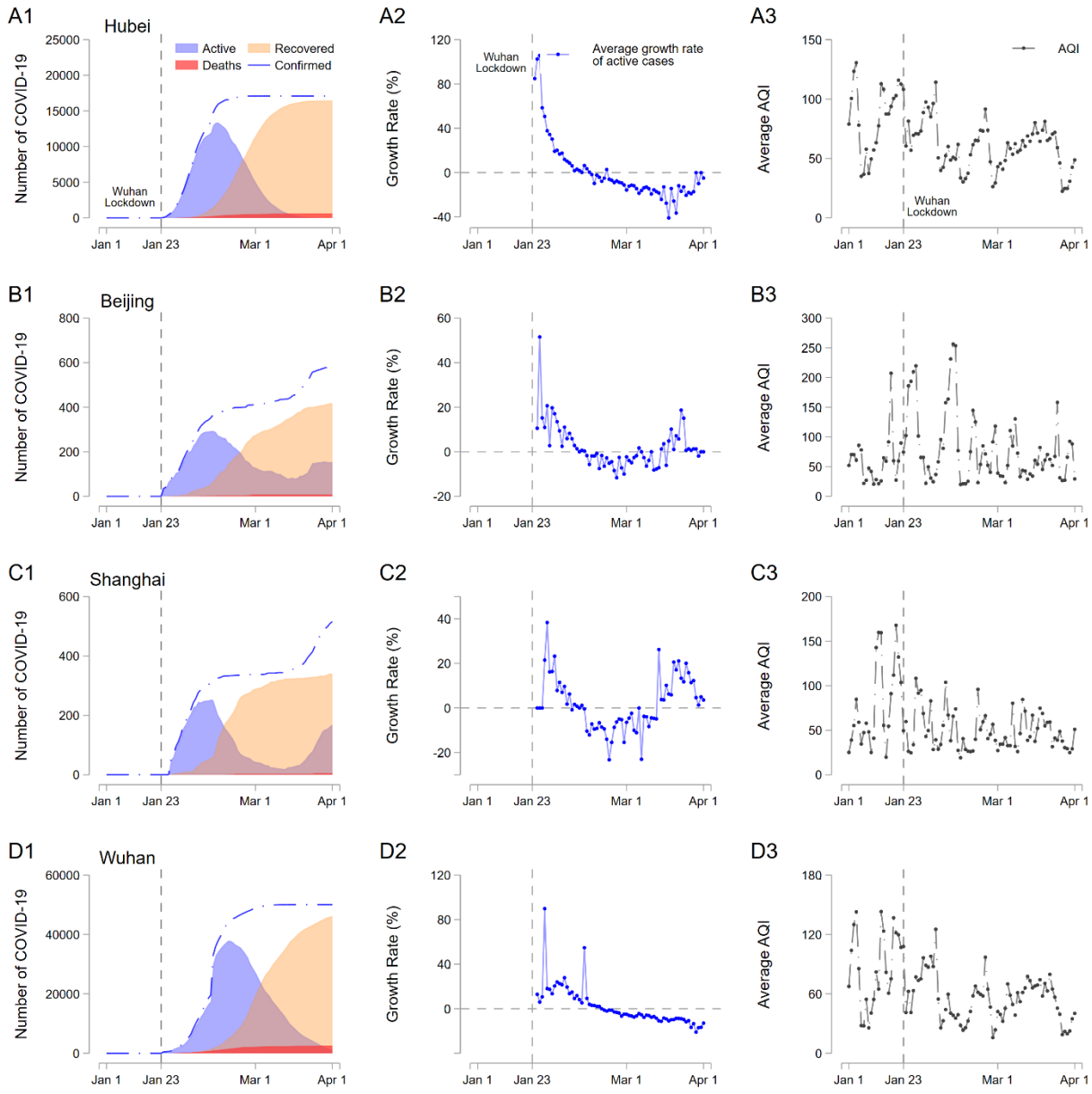
Appendix Figures

Appendix Figure 1. Air Quality Index is correlated with the most important pollutants



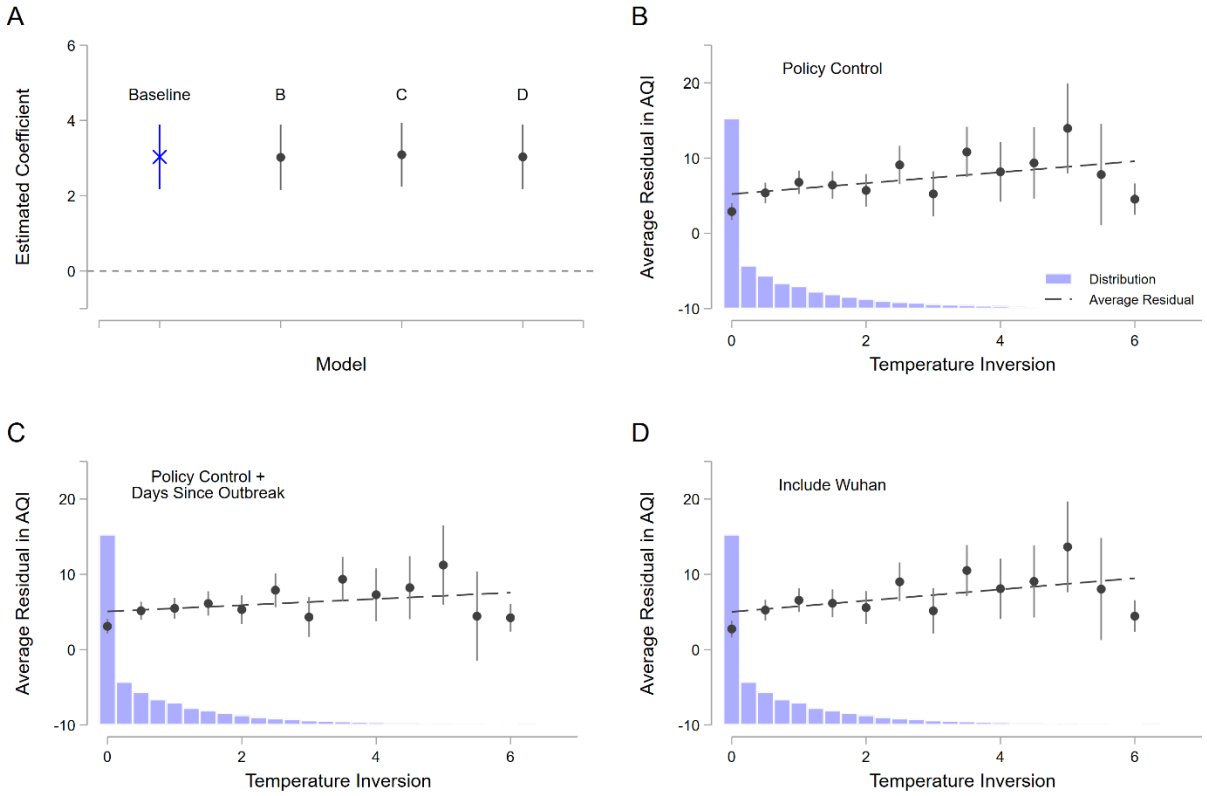
Notes: Panels A to F respectively represents the relationship between the AQI and each pollutant, including $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3, which are components of the index. During the wintertime in China, $PM_{2.5}$, PM_{10} are the primary pollutants in most cities and thus determine the levels of API. There is a strong correlation between the AQI and each pollutant, except for O_3 (ozone). This might be because ozone is mechanically negatively correlated with some of the primary pollutants. We trim the observations below 1 percentile and above 99 percentile in each pollutant. See Extended Data Table 1 for the definition of the AQI.

Appendix Figure 2. COVID-19, its growth rate, and the Air Quality Index in Hubei Province, Beijing, Shanghai, and Wuhan



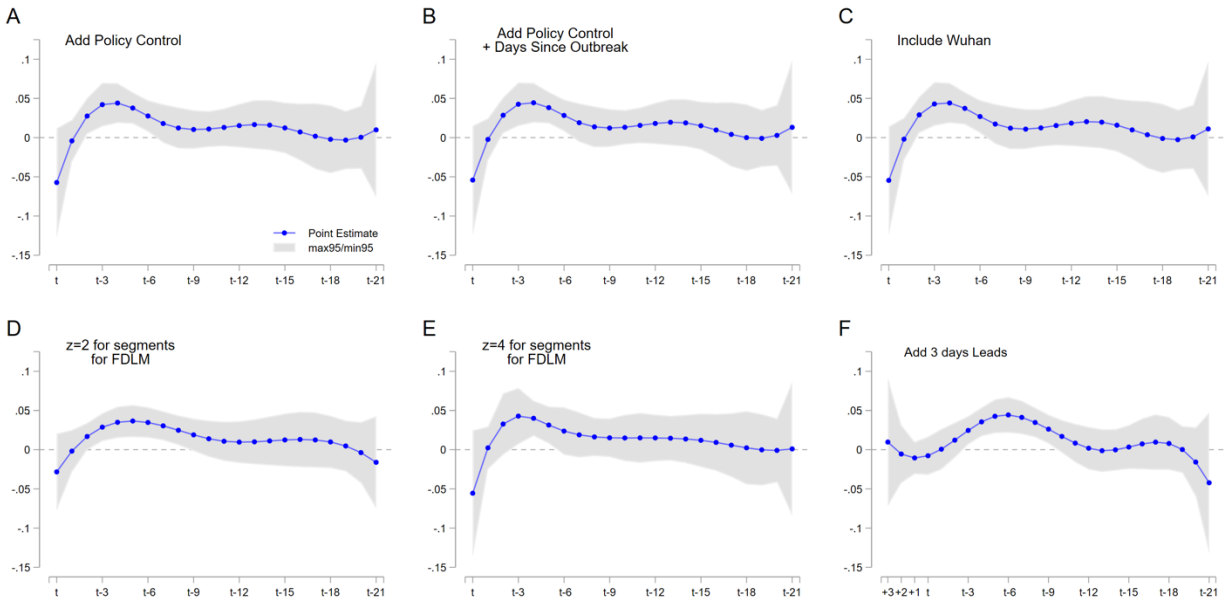
Notes: These figures show the COVID-19 outbreak (number of active, recovered, and deceased from COVID-19), the infection growth rate, and AQI in each region in different cities.

Appendix Figure 3. Strong correlation between thermal inversion and air pollution is robust to different specifications



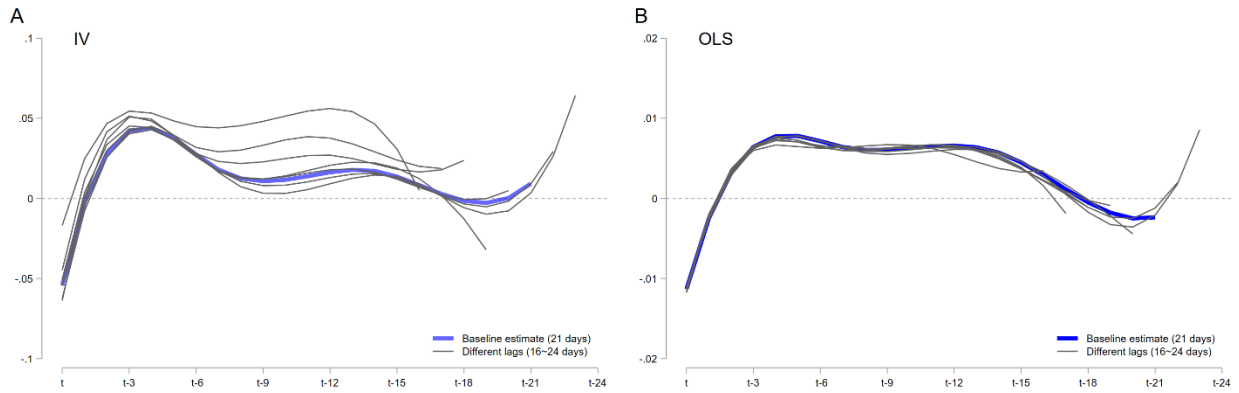
Notes: These figures show the correlations between temperature inversion and residuals in the Air Quality Index, after controlling for the weather variables, city fixed effects, and date fixed effects. Panel A plots the coefficients of the contemporaneous relationship for each model, which corresponds to Figure 3A. Panes B to D show the distribution of thermal inversions and the AQI residual. Panel B includes lockdown status as a control variable, Panel C includes lockdown status and days since the outbreak (first confirmed cases) as control variables. Panel D includes data from Wuhan city in the regression.

Appendix Figure 4. The effects of air pollution on the COVID-19 growth rate using the IV estimates are robust to a number of model specifications



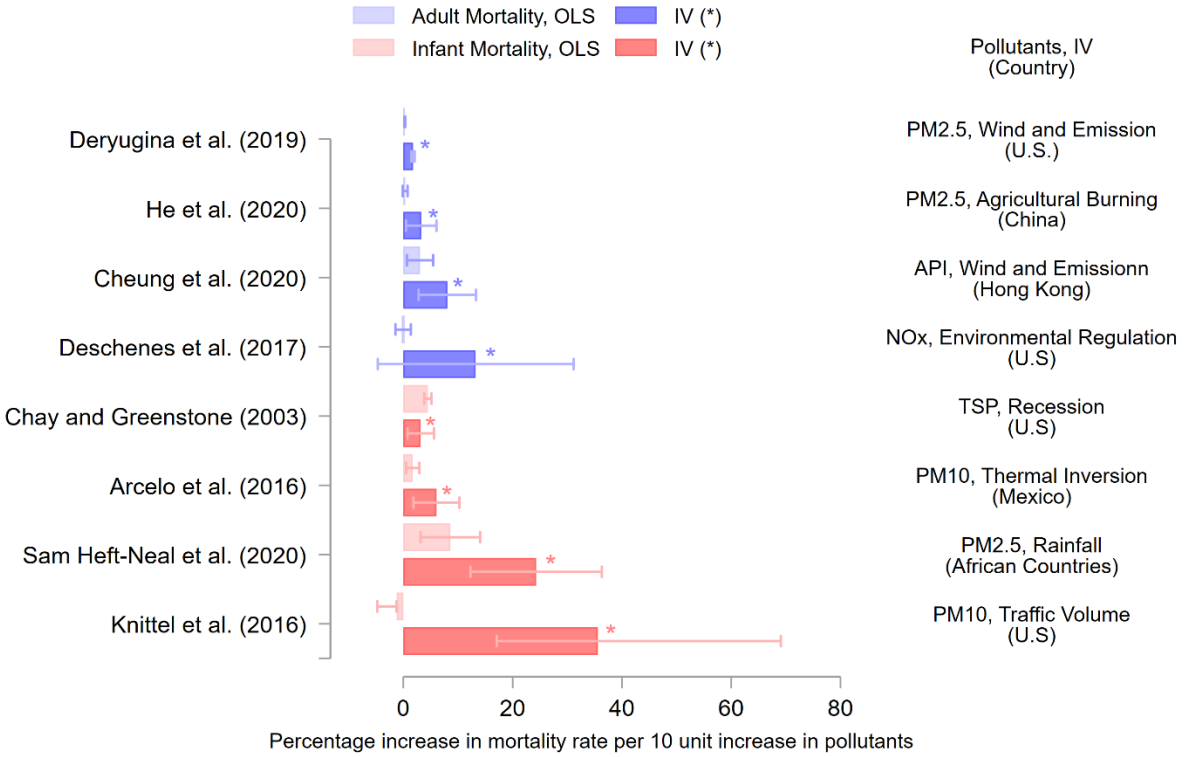
Notes: In Panel A, lockdown status is added as a control variable. In Panel B, the lockdown status and days since the outbreak (the first case confirmed) are added in the regression. In Panel C, we include Wuhan. In Panels D and E, we adopt different segments for the Flexible Distributed Lag Model. In Panel F, we add up to three days of leads of daily air pollution levels in the regression. The coefficients for the three leads are all close to zero, suggesting that future air pollution does not affect the current disease growth rate. Thermal inversions and the same controls are used for all first-stage regressions. All regressions include weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects. Standard errors are clustered at the city level.

Appendix Figure 5. The effects of air pollution on the COVID-19 growth rate using different lags



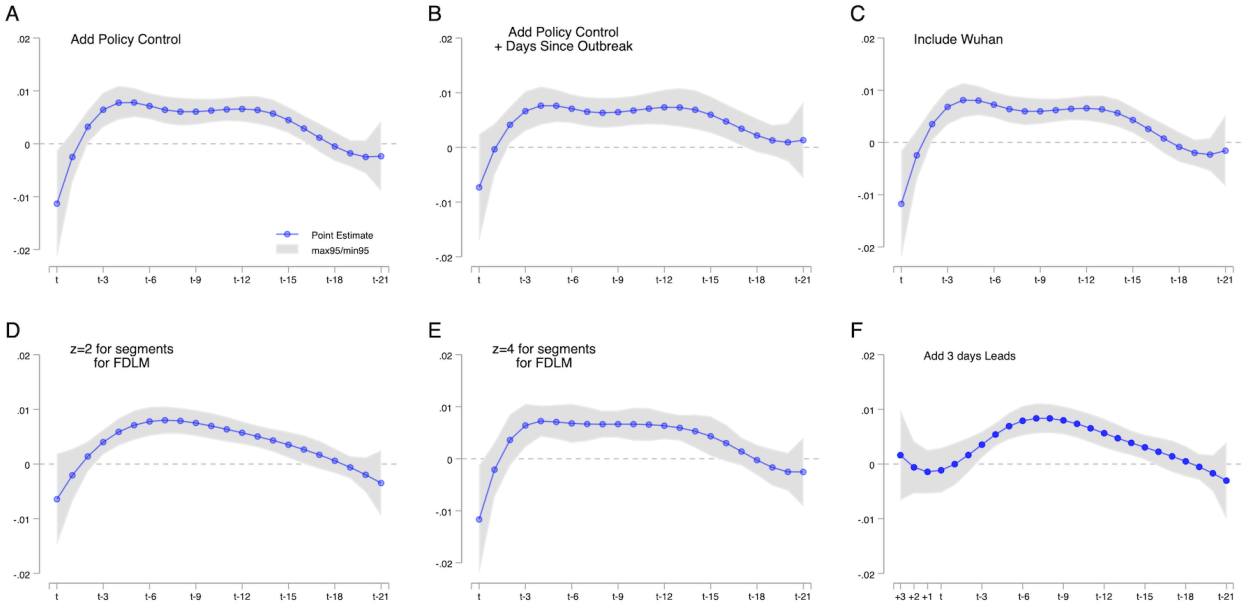
Notes: Panel A represents the IV estimates. We use different lengths of lags from 16 days to 24 days. The blue line plots the baseline estimates (21 days), and the gray line shows the results with different lags. Panel B shows the results using OLS estimates.

Appendix Figure 6. Instrumental variables (IV) estimates are consistently larger than OLS estimates in existing studies linking air pollution and health outcomes



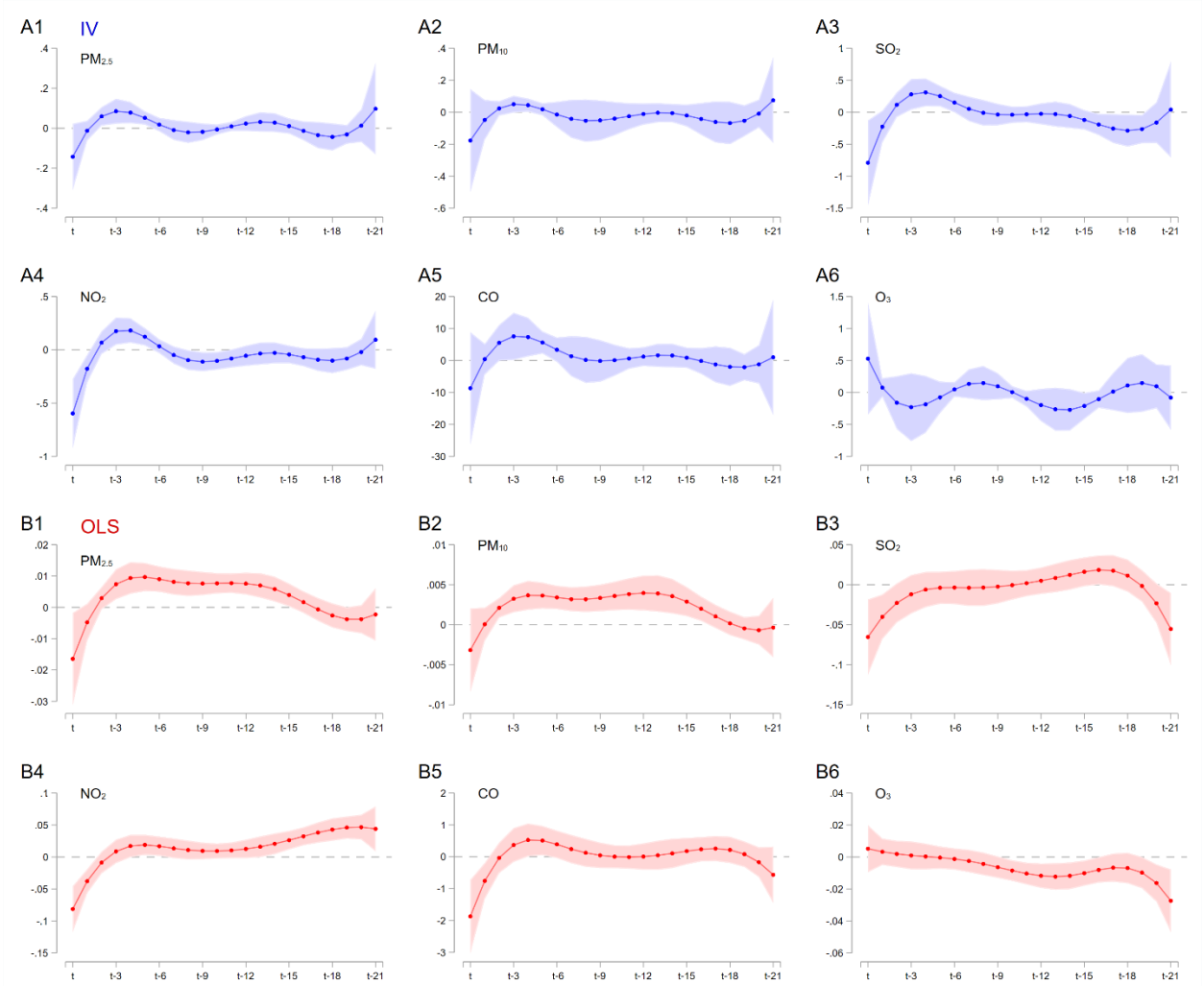
Notes: This figure plots the estimates of the effect per 10 unit increase in air pollution on mortality rate (%). Deryugina et al. (2019) use mortality among aged above 65. Except for Chay and Greenstone (2003), existing studies report the OLS estimates are smaller than the IV estimates. Note that different studies focus on different pollutants and use different instrumental variables. Therefore, we do not compare estimates across different studies but across different methods within each study.

Appendix Figure 7. The correlations between air pollution and COVID-19 growth rate using OLS estimates are robust to a number of model specifications



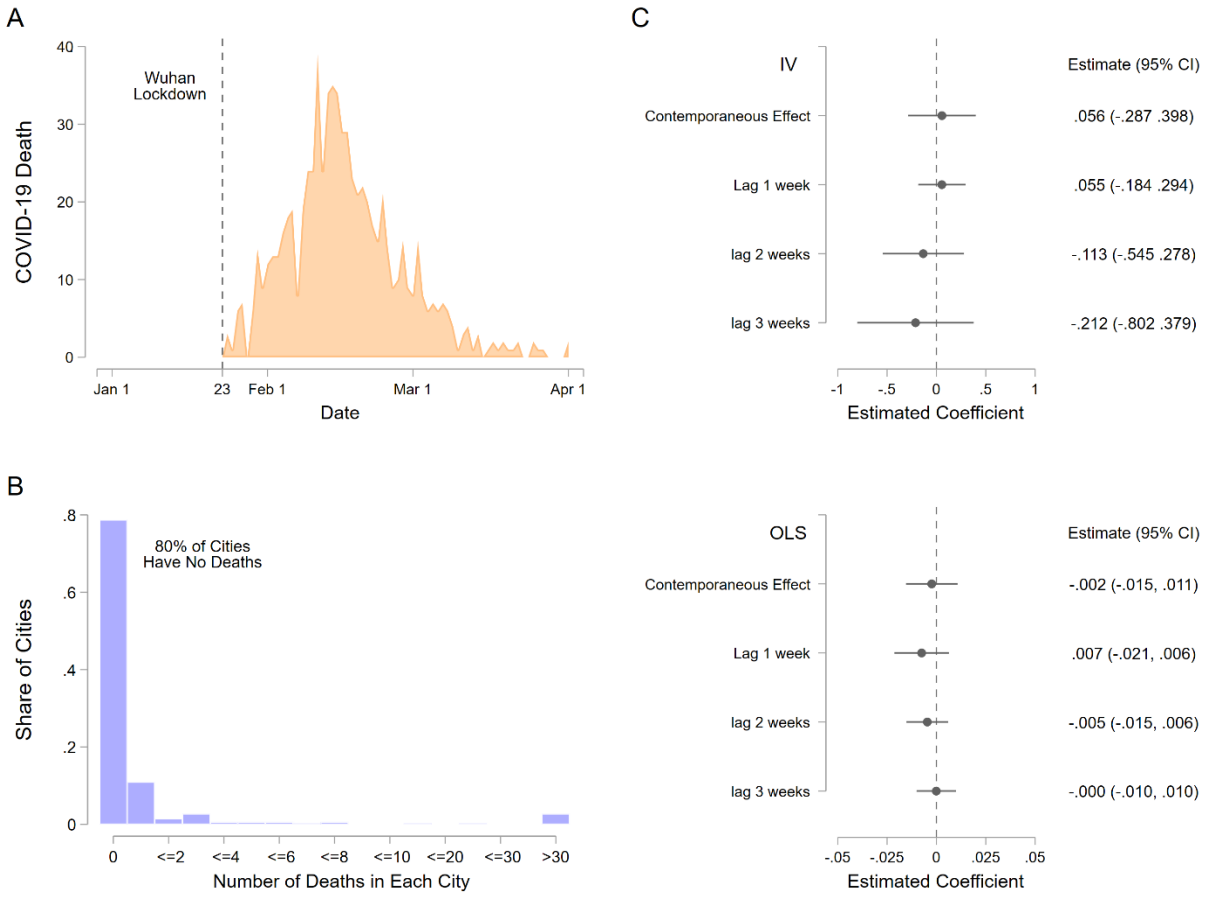
Notes: In Panel A, we include the lockdown status as a control variable. In Panel B, we include the lockdown status and days since the outbreak (the first case confirmed) as control variables. In Panel C, we include data from Wuhan in the regression. In Panels D and E, we use different segments for the Flexible Distributed Lag Model. In Panel F, we add three days of leads daily air quality levels in the regression. Thecoefficients for the three leads are all close to zero, suggesting that future air pollution levels are not correlated with the current disease growth rate. All regressions include weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects. Standard errors are clustered at the city level.

Appendix Figure 8. The effects of air pollution on the COVID-19 growth rate using different air pollutants.



Notes: Panels A1 to A6 plot the IV estimates. Each figure shows the coefficient of a 1 point increase in each pollutant in each day. Weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects are included in both the first and second stage regression. The blue dots represent the point estimates, while the blue areas denote the 95% confidence intervals. Panel Bs plot the OLS estimates with the same set of controls. The red dots represent the point estimates, while the red areas denote the 95% confidence intervals. In all regressions, standard errors are clustered at the city level.

Appendix Figure 9. Impacts of air pollution on the COVID-19 death rate



Notes: Panel A plots the trend in the number of the COVID-19 deaths over time outside Wuhan. Panel B plots the distribution of deaths in each city. Panel C on the right compares IV estimates with OLS estimates. We aggregate data to the week-by-city level for this analysis. We do not find a statistically significant relationship between AQI and the death rate.

Appendix Tables

Appendix Table 1. The relationship between the AQI and different air pollutants.

AQI	Each Air Pollutant						Air Quality Levels
	PM_{10} (24hr)	$PM_{2.5}$ (24hr)	NO_2 (24hr)	O_3 (8hr)	CO (24hr)	SO_2 (24hr)	
0-50	0-50	0-35	0-40	0-100	0-2	0-50	Excellent
50-100	50-150	35-75	40-80	100-160	2-4	50-150	Good
100-200	150-350	75-150	80-280	160-265	4-24	150-800	Slightly Polluted
200-300	350-420	150-250	280-565	265-800	24-36	800-1600	Moderately Polluted
300-400	420-500	250-350	565-750	/	36-48	1600-2100	Severely Polluted
400-500	500-600	350-500	750-940	/	48-60	2100-2620	Severely Polluted

Notes: This table reports the AQI sub-index levels for each air pollutant. The sub-index with the highest value will then be used as the AQI. For CO, the unit is mg/m^3, and for other pollutants, the units are $\mu g/m^3$.

Appendix Table 2. Summary Statistics

	Variable	Obs	Mean	Std. Dev.	Min	Max
A.	*COVID-19*					
	Growth Rate	30,268	18	109	-1	2760
B.	*Air Quality*					
	Air Quality Index	29,701	69.7	49.7	9.08	500
	$PM_{2.5}$ ($\mu g/m^3$)	29,705	46.3	39.7	1.83	1324
	PM_{10} ($\mu g/m^3$)	29,705	70.4	69.3	3.33	2342
	SO_2 ($\mu g/m^3$)	29,705	11.6	9.3	1	129
	NO_2 ($\mu g/m^3$)	29,705	24.7	15	1.67	115
	CO (mg/m^3)	29,705	0.869	0.449	0.1	6.26
	O_3 ($\mu g/m^3$)	29,701	53.6	20.9	1.55	200
C.	*Weather*					
	Thermal Inversion (°C)	29,029	0.864	1.23	0	10.2
	Temperature (°C)	30,268	5.95	8.52	-30.7	28.9
	Precipitation (mm)	30,268	23.3	61.4	0	1498
	Snow (depth, mm)	30,268	54.9	31.9	10	412

Notes: Growth rate is computed by taking the first difference in the natural logarithm of daily confirmed active cases.

Appendix Table 3. List of locked-down cities

Starting Date	Cities
23-Jan	Wuhan
24-Jan	Huangshi, Shiyan, Yichang, Ezhou, Jingmen, Xiaogan, Huanggang, Xianning, Enshi
25-Jan	Qinhuangdao
26-Jan	Xiangyang, Jingzhou, Xiantao
28-Jan	Tangshan
30-Jan	Dongying
31-Jan	Chongqing, Yinchuan, Wuzhong
2-Feb	Wenzhou
3-Feb	Wuxi, Jining
4-Feb	Harbin, Nanjing, Xuzhou, Changzhou, Nantong, Hangzhou, Ningbo, Fuzhou, Jingdezhen, Zaozhuang, Linyi, Zhengzhou, Zhumadian
5-Feb	Shenyang, Dalian, Anshun, Fushun, Benxi, Dandong, Jinzhou, Fuxin, Liaoyang, Panjin, Tieling, Chaoyang, Huludao, Yangzhou, Hefei, Quanzhou, Nanchang, Jinan, Qingdao, Taian, Rizhao, Laiwu, Nanning
6-Feb	Tianjin, Shijiazhuang, Suzhou, Pingxiang, Jiujiang, Xinyu, Yingtan, Ganzhou, Ji'an, Yichun, Fuzhou, Shangrao, Neijiang, Yibin, Xinyang
7-Feb	Suzhou, Guangzhou
8-Feb	Shenzhen, Foshan, Fangchenggang,
9-Feb	Cangzhou, Huaibei
10-Feb	Beijing, Shanghai
13-Feb	Hohhot, Baotou, Wuhai, Chifeng, Tongliao, Ordos, Hulun Buir, Bayan Nur, Ulanqab, Xing'an League, Xilingol League, Alxa League

Notes: The lockdown information is from local government and various media news in 2020.

Appendix Table 4. Full results of the effect of air quality on the COVID-19 growth rate

Dependent Variable: Daily disease growth rate	OLS		IV	
	coefficients	s.e.	coefficients	s.e.
	(1)	(2)	(3)	(4)
Current Day	-0.011	[0.005]	-0.054	[0.035]
lag: 1 day	-0.003	[0.002]	-0.002	[0.014]
lag: 2 days	0.003	[0.002]	0.028	[0.012]
lag: 3 days	0.006	[0.002]	0.042	[0.014]
lag: 4 days	0.008	[0.002]	0.044	[0.013]
lag: 5 days	0.008	[0.001]	0.038	[0.010]
lag: 6 days	0.007	[0.001]	0.028	[0.010]
lag: 7 days	0.006	[0.001]	0.018	[0.013]
lag: 8 days	0.006	[0.001]	0.012	[0.013]
lag: 9 days	0.006	[0.001]	0.011	[0.013]
lag: 10 days	0.006	[0.001]	0.012	[0.012]
lag: 11 days	0.006	[0.001]	0.014	[0.012]
lag: 12 days	0.007	[0.001]	0.016	[0.014]
lag: 13 days	0.006	[0.001]	0.018	[0.016]
lag: 14 days	0.006	[0.001]	0.017	[0.016]
lag: 15 days	0.004	[0.001]	0.014	[0.016]
lag: 16 days	0.003	[0.001]	0.008	[0.019]
lag: 17 days	0.001	[0.001]	0.003	[0.021]
lag: 18 days	-0.000	[0.001]	-0.001	[0.022]
lag: 19 days	-0.002	[0.001]	-0.003	[0.019]
lag: 20 days	-0.002	[0.002]	0.000	[0.020]
lag: 21 days	-0.002	[0.003]	0.009	[0.044]
z-Segment, k-order	3, 3		3, 3	
Observations (cities)	22,701 (329)		22,701 (329)	
Weather Control	Y		Y	
Date fixed effects	Y		Y	
City fixed effects	Y		Y	

Notes: The results correspond to Figure 4. The dependent variable is the day-by-city level growth rate of the activated COVID-19 cases. Each estimate indicates the effect of the current and past air pollution (Air Quality Index) on the growth rate of COVID-19. Weather controls include temperature, precipitation, and snow depth. Standard errors are clustered at the city level and shown in the right-side brackets.