Online Appendix for Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China

By Jiaxiu He, Haoming Liu and Alberto Salvo

A. Further discussion of institutions and data

The work day choice. In each 30-day period of department activity (that is, excluding department holidays with zero aggregate output), we observe workers typically producing on 21 to 25 12-hour shifts in Jiangsu, and 17 to 21 12-hour shifts in Henan (Figure A.1).[1] Recalling Henan's standard shift design—work 12 hours, rest 24 hours—and that mathematically there are 20 such 36-hour cycles in a 30-day period, the evidence indicates that worker absenteeism is low. In Jiangsu—work 12 hours, rest 12 hours, and take leave one day every two to four consecutive work days—hours worked are higher than in Henan. We do not observe which of a given worker's days with zero output were scheduled (predetermined) leave days, and which days were unplanned absences. However, to judge by the higher work hours than in Henan, worker absenteeism in Jiangsu is similarly low.

Our empirical analysis considers the possibility of selection on pollution in our worker-day output samples (see the next section). A worker's absence on a given day could be due to pollution-related sickness, or to weather (a correlate of pollution) shifting the value of the outside option, such as clear skies inducing the worker to spend a leisure day outdoors (Shi and Skuterud, 2015). We do not find a significant association between pollution and the probability of positive versus zero output on a given day (Table A.5 and Figure A.4). This is not surprising given the low worker absenteeism. Our interpretation is that these are relatively healthy Chinese workers who live close to their employer, an employer that has much information about the employee, say to verify a shirking employee's false claim of sickness on justifying an unplanned absence. Moreover, scheduled leaves, like department holidays, are predetermined, so they are unlikely to depend on short-run variation in environmental quality. In practice, we find that attempting to control for selection in the output equation, using exogeneity restrictions based on concurrent weather, makes little difference to our estimates of the relationship between pollution and productivity (Table 4 versus Table A.6).

[1] We ignore the density at 30 days for Henan, as this consists of 15 temporary workers who appear briefly in the records, with positive output for up to three consecutive days only (i.e., 3 days with output / 3 days from worker's first day to last day in sample × 30 days in a 30-day period).

Jiangsu's monthly work shift transition. To preserve the continuity of operations, the monthly shift transition (day shift versus night shift) happens at the end of each month over a few days, rather than abruptly on day 1 of the subsequent month. For example, October 2014 records indicate that Team 1 worked the day shift and Team 2 worked the night shift. However, on October 30, 2014, we observe a total of 12 Team 1 workers and 5 Team 2 workers with non-zero output. Given the required staff of 8 to 9 workers per shift (17 machines at 2 machines per worker), the asymmetry in the number of workers from each team who worked on October 30 is due to the fact that, while some Team 2 workers took leave, some Team 1 workers already transitioned to working the night shift alongside the other Team 2 workers. Since output records do not mention who the Team 1 workers newly transitioned to the night shift were, nor when the transition started (e.g., October 28 or 29), our regression models include indicators for the last three days of the month interacted with shift, to account for measurement error in the time of day for these few days and for some workers (three Team 1 workers in the example). It is during the last three days of each month that the numbers of Team 1 and Team 2 workers with non-zero output are most unbalanced. Estimates are robust to alternatively dropping these month-end observations.

Worker tasks, products and piece rates. A Jiangsu worker operates machines (most often two) as they wind yarn threads from long narrow ring bobbins into large cones. Over time and following a production schedule, the team leader assigns workers to products and machines, but the worker keeps to her workstation, with the assigned products and machines, for at least an entire shift. As days, weeks, or months pass, a worker may operate different machines, but these machines are similar. Each machine has 16 magazines and each magazine can hold 5 ring bobbins at a time. It takes 10-15 minutes for a machine to completely wind the yarn, depending on the thickness of the yarn, at which point the worker has to replace the bobbins. To do so, one end of thread in a new bobbin needs to be pulled out and attached to the magazine, which requires skill and effort from the worker. In addition, threads naturally break apart during machine operation. The Jiangsu machines—which are more sophisticated than at the Henan department—can connect broken threads automatically, taking 5-10 seconds each time. However, this process may fail, in which case a red light alerts the worker, who needs to intervene and manually reconnect the thread as quickly as possible, for the machine to resume operation. The Jiangsu department produces products with some differentiation, with 80s accounting for the largest share of output in our sample (a mean of 32% of the within-shift department output across date-shifts), followed by 50s (22%), 60s (18%), 45s (14%), and 100s (12%). These five products thus account for 98% of combined compensation-adjusted output. Another four products (34s, 40s, 70s, and 120s) together account for only 2% of output.

Henan workers perform similar tasks. They walk up and down the aisles in their workstation, each worker attending to about five machines, and reconnecting the threads that break during machine operation. Again, a team leader assigns workers to products and machines, and a worker keeps to the assigned products and machines through the shift. One product, 32s, accounts for about half the department's output (a mean of 48% of the aggregate output across date-shifts), followed by 40s (18%), 21s (13%), and 24s (10%). Jointly, these four products account for 89% of output, and six other products account for the remainder (7s, 16s, 18s, 19.5s, 26s, and 32.5s). As in the Jiangsu department, a worker's product-machine assignment is persistent over days and weeks.

Piece rates in Henan, expressed in CNY/kg, are: 0.41 for output of product 40s, 0.37 for 32s on an extended machine, 0.29 for 32s on a regular machine, 0.25 for 24s, 0.235 for 26s, 0.205 for 21s, 0.19 for 19.5s and 18s, 0.165 for 16s, and 0.10 for 7s. In addition to these products, a "preparatory task" whose output is recorded is compensated at piece rates of 0.22 CNY/kg on an extended machine and 0.205 CNY/kg on a regular machine.

To determine compensation, the Jiangsu department aggregates output across individual products using the following weights, expressed in adjusted cases/physical case for each product: 1.50 for output of 120s, 1.40 for 100s, 1.30 for 80s, 1.15 for 70s, 1.06 for 60s, 1.00 for 50s, 0.98 for 45s and 40s, and 0.96 for 34s. A worker's aggregate number of adjusted cases is converted into "counts" at a rate of 105 counts/adjusted case. For output produced in the 12-hour shift, expressed in counts, a worker enjoys a bonus of 0.01 CNY per count in excess of 11,000 counts but no greater than 12,000 counts (i.e., 10 CNY per 1,000 excess counts), a bonus of 0.02 CNY per count in excess of 12,000 counts but no greater than 13,000 counts, and a bonus of 0.03 CNY per count for output in excess of 13,000 counts. The penalty scheme for 12-hour output that falls short of 11,000 counts is symmetric to the bonus scheme around the reference output of 11,000 counts. This variable compensation scheme can be visualized in Figure A.2.

Differences in piece rates across products reflect differences in the standard rates of output from variation in product thread. Worker tasks are very similar across products. Figure A.3 shows that the assignment of product varieties to workers exhibits strong day-to-day persistence at both work sites. To prepare the plots by work site, we take the four or five varieties in a work site with double-digit output shares, and aggregate the remaining varieties into a residual category labeled "Other." We consider pairs of consecutive work shifts by a worker, with 12 hours and up to 24 hours of rest in between work in Jiangsu and Henan, respectively, given the standard shift designs. We then compute the empirical probability that, conditional on a worker producing a given variety on a shift, she also produced that variety in her preceding shift (likely among others—Table 2). These product-level transition probabilities within worker are of the order of 60 to 80%.

Cleaning the worker output samples. For each department, we observe worker output by date-shift pair. Table A.1 describes the minimal data cleaning we apply to the raw records. We drop the one shift that precedes (resp., follows) a department holiday, during which machines are turned off early (resp., turned on late) for cleaning (resp., setting up). We drop a few date-shifts for which exceptional conditions were annotated, such as the occurrence of a power outage that affects all workers who worked on that shift. For the analysis of individual output, we additionally drop a few observations for which exceptional conditions were annotated, such as the worker being unusually assigned multiple tasks during the shift, beyond working at her single workstation, or was not working in the main ground-level production facility (Henan).

Air pollution data availability. Hourly PM2.5 and SO_2 mass concentration measurements at official air monitoring sites in both towns that host the workplaces we study are quite complete during the sample periods.[2] The exception is the Henan workplace during 2014, for which ambient air data are available only at three neighboring cities at most 60 km away. As stated below, the few missing observations are fairly evenly distributed throughout the sample periods and do not cluster on specific days, suggesting that the sites are well maintained.

For the PM2.5 monitor 3.7 km from the Jiangsu workplace, only 3% of all 7272 hours during the period August 2, 2014 (30 days prior to the start of the output sample) to May 31, 2015 exhibit missing concentrations in the raw data, say due to equipment failure. We impute the relatively few hourly values that are missing for the closest monitor using the mean value recorded contemporaneously at three other PM2.5 monitors located in the same town. Pairwise correlation coefficients for PM2.5 measured hour by hour at the closest monitor and at each of the other three same-town monitors range between 0.93 and 0.96, indicating high levels of spatial correlation within the Jiangsu town.

Similarly, for the SO_2 monitor 3.7 km from the Jiangsu workplace, 4% of all hourly observations are missing during the period, and we impute the few missing values using the mean value recorded at three other SO_2 monitors located in the same town. We noticed that the SO_2 series for a given monitor can exhibit short-lived and isolated spikes, fluctuations that are often not observed around the same hours at the other SO_2 monitors. For example, the five hourly readings starting at 8 am on August 15, 2014 at the closest monitor are 13, 15, 219, 30 and 39 $\mu g/m^3$. For this reason, when using SO_2 in our specifications, we take mean SO_2 levels across monitors in the town. The correlation coefficient between mean SO_2 across monitors in the town and PM2.5 at the monitor closest to the work site is 0.61 (again, hourly series).

For the PM2.5 monitor 3.4 km from the Henan workplace, only 2% of all 3624 hourly records during the period January 1, 2015 to May 31, 2015 are missing. We recode to missing 24 successive hourly records starting at noon on January 18, 2015 that are fixed at 60 $\mu g/m^3$, when observations at three other monitors in

[2]We acquired local pollutant concentrations from the Chinese Ministry of Environmental Protection, via `http://www.pm25.in`. PM2.5 concentrations from the US State Department for its regional US embassies, which we describe in the text, are available via `www.stateair.net/web/post/1/1.html`.

the same town vary from hour to hour and in step. As in Jiangsu, we then impute the few values that are missing for the workplace's closest monitor using the mean contemporaneous value at the same town's three other PM2.5 monitors. Again, hourly PM2.5 readings exhibit high levels of spatial correlation; pairwise correlation coefficients for measurements at the closest monitor and measurements at the other same-town monitors are at least 0.90. Starting in March 2014 (one month prior to the start of the output sample), hourly data availability is similarly high at PM2.5 monitors in three neighboring cities at most 60 km away, 17 monitors in all. For the overlapping 2015 period, the correlation coefficient for PM2.5 measured 3.4 km from the Henan workplace and mean PM2.5 across the neighboring cities is 0.80.

Similar comments to those above apply to hourly SO_2 concentrations measured in the town that hosts the Henan workplace (4 monitors, the closest being 3.4 km away) and at the neighboring cities (17 monitors). For the overlapping 2015 period, the correlation coefficient for mean SO_2 measured at the host town and mean SO_2 measured across the neighboring cities is 0.76.

Surface and atmospheric meteorological data availability. As stated in the text, we obtain local surface-level temperature, humidity and precipitation readings, at three-hour intervals, from NASA. For each workplace, we select the geographic cell (0.25 degree by 0.25 degree) in the NASA data that is centered on its host town. We obtain surface-level readings for wind speed and direction, and for atmospheric temperature differences, at 12-hour intervals, from NOAA. For each workplace, we select the reference location in the NOAA data that is closed to the host town, namely station identification codes CHM00058362 for the Jiangsu site and CHM00057083 for the Henan site.[3]

Both datasets are quite complete during the sample periods. For example, surface-level data in the NOAA data for stations CHM00058362 and CHM00057083 are missing for one and six 12-hour readings, respectively. Whenever the wind direction or wind speed readings on the surface are missing, we take the corresponding parameter reading above the surface at an atmospheric pressure point of 1000 mb, if the latter is available; failing that, we take the wind direction or speed reading further above the surface at an atmospheric pressure point of 925 mb.

B. A selection model.

The empirical design is based on a more general model of the worker's problem, where the worker's productive but costly effort choice e, conditional on attending work, is embedded in the choice of attending work over an unscheduled absence

[3]NASA and NOAA data are available via `http://daac.gsfc.nasa.gov/` and `http://www1.ncdc.noaa.gov/pub/data/igra/`.

with reservation utility ϕ:

$$\max\left\{\max_e wq(e,a) - c(e,Z), \phi(Z,W)\right\}. \tag{B.1}$$

Conditional on selecting into work on a given day, the worker is paid a piece rate w per unit of individual output quantity q. Besides its dependence on effort, output can shift with the worker's ability a. The worker's exposure to ambient air pollution Z, both concurrent to and preceding the day of work, shifts her cost of effort and, thus, the payoff from work.

Pollution exposure Z may also affect selection into work on a given day.[4] The weather-sheltered workplace in our settings—temperature and humidity control, indoor protection from rain—provides a natural exclusion restriction for identifying the selection equation, since concurrent weather W can shift the value of outdoor leisure ϕ but not the payoff from work. The epidemiological studies cited in the text do not model a distributed lag for weather.

The empirical counterpart to framework (B.1) is then:

$$d^*_{it} = \gamma_0 + Z_t\delta + W_t\gamma_1 + \gamma_t + \gamma_i + \zeta_{it}, \quad d_{it} = 1[d^*_{it} > 0], \tag{B.2}$$

$$q_{it} = \alpha_0 + Z_t\beta + X_{it}\alpha_1 + \alpha_t + \alpha_i + \epsilon_{it} \;\text{ if } d_{it} = 1, \tag{B.3}$$

where $d_{it} = 1[d^*_{it} > 0]$ indicates worker i's choice of coming to work on day t, and q_{it} is her conditional output. Vector Z_t consists of concurrent pollution and, more generally, lagged-day pollution levels. W_t are concurrent weather covariates, namely, temperature, humidity and rain that may shift the worker's reservation utility. As explained in the text, X_{it} is a vector of observed worker-day output controls. To account for systematic seasonality and worker heterogeneity, both selection and output equations include time fixed effects, γ_t and α_t, and individual worker fixed effects, γ_i and α_i. ζ_{it} and ϵ_{it} are idiosyncratic error terms.

In a regression of (B.3) without the worker fixed effects α_i, a sufficient condition for consistently estimating β, pollution's effect on output, using pooled OLS is $E[\alpha_i+\epsilon_{it}|Z_t, X_{it}, \alpha_t, d_{it} = 1] = 0$ (e.g., Dustmann and Rochina-Barrachina, 2007). The OLS estimator will be biased downward, for example, if workers with higher ability α_i are more likely to call in absent when faced with high pollution exposure Z_t. Or say that more vulnerable workers are more likely to call in sick when Z_t rises, and these workers are less productive than other workers, in which case the estimate for β will biased upward.

The inclusion of worker fixed effects in (B.3) can remove the bias caused by selection on α_i. To obtain a consistent estimate of β, a sufficient condition is (Wooldridge, 1995):

$$E[\epsilon_{it} - \epsilon_{is}|Z_t, X_{it}, \alpha_t, Z_s, X_{is}, \alpha_s, d_{it} = d_{is} = 1] = 0 \;\text{ for periods } s \neq t,$$

This condition will only be violated if there is selection on ϵ_{it}. For example, a

[4]Since the work shift starts at a predetermined time for the worker's team and is of fixed 12-hour duration, we ignore these additional margins of labor supply.

worker who experiences a positive output shock on day t (high ϵ_{it}) is more likely to call in absent when exposed to high rather than low pollution. Further, the worker would need to observe ϵ_{it} prior to making the work versus non-work choice. We can then use (B.2) to estimate the probability that a worker chooses work on each day, and correct for this probability in (B.3).

C. Further results (more prolonged exposure)

Worker attendance. Panels in the third column of Figure A.4 illustrate the relationship between attendance and cumulative pollution exposure (a P-day lag structure). The outcome variable is one if the worker produced output on a department work day, and zero otherwise. We consider a sample of worker by day observations in which the worker either produced output or her first day of zero output immediately following a day she worked. As in the concurrent pollution analysis of Table A.5, concurrent weather W is allowed to shift attendance. We plot 2SLS estimates,[5] using the full sample period for Henan. Point estimates of the effect on attendance from cumulative pollution exposure, as we raise P, are negative in Jiangsu and mostly positive in Henan, but in both cases they are at best marginally statistically significant and economically modest, peaking at -1% and +1%, respectively, per +10 μg/m^3 sustained increase in PM2.5. Effects for SO_2 are higher than for PM2.5 as SO_2 levels in μg/m^3 are lower.

Individual heterogeneity. We estimate output equation (1) allowing the response to prolonged pollution exposure to vary by worker i, i.e., write β_i instead of β. The estimation sample, by site, includes the 25 workers with the most output observations. We interact each exposure variable $Z_{p,t}$ (24-hour PM2.5 or SO_2 at daily lag p of output day t) with a set of worker fixed effects for the 25 individuals. Within worker, exposure coefficients $\beta_{i,p}$ are constrained over p according to her own PDL$(P, 4)$. The Henan sample is the full April 2014 to May 2015 period. Panels in the second column of Figure A.4 plot the 25th, 50th and 75th percentiles in the cross-worker distributions of 2SLS estimates for $\sum_{p=0}^{P} \hat{\beta}_{i,p}$ as we vary P. The median effect across workers is similar to the mean effect in panels (a) to (d) of Figure 6(I)-6(II); at the peak, the 75th percentile can reach twice the median. Results for specific workers, of varying output sensitivity to pollution, are summarized in panels (e) and (f) of Figure 6(I)-6(II) (sensitivity is ranked based on the $P = 20$ model).

Figure A.5 further examines the individual worker response to sustained pollution exposure. Panels (a) and (b) show that a worker's rank in terms of output sensitivity to pollution exposure ("q sens.") is similar whether we use PM2.5 or SO_2. We also rank workers according to their mean output in the sample and, in panels (c) to (f), find that a worker's output rank tends to correlate positively with her output sensitivity to pollution. We further rank workers according to their attendance sensitivity to pollution ("d sens."), similarly based on a PDL$(20, 4)$, but

[5] Atmospheric ventilation V influences pollution but is excluded from the attendance equation (B.2).

do not find an association with the worker's output sensitivity rank—see panels (g) to (j).

REFERENCES

Dustmann, Christian, and María Engracia Rochina-Barrachina. 2007. "Selection correction in panel data models: An application to the estimation of females' wage equations." *Econometrics Journal*, 10(2): 263–293.

Shi, J., and M. Skuterud. 2015. "Gone fishing! Reported sickness absenteeism and the weather." *Economic Enquiry*, 53(1): 388–405.

Wooldridge, Jeffrey M. 1995. "Selection corrections for panel data models under conditional mean independence assumptions." *Journal of Econometrics*, 68(1): 115–132.

Table A.1—: Procedure to prepare the raw worker output data

Procedure	Observations: Dropped	Observations: Remaining	No. of workers: Dropped	No. of workers: Remaining
(I) **Jiangsu**				
Initial number of observations (worker by day pairs)		4579		47
(−) One shift prior to/after dept. holiday (to clean/setup)	32	4547	0	47
(=) **Number of observations for aggregate analysis**		4547		47
(−) Observations with exceptional duties recorded	47	4500	1	46
(=) **Number of observations for individual analysis**		4500		46
Proportion of raw data in the final sample		98%		98%
(II) **Henan**				
Initial number of observations (worker by day pairs)		5614		83
(−) One shift prior to/after dept. holiday (to clean/setup)	124	5490	0	83
(−) Shifts with power outage recorded	52	5438	0	83
(=) **Number of observations for aggregate analysis**		5438		83
(−) Observations with exceptional duties recorded	131	5307	3	80
(=) **Number of observations for individual analysis**		5307		80
Proportion of raw data in the final sample		95%		96%

Table A.2—: Summary statistics for atmospheric ventilation conditions, by work site

Variables	N	Mean	Std.Dev.	Min.	Max.
Jiangsu					
Wind speed at the surface (m/s)	525	2.66	1.46	0.00	8.00
Wind direction at the surface (all hours from a given direction=1)					
... from North	525	0.20	0.36	0.00	1.00
... from East	525	0.42	0.46	0.00	1.00
... from South	525	0.12	0.30	0.00	1.00
... from West	525	0.26	0.43	0.00	1.00
Temperature difference (°C) for increasing altitudes at standard atmospheric pressure levels					
... from surface to 1000 mb	525	-0.43	1.75	-3.20	9.00
... from 1000 to 925 mb	525	-3.50	2.15	-6.60	8.00
... from 925 to 850 mb	525	-2.58	2.19	-6.60	5.40
... from 850 to 700 mb	525	-5.50	3.66	-12.90	5.60
... from 700 to 500 mb	525	-14.12	2.79	-20.90	-3.80
Henan					
Wind speed at the surface (m/s)	778	2.00	1.18	0.00	8.00
Wind direction at the surface (all hours from a given direction=1)					
... from North	784	0.12	0.28	0.00	1.00
... from East	784	0.41	0.45	0.00	1.00
... from South	784	0.25	0.38	0.00	1.00
... from West	784	0.21	0.37	0.00	1.00
Temperature difference (°C) for increasing altitudes at standard atmospheric pressure levels					
...from surface to 1000 mb	778	0.35	1.30	-2.00	8.90
...from 1000 to 925 mb	784	-3.18	2.25	-6.40	7.80
...from 925 to 850 mb	784	-3.73	1.82	-7.20	4.30
...from 850 to 700 mb	784	-8.29	3.48	-15.60	3.00
...from 700 to 500 mb	784	-14.72	3.03	-22.60	-1.40

Note: An observation is a date-shift in the corresponding output sample. Sample periods are September 1, 2014 to May 31, 2015 (Jiangsu) and April 1, 2014 to May 31, 2015 (Henan).

Table A.3—: Output and concurrent pollution: Robustness to untrimmed individual-level samples

Jiangsu work site	OLS categorical		OLS linear		2SLS linear			
Dep. variable is individual worker output	(1) PM2.5	(2) SO_2	(3) PM2.5	(4) SO_2	(5) PM2.5	(6) SO_2		
PM2.5 $\in$ [50,100) or $SO_2 \in$ [25,50) μg/m^3 (yes=1)	0.17	-0.48						
	(0.30)	(0.32)						
PM2.5 $\geq$ 100 or $SO_2 \geq$ 50 μg/m^3 (yes=1)	-0.25	-0.85*						
	(0.50)	(0.51)						
PM2.5 or SO_2 (per 10 μg/m^3)			0.00	-0.06	0.01	-0.23		
			(0.05)	(0.12)	(0.05)	(0.28)		
Observations	4,491	4,491	4,491	4,491	4,491	4,491		
Number of regressors	71	71	70	70	70	70		
First-stage F-statistic (excluded instruments)					225	183		
R-squared	0.84	0.84	0.84	0.84	0.84	0.84		
Mean value of dependent variable (cases worker^{-1} day^{-1})	111.6	111.6	111.6	111.6	111.6	111.6		
Henan work site	OLS categorical		OLS linear		2SLS linear		2SLS lin., neighbor poll.	
Dep. variable is individual worker output	(7) PM2.5	(8) SO_2	(9) PM2.5	(10) SO_2	(11) PM2.5	(12) SO_2	(13) PM2.5	(14) SO_2
PM2.5 $\in$ [50,100) or $SO_2 \in$ [25,50) μg/m^3 (yes=1)	0.76	1.14						
	(0.73)	(1.00)						
PM2.5 $\in$ [100,150) or $SO_2 \in$ [50,75) μg/m^3 (yes=1)	0.90	1.99*						
	(1.20)	(1.16)						
PM2.5 $\geq$ 150 or $SO_2 \geq$ 75 μg/m^3 (yes=1)	-0.82	-0.05						
	(1.65)	(1.40)						
PM2.5 or SO_2 (per 10 μg/m^3)			-0.00	0.13	-0.09	0.04	-0.00	0.19
			(0.10)	(0.10)	(0.13)	(0.12)	(0.10)	(0.20)
Observations	2,110	2,110	2,110	2,110	2,110	2,110	5,117	5,117
Number of regressors	67	67	65	65	65	65	113	113
First-stage F-statistic (excluded instruments)					186	53	96	72
R-squared	0.80	0.80	0.79	0.79	0.79	0.79	0.72	0.72
Mean value of dependent variable (CNY worker^{-1} day^{-1})	118.5	118.5	118.5	118.5	118.5	118.5	125.4	125.4
Worker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Worker-date-shift characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-of-day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The same notes as in Table 3(I)-3(II) apply except that the individual analyses here, for the Jiangsu site in the top panel and the Henan site in the bottom panel, do not trim samples at three standard deviations of the sample mean.

TABLE A.4—: MULTI-POLLUTANT MODELS (PM2.5 AND SO_2) OF CONCURRENT EXPOSURE AND OUTPUT

Dependent variable is average output per worker	**Jiangsu work site**		**Henan work site**		
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) 2SLS, neighbor poll.
PM2.5 (per 10 μg/m^3)	-0.01	0.07	-0.03	-0.12	0.00
	(0.04)	(0.08)	(0.13)	(0.17)	(0.13)
SO_2 (per 10 μg/m^3)	-0.17	-0.47	0.14	0.24	0.40*
	(0.13)	(0.39)	(0.16)	(0.18)	(0.24)
Combined effect 1: +10 μg/m^3 PM2.5 and +10 μg/m^3 SO_2	-0.18	-0.40	0.11	0.12	0.40
	(0.13)	(0.33)	(0.11)	(0.09)	(0.25)
Combined effect 2: +10 μg/m^3 PM2.5 and +5 μg/m^3 SO_2	-0.09	-0.17	0.04	0.00	0.20
	(0.08)	(0.14)	(0.09)	(0.10)	(0.16)
Date-shift characteristics	Yes	Yes	Yes	Yes	Yes
Observations	524	524	258	258	756
Number of regressors	24	24	21	21	30
First-stage F-statistic (excluded instruments)		53		134	43
R-squared	0.85	0.85	0.76	0.76	0.77
Mean value of dependent variable (cases or CNY worker^{-1} day^{-1})	111.4	111.4	119.6	119.6	127.4
Dependent variable is individual worker output	**Jiangsu work site**		**Henan work site**		
	(6) OLS	(7) 2SLS	(8) OLS	(9) 2SLS	(10) 2SLS, neighbor poll.
PM2.5 (per 10 μg/m^3)	0.02	0.11	-0.09	-0.19	-0.08
	(0.04)	(0.08)	(0.13)	(0.14)	(0.10)
SO_2 (per 10 μg/m^3)	-0.08	-0.45	0.16	0.14	0.26
	(0.10)	(0.37)	(0.13)	(0.16)	(0.21)
Combined effect 1: +10 μg/m^3 PM2.5 and +10 μg/m^3 SO_2	-0.06	-0.34	0.07	-0.05	0.18
	(0.10)	(0.32)	(0.10)	(0.14)	(0.20)
Combined effect 2: +10 μg/m^3 PM2.5 and +5 μg/m^3 SO_2	-0.02	-0.11	-0.01	-0.12	0.05
	(0.06)	(0.14)	(0.10)	(0.12)	(0.12)
Worker fixed effects	Yes	Yes	Yes	Yes	Yes
Worker-date-shift characteristics	Yes	Yes	Yes	Yes	Yes
Observations	4,443	4,443	2,085	2,085	5,054
Number of regressors	70	70	65	65	106
First-stage F-statistic (excluded instruments)		55		186	46
R-squared	0.83	0.83	0.79	0.79	0.70
Mean value of dependent variable (cases or CNY worker^{-1} day^{-1})	112.4	112.4	119.6	119.6	126.6
Year-month fixed effects	Yes	Yes	Yes	Yes	Yes
Day-of-week fixed effects	Yes	Yes	Yes	Yes	Yes
Time-of-day fixed effects	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is average output per worker in columns (1) to (5), or a worker's output in columns (6) to (10). An observation is a date-shift (top panel), or a worker by date-shift with non-zero output (bottom panel; we trim the sample at three standard deviations of the sample mean). Sample periods are September 1, 2014 to May 31, 2015 (Jiangsu), January 1, 2015 to May 31, 2015 (Henan) or April 1, 2014 to May 31, 2015 (Henan with neighboring-city pollution). Pollutant concentrations are concurrent means during each 12-hour date-shift. See the notes to Table 3(I) (Jiangsu) and Table 3(II) (Henan) for date-shift characteristics and worker-date-shift characteristics. OLS estimates or 2SLS estimates, where measured pollution, Z, is instrumented using pollution predicted from atmospheric ventilation conditions, $\hat{Z}$ (see note 20). To reflect common emission sources, we report an additional "combined effect" based on a mean PM2.5 to SO_2 concentration ratio of about 2 (Table 2). Standard errors (in parentheses) are Newey-West with a 30-day lag, in the top panel, and additionally account for correlation across workers within day (day clusters), in the bottom panel. ***Significant at 1%, **at 5%, *at 10%.

TABLE A.5—: THE RELATIONSHIP BETWEEN ATTENDANCE AND IMMEDIATE EXPOSURE TO POLLUTION

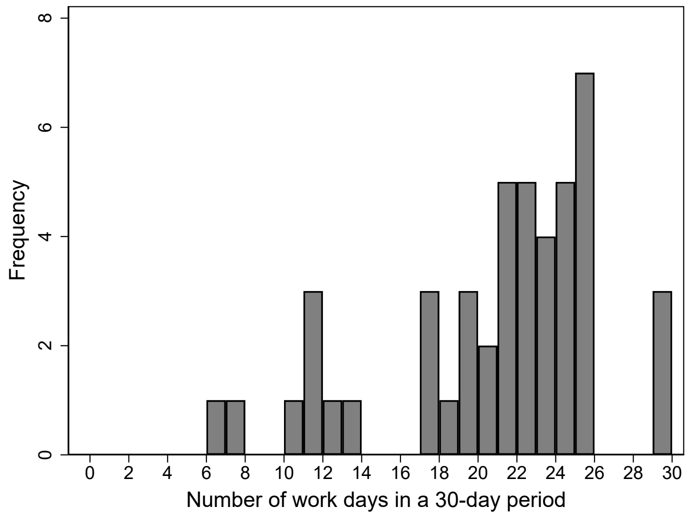
Jiangsu work site	OLS categorical		OLS linear		2SLS linear			
Dependent variable is individual works on day (yes=1)	(1) PM2.5	(2) SO_2	(3) PM2.5	(4) SO_2	(5) PM2.5	(6) SO_2		
PM2.5 $\in$ [50,100) or $SO_2 \in$ [25,50) μg/m^3 (yes=1, previous day)	-0.02**	-0.01						
	(0.01)	(0.01)						
PM2.5 $\geq$ 100 or $SO_2 \geq$ 50 μg/m^3 (yes=1, previous day)	-0.01	-0.01**						
	(0.01)	(0.01)						
PM2.5 or SO_2 (per 10 μg/m^3, previous day)			-0.00	-0.00	-0.00	0.00		
			(0.00)	(0.00)	(0.00)	(0.01)		
Indicator for an initial two months at the firm	Yes	Yes	Yes	Yes	Yes	Yes		
Observations	5,671	5,671	5,671	5,671	5,671	5,671		
Number of regressors	88	88	87	87	87	87		
First-stage F-statistic (excluded instruments)					159	155		
R-squared	0.03	0.03	0.03	0.03	0.03	0.03		
Mean value of dependent variable (works on day=1)	0.79	0.79	0.79	0.79	0.79	0.79		
Henan work site	OLS categorical		OLS linear		2SLS linear		2SLS lin., neighbor poll.	
Dependent variable is individual works on day (yes=1)	(7) PM2.5	(8) SO_2	(9) PM2.5	(10) SO_2	(11) PM2.5	(12) SO_2	(13) PM2.5	(14) SO_2
PM2.5 $\in$ [50,100) or $SO_2 \in$ [25,50) μg/m^3 (yes=1, previous day)	0.04	-0.03*						
	(0.04)	(0.02)						
PM2.5 $\in$ [100,150) or $SO_2 \in$ [50,75) μg/m^3 (yes=1, previous day)	0.07	-0.01						
	(0.05)	(0.02)						
PM2.5 $\geq$ 150 or $SO_2 \geq$ 75 μg/m^3 (yes=1, previous day)	0.09	-0.01						
	(0.09)	(0.03)						
PM2.5 or SO_2 (per 10 μg/m^3, previous day)			0.00	0.00	0.01	0.00	0.00	-0.00
			(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
Observations	3,230	3,230	3,230	3,230	3,230	3,230	8,123	8,123
Number of regressors	79	79	77	77	77	77	132	132
First-stage F-statistic (excluded instruments)					115	26	55	55
R-squared	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04
Mean value of dependent variable (works on day=1)	0.66	0.66	0.66	0.66	0.66	0.66	0.65	0.65
Weather W on current & previous day (a set for each day)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Worker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is an indicator that the worker produced output on a department work day. An observation is a worker by day in which the worker either produced output or her first day of leave excluding department holidays. Sample periods are September 1, 2014 to May 31, 2015 in the top panel, and January 1, 2015 (April 1, 2014 in columns (13)-(14)) to May 31, 2015 in the bottom panel. Pollutant concentrations are 24-hour means in the preceding day. Weather W on current and previous day includes 24-hour means for temperature (in 3 °C bins) and 24-hour means for humidity and rainfall (linear and quadratic terms). OLS estimates or 2SLS estimates, where measured pollution is instrumented using pollution predicted from atmospheric ventilation conditions (see note 20). Standard errors, in parentheses, account for correlation across workers within day (day clusters) and over time (autocorrelation bandwidth of 30 days, Bartlett kernel). ***Significant at 1%, **at 5%, *at 10%.

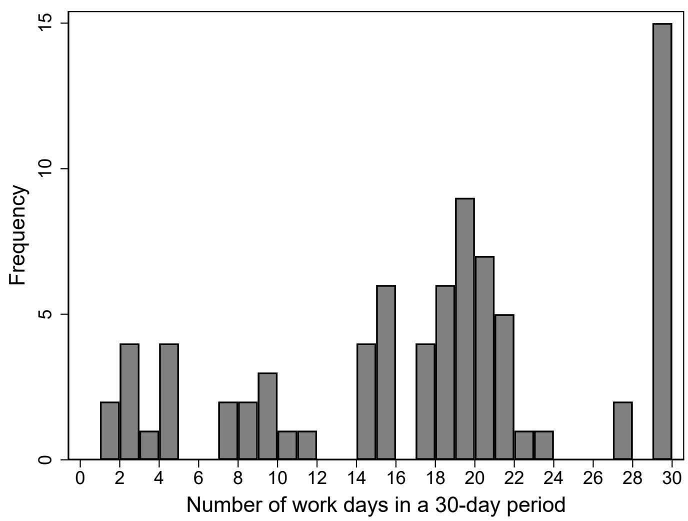
Table A.6—: Robustness to adding a selection correction in the output equation

Lag structure	Jiangsu		Henan		Henan, neighbor poll.	
(P)	PM2.5	SO_2	PM2.5	SO_2	PM2.5	SO_2
5	-0.06	-0.55**	0.02	0.33	-0.09	0.35
	(0.14)	(0.26)	(0.23)	(0.33)	(0.14)	(0.33)
10	-0.19	-0.19	-0.66**	0.46	-0.17	-0.39
	(0.19)	(0.29)	(0.33)	(0.51)	(0.18)	(0.44)
15	-0.30	-0.11	-0.75	-0.70	-0.05	-1.05**
	(0.22)	(0.27)	(0.57)	(0.81)	(0.20)	(0.52)
20	-0.47**	-0.12	-1.24	-0.70	-0.29	-2.08***
	(0.22)	(0.40)	(0.80)	(1.07)	(0.19)	(0.65)
25	-0.62***	-1.27**	-1.26	-1.83	-0.44*	-2.26***
	(0.21)	(0.54)	(1.03)	(1.39)	(0.27)	(0.63)
30	-0.34*	-1.23**	-1.97*	-4.89***	-0.67*	-2.53***
	(0.20)	(0.50)	(1.06)	(1.54)	(0.35)	(0.71)

Note: 2SLS estimates of PDL($P, 4$) output equations that include a selection correction imputed from a UDL(P) probit attendance regression, where P is the number of lags in days. Each row and column reports the coefficient and standard error (in parentheses, clustered on day) on the cumulative effect on worker output from concurrent and lagged pollution exposure, $\sum_{p=0}^{P} \beta_p$, estimated from a separate quartic distributed lag model, and pollution variables are daily 24-hour means. Sample periods to May 31, 2015 start on September 1, 2014 (Jiangsu), January 1, 2015 (Henan with host-town pollution) or April 1, 2014 (Henan with neighboring-city pollution). As in Table 4, we report estimates as a proportion of mean worker output. ***Significant at 1%, **at 5%, *at 10%.



(a) Jiangsu worker sample



(b) Henan worker sample

figure A.1. : Number of 12-hour shifts worked per 30-day period

Notes: Frequency chart across workers (46 in Jiangsu, 80 in Henan) of the number of 12-hour shifts worked per 30-day period of department activity. This is calculated as a worker's number of positive output records (work days) in the sample divided by the number of days with department activity between the worker's first and last work day in the sample, multiplied by 30 days (i.e., per 30-day period).

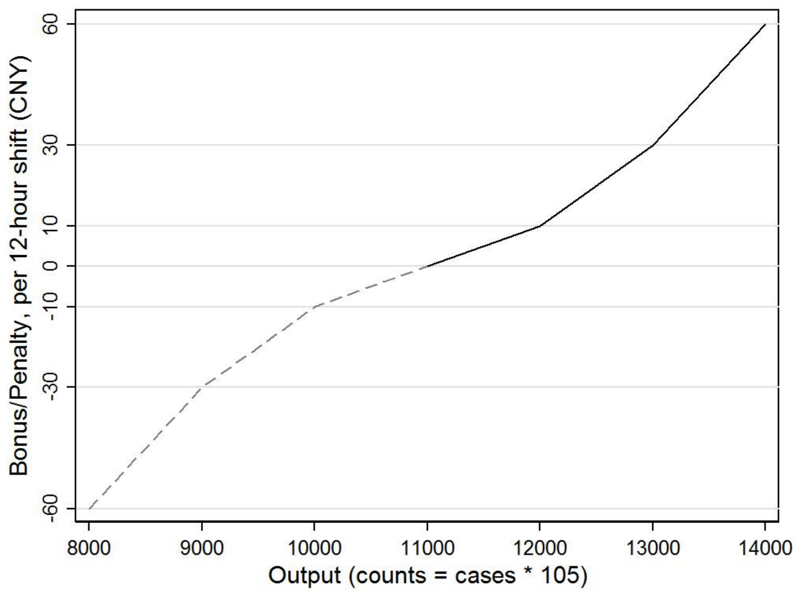
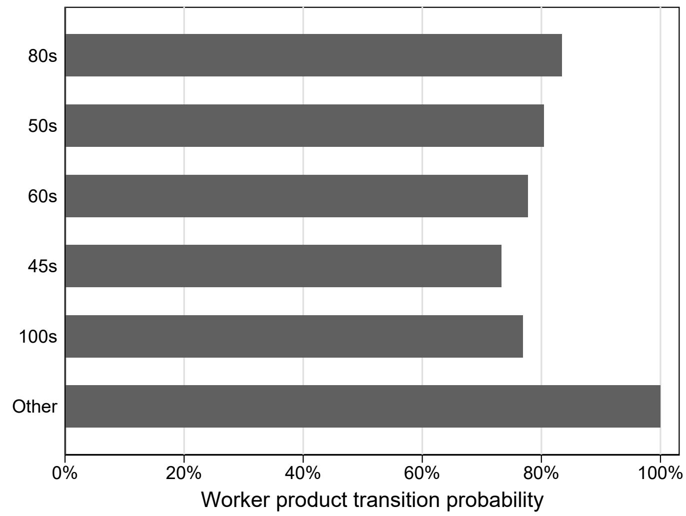
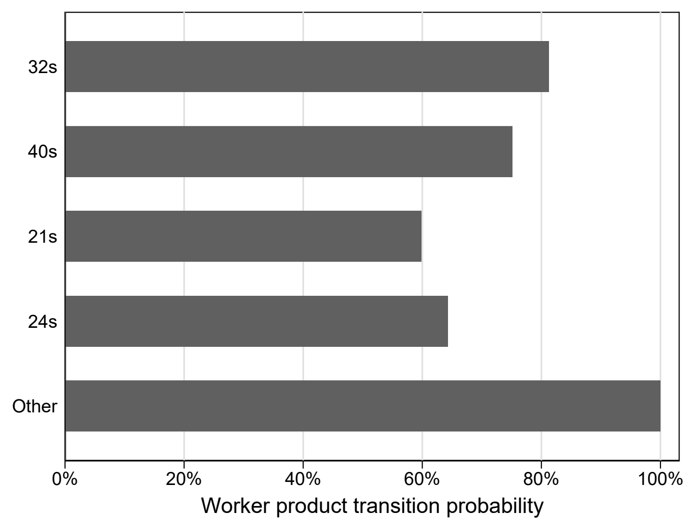


FIGURE A.2. : VARIABLE COMPENSATION SCHEME IN JIANGSU

Notes: A Jiangsu department worker's output during a 12-hour shift is aggregated across individual products into "adjusted cases" using the weights listed in the Appendix text, and then converted into "counts" at a rate of 105 counts/adjusted case. The worker is paid a bonus of: 0.01 CNY per count in excess of 11,000 counts but no greater than 12,000 counts, 0.02 CNY per count in excess of 12,000 counts but no greater than 13,000 counts, and 0.03 CNY per count for output in excess of 13,000 counts. The penalty scheme for output that falls short of 11,000 counts is symmetric to the bonus scheme around the reference output of 11,000 counts.



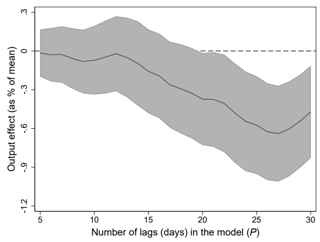
(a) Jiangsu worker sample



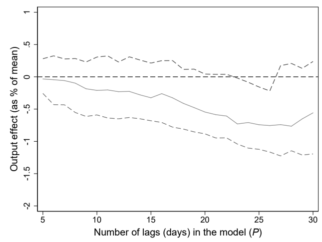
(b) Henan worker sample

FIGURE A.3. : A worker's shift-to-shift own-product transitions

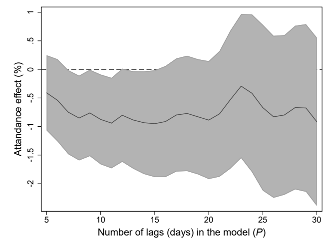
Notes: To prepare each plot, we take the five and four varieties in each work site with double-digit output shares, and aggregate the remaining varieties into a residual category labeled "Other." We consider pairs of consecutive work shifts by a worker, with 12 hours and up to 24 hours of rest in between work in Jiangsu and Henan, respectively, given the standard shift designs. We then compute the empirical probability that, conditional on a worker producing a given variety on a shift, she also produced that variety in her preceding shift.



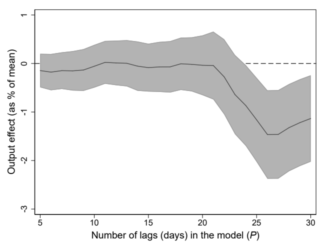
(a) Jiangsu, PM2.5, OLS, output



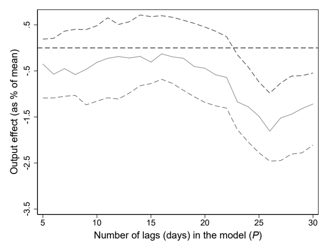
(b) Jiangsu, PM2.5, 2SLS, p25-p75



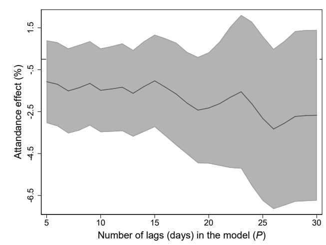
(c) Jiangsu, PM2.5, 2SLS, attendance



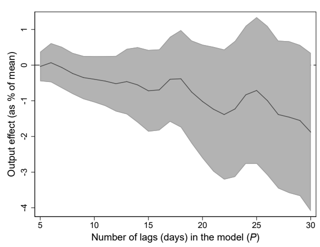
(d) Jiangsu, SO_2, OLS, output



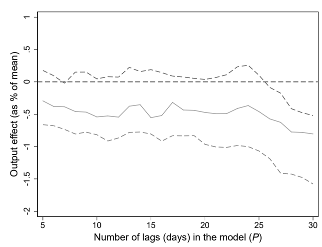
(e) Jiangsu, SO_2, 2SLS, p25-p75



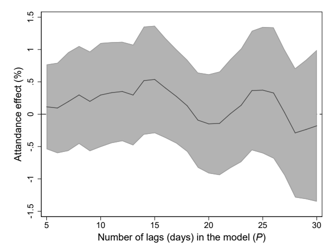
(f) Jiangsu, SO_2, 2SLS, attendance



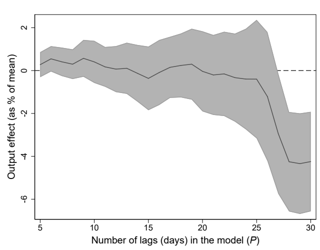
(g) Henan, PM2.5, OLS, output



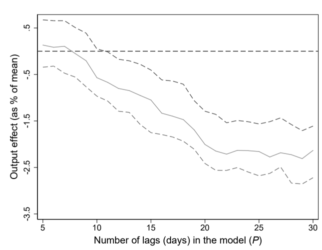
(h) Henan, PM2.5, 2SLS, p25-p75



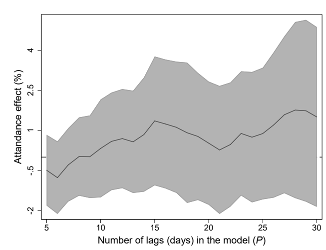
(i) Henan, PM2.5, 2SLS, attendance



(j) Henan, SO_2, OLS



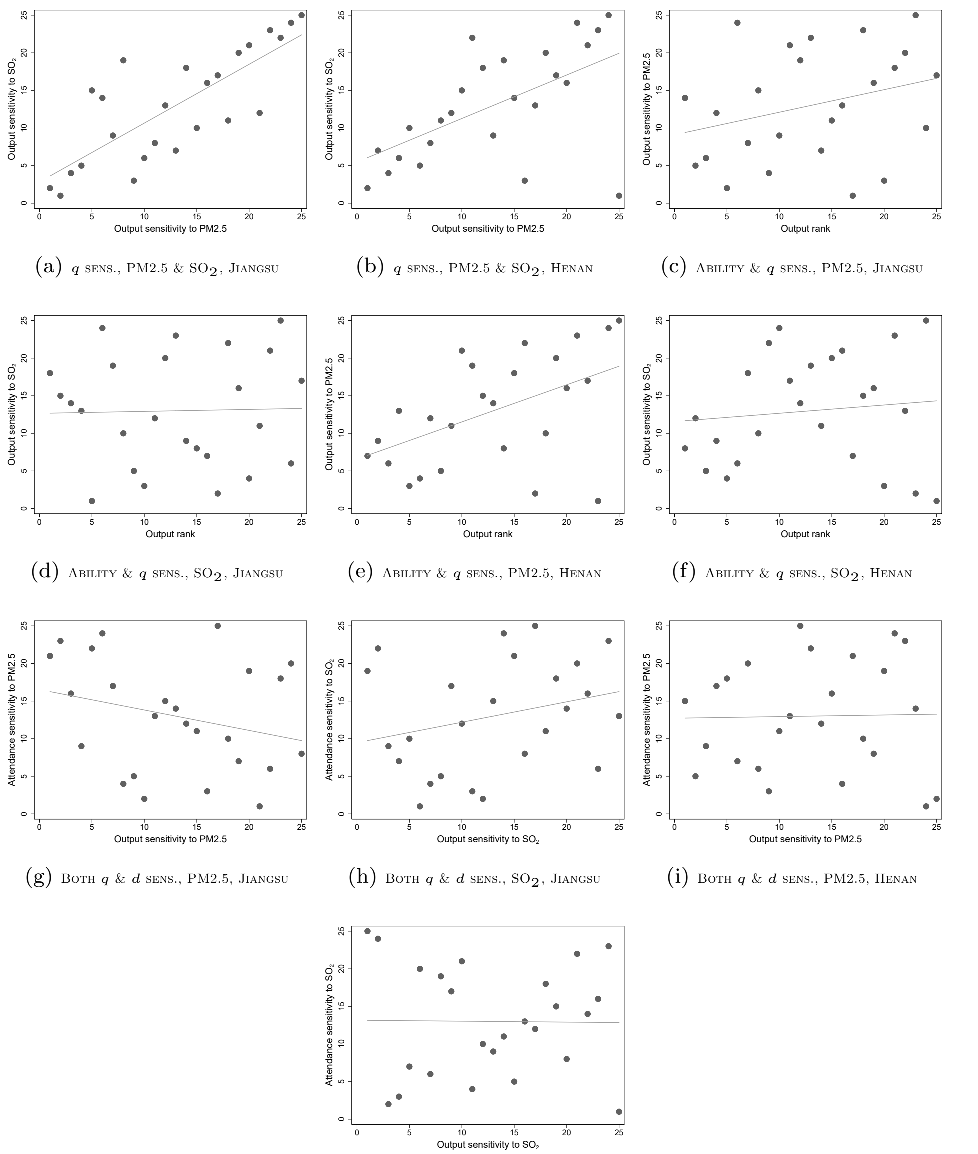
(k) Henan, SO_2, 2SLS, p25-p75



(l) Henan, SO_2, 2SLS, attendance

FIGURE A.4. : THE CUMULATIVE IMPACT OF EXPOSURE TO PM2.5 OR SO_2

Notes: Robustness tests of the cumulative impact on worker supply of $+10\ \mu g/m^3$ exposure to pollution (PM2.5 or SO_2) on each of the concurrent and previous P days. Panels in the first column report the **output effect estimated by OLS** (expressed as a proportion of mean output in the respective sample). Panels in the second column report **individual output effects** (25th, 50th and 75th percentiles). Panels in the third column report the **attendance effect** (in percentage points). In the two last columns, 2SLS uses $\hat{Z}$ to instrument for Z. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different quartic distributed lag model, PDL$(P, 4)$, as we raise P along the horizontal axis.



(a) q SENS., PM2.5 & SO_2, JIANGSU

(b) q SENS., PM2.5 & SO_2, HENAN

(c) ABILITY & q SENS., PM2.5, JIANGSU

(d) ABILITY & q SENS., SO_2, JIANGSU

(e) ABILITY & q SENS., PM2.5, HENAN

(f) ABILITY & q SENS., SO_2, HENAN

(g) BOTH q & d SENS., PM2.5, JIANGSU

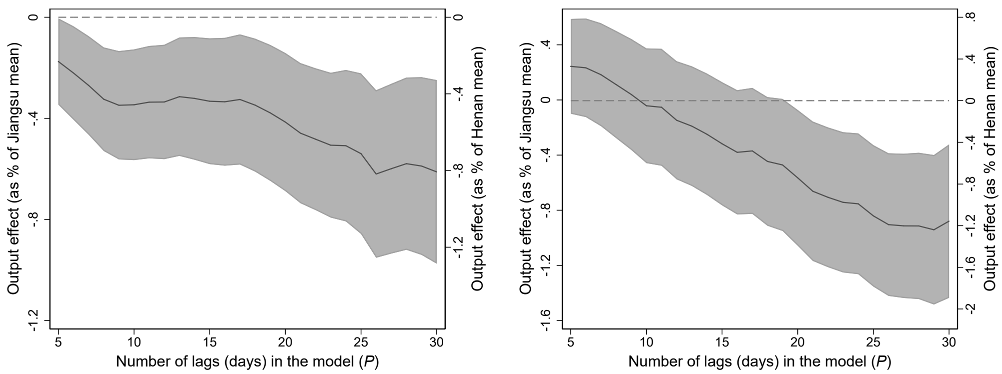
(h) BOTH q & d SENS., SO_2, JIANGSU

(i) BOTH q & d SENS., PM2.5, HENAN

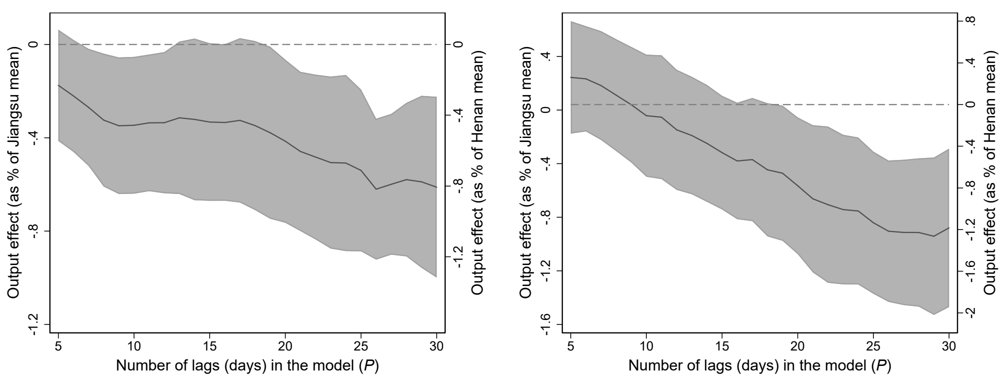
(j) BOTH q & d SENS., SO_2, HENAN

FIGURE A.5. : INDIVIDUAL HETEROGENEITY

Notes: In each panel, an observation is a worker in the sample of 25 workers with most output observations at the given work site. Workers are ranked (i.e., 1, 2, ..., 25) according to their mean output in the sample ("output rank"), and their output sensitivity ("q sens.") or their attendance sensitivity ("d sens.") to cumulative pollution exposure, according to 20-day quartic distributed lag models, PDL$(20, 4)$, estimated by 2SLS.



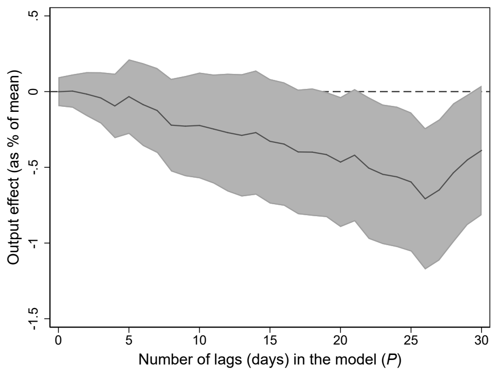
(a) PM2.5, two-way cluster (day, worker) (b) SO_2, two-way cluster (day, worker)



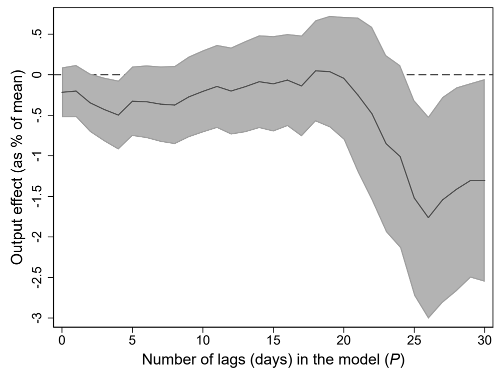
(c) PM2.5, cluster by day & 30-day autocorr. (d) SO_2, cluster by day & 30-day autocorr.

FIGURE A.6. : FORMS OF AUTOCORRELATION

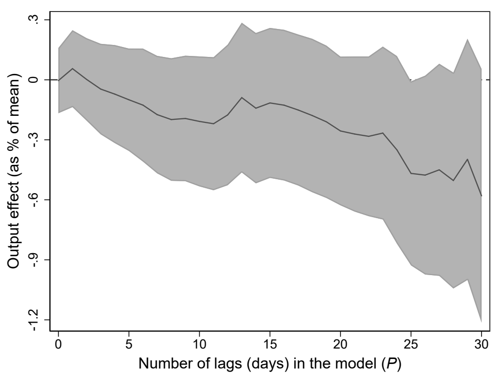
Notes: Robustness to allowing for alternative forms of autocorrelation in the analysis of prolonged pollution exposure. 2SLS output regressions implemented on the pooled sample that combines observations from both Jiangsu and Henan sites, with $\hat{Z}$ instrumenting for Z. As in Figure 7, we show the cumulative impact on output of $+10\ \mu g/m^3$ exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Panels (a) (PM2.5) and (b) (SO_2) compute standard errors using two-way clusters on day and on worker. Panels (c) (PM2.5) and (d) (SO_2) compute standard errors using day clusters (within-day correlation across workers) and an autocorrelation bandwidth of 30 days (correlation over time). Other notes to Figure 7 apply here.



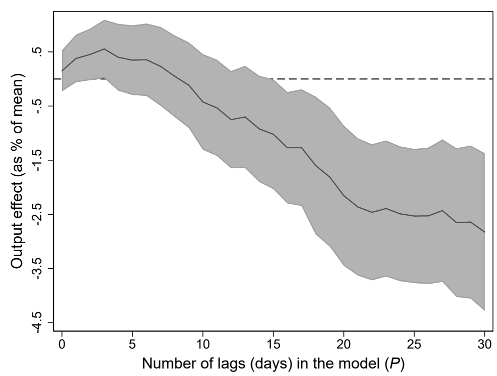
(a) PM2.5, Jiangsu



(b) SO_2, Jiangsu



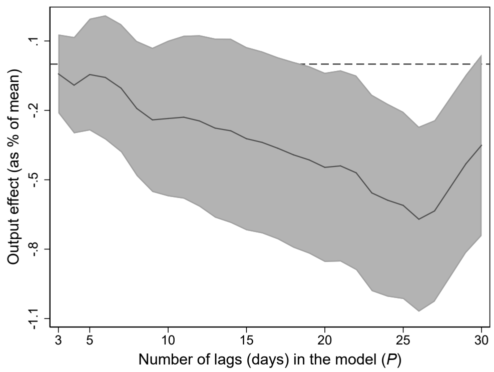
(c) PM2.5, Henan



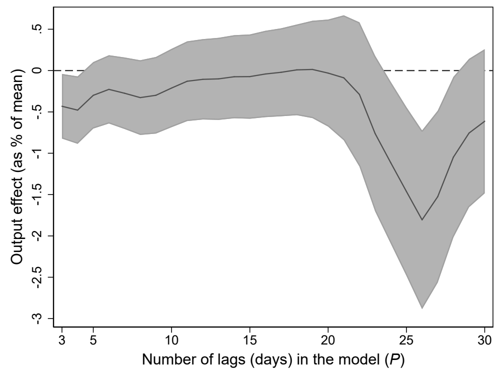
(d) SO_2, Henan

FIGURE A.7. : UNCONSTRAINED DISTRIBUTED LAG

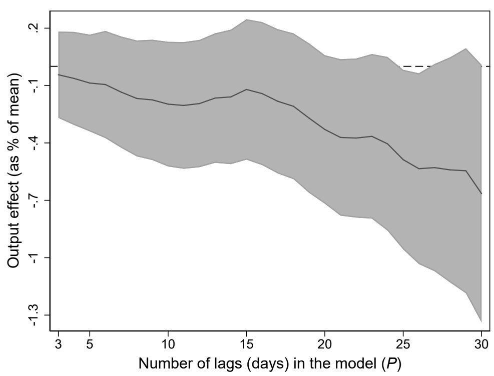
Notes: Robustness to removing the smoothness constraint on the lagged pollution coefficients, β_p: Unconstrained distributed lag, UDL(P). 2SLS output regressions implemented separately by site (full output sample with neighboring-city pollution in the case of Henan), using $\hat{Z}$ to instrument for Z. As in Figure 6(I)-6(II), we show the cumulative impact on output of $+10$ μg/m^3 exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different UDL(P) model, as we raise P along the horizontal axis. We report the effect on output as a proportion of mean output in the respective sample.



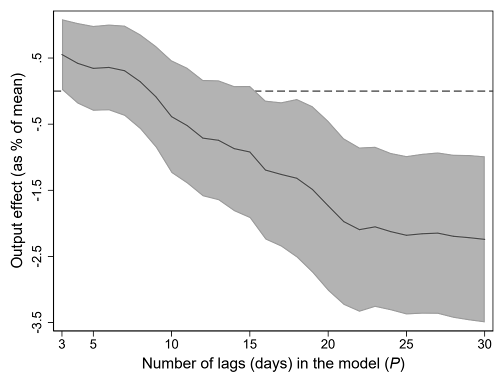
(a) PM2.5, Jiangsu



(b) SO_2, Jiangsu



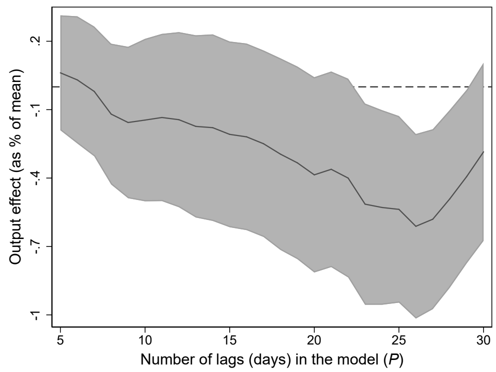
(c) PM2.5, Henan



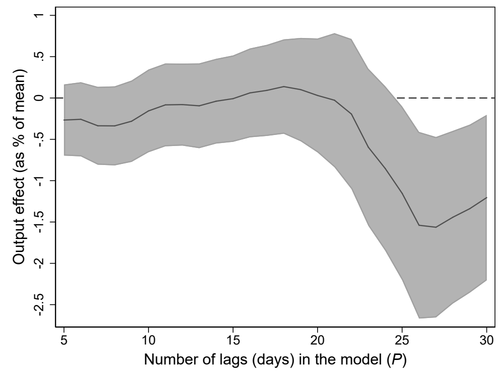
(d) SO_2, Henan

FIGURE A.8. : QUADRATIC DISTRIBUTED LAG

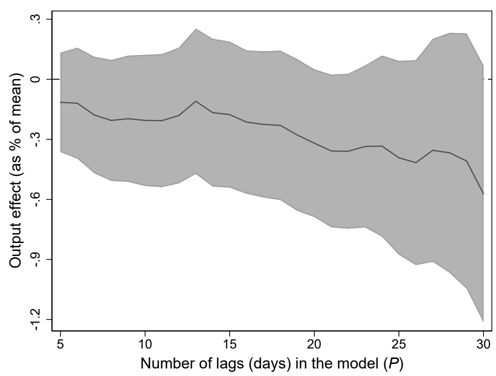
Notes: Robustness to constraining the lagged pollution coefficients, β_p, further: Quadratic distributed lag, PDL$(P, 2)$. 2SLS output regressions implemented separately by site (full output sample with neighboring-city pollution in the case of Henan), using $\hat{Z}$ to instrument for Z. As in Figure 6(I)-6(II), we show the cumulative impact on output of $+10\ \mu$g/m^3 exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different PDL$(P, 2)$ model, as we raise P along the horizontal axis. We report the effect on output as a proportion of mean output in the respective sample.



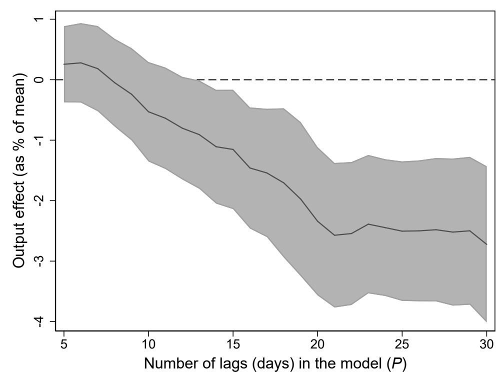
(a) PM2.5, Jiangsu



(b) SO_2, Jiangsu



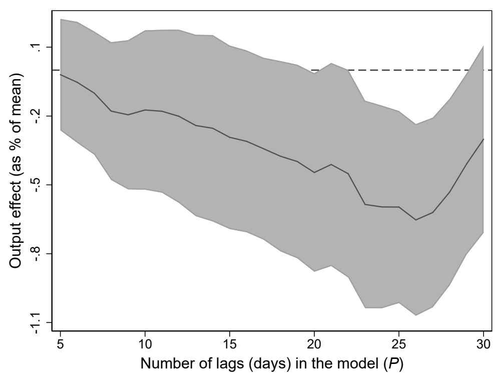
(c) PM2.5, Henan



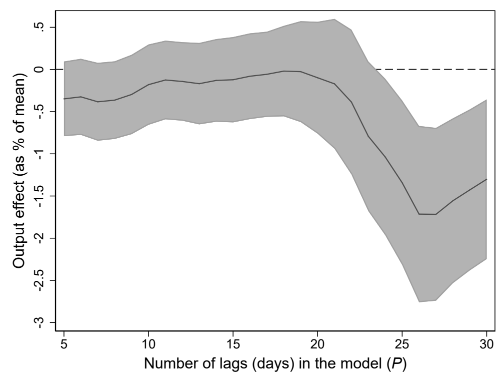
(d) SO_2, Henan

FIGURE A.9. : OUTPUT SAMPLE UNTRIMMED

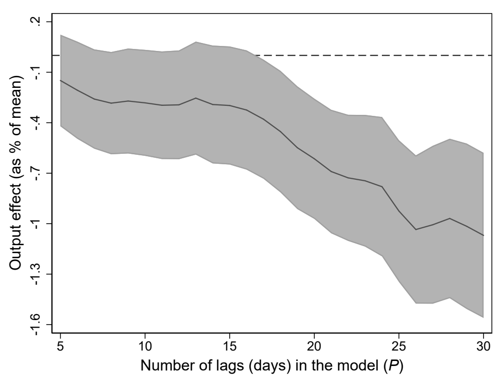
Notes: Robustness to not trimming the output sample at three standard deviations of the sample mean. 2SLS output regressions implemented separately by site (full output sample with neighboring-city pollution in the case of Henan), using $\hat{Z}$ to instrument for Z. As in Figure 6(I)-6(II), we show the cumulative impact on output of $+10$ μg/m^3 exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different quartic distributed lag model, PDL$(P, 4)$, as we raise P along the horizontal axis. We report the effect on output as a proportion of mean output in the respective sample.



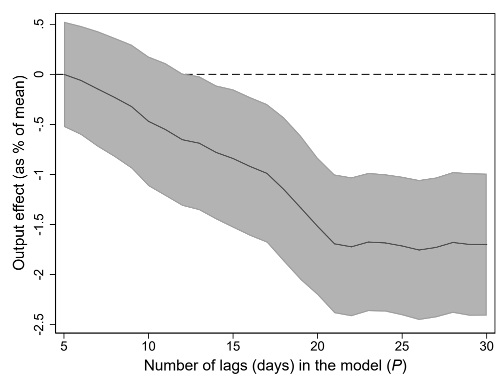
(a) PM2.5, Jiangsu



(b) SO_2, Jiangsu



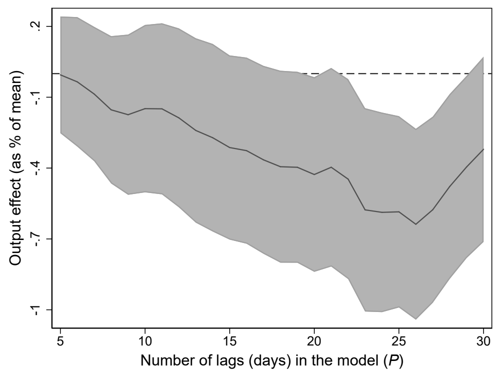
(c) PM2.5, Henan



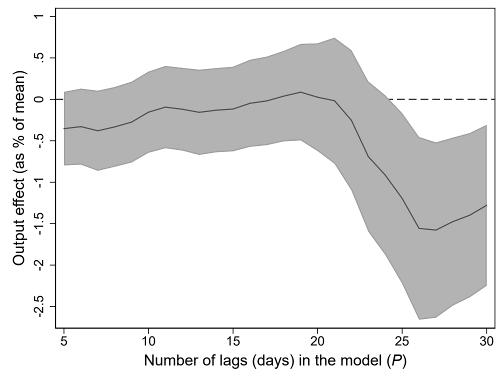
(d) SO_2, Henan

FIGURE A.10. : SEASONALITY CONTROLS

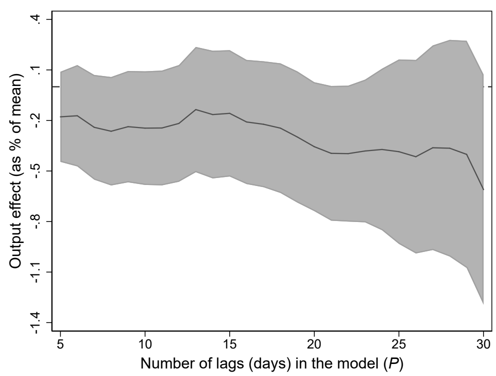
Notes: Robustness to controls for seasonality: Replacing year-month fixed effects by a quadratic time trend coupled with indicators for winter months, November to February, during which pollution tends to be most severe. 2SLS output regressions implemented separately by site (full output sample with neighboring-city pollution in the case of Henan), using $\hat{Z}$ to instrument for Z. As in Figure 6(I)-6(II), we show the cumulative impact on output of $+10\ \mu g/m^3$ exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different quartic distributed lag model, PDL$(P, 4)$, as we raise P along the horizontal axis. We report the effect on output as a proportion of mean output in the respective sample.



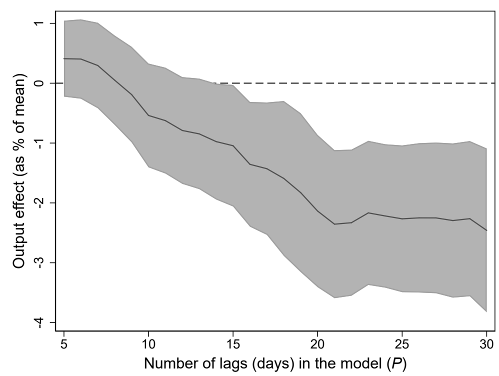
(a) PM2.5, Jiangsu



(b) SO_2, Jiangsu



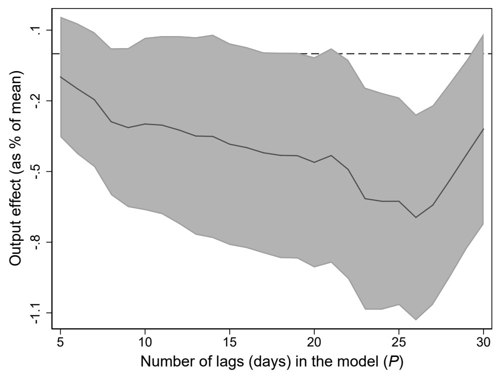
(c) PM2.5, Henan



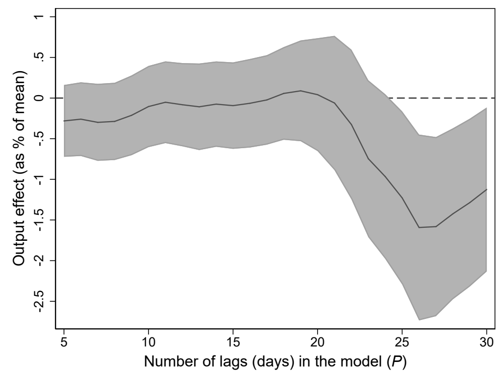
(d) SO_2, Henan

FIGURE A.11. : EXOGENOUS VENTILATION

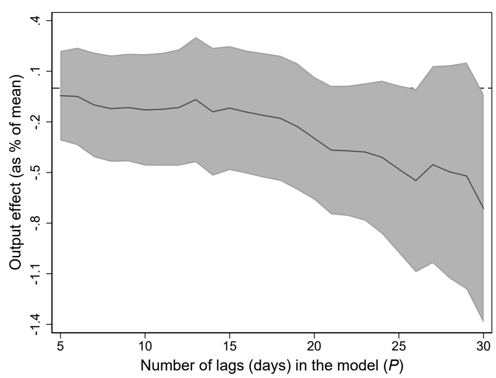
Notes: Robustness to the set of exogenous ventilation conditions, in which we drop wind direction from the ventilation conditions V that are used to form an instrument $\hat{Z}$ for measured pollution Z. 2SLS output regressions implemented separately by site (full output sample with neighboring-city pollution in the case of Henan). As in Figure 6(I)-6(II), we show the cumulative impact on output of $+10$ μg/m^3 exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different quartic distributed lag model, PDL$(P, 4)$, as we raise P along the horizontal axis. We report the effect on output as a proportion of mean output in the respective sample.



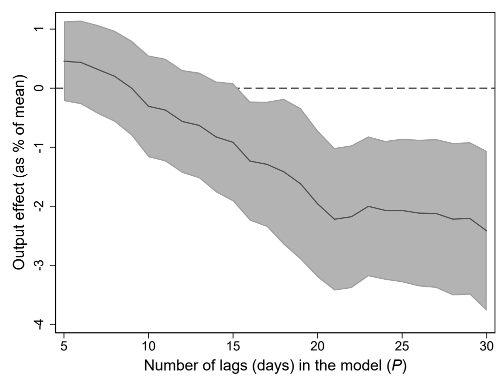
(a) PM2.5, Jiangsu



(b) SO_2, Jiangsu



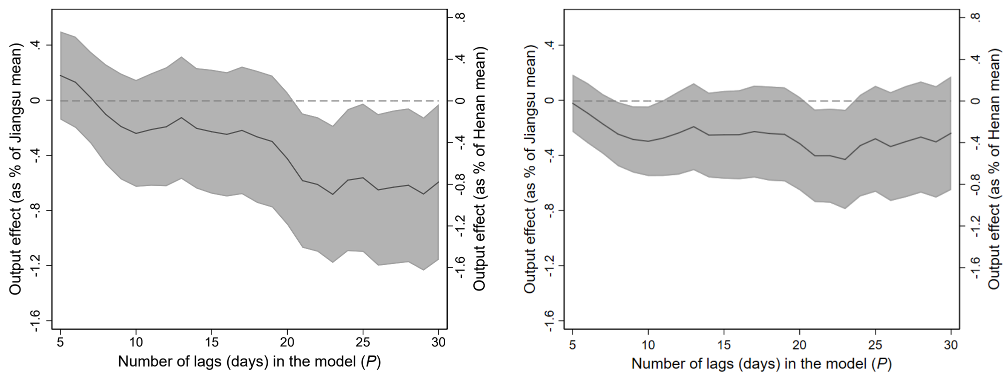
(c) PM2.5, Henan



(d) SO_2, Henan

FIGURE A.12. : THE LOGARITHM OF OUTPUT

Notes: Robustness to specifying the dependent variable, individual worker output, in logarithms instead of levels. 2SLS output regressions implemented separately by site (full output sample with neighboring-city pollution in the case of Henan), using $\hat{Z}$ to instrument for Z. As in Figure 6(I)-6(II), we show the cumulative impact on output of $+10$ μg/m^3 exposure to a pollutant (PM2.5 or SO_2) on each of the concurrent and previous P days. Each value reports the point estimate and 95% CI on $\sum_{p=0}^{P} \beta_p$ from a different quartic distributed lag model, PDL$(P, 4)$, as we raise P along the horizontal axis. We report the effect on output in log points ($\times 100$).



(a) Combined PM2.5 & SO_2: +10 & +10 μg/m^3 (b) Combined PM2.5 & SO_2: +10 & +5 μg/m^3

FIGURE A.13. : Multi-pollutant models

Notes: Multi-pollutant models (PM2.5 and SO_2) of more prolonged exposure and output. We implement each PDL$(P, 4)$ output regression on the pooled sample that combines observations from both Jiangsu and Henan sites, with $\hat{Z}$ (PM2.5 and SO_2) instrumenting for Z (both PM2.5 and SO_2 in the regression equation). Panel (a) shows the cumulative impact on output (point estimate and 95% CI) of combined doses of +10 μg/m^3 PM2.5 and +10 μg/m^3 SO_2 on each of the concurrent and previous P days. To reflect common emission sources and a mean PM2.5 to SO_2 concentration ratio of about 2, panel (b) shows the cumulative impact on output of combined doses of +10 μg/m^3 PM2.5 and +5 μg/m^3 SO_2 on each of the concurrent and previous P days. Other notes to Figure 7 (for single-pollutant models) apply here.