Data availability for MS AER-2010-0533.R2: The Impact of Pollution on Worker Productivity

Based on our contractual arrangement with the farm, the data on productivity used in this analysis must remain confidential, and thus can not be shared. Although the merged weather and pollution data are publicly available, we can not share this data either because the location of the monitors can potentially reveal the location of the farm. Hence, this description provides the steps one would need to take if they were granted permission to use the farm data. All processing and analysis was done in Stata.

1. Pollution and weather data

A. Download data

Both data sets were downloaded directly from The California Air Resources Board:

Weather: http://www.arb.ca.gov/aqmis2/metselect.php
Pollution: http://www.arb.ca.gov/aqmis2/aqdselect.php

After selecting the 'hourly data' tab, perform the following steps:
Step 1: Choose "select a parameter", and one-by-one choose the following parameters: ozone (ppb), particulate matter (pm 2.5), temperature, dew point, relative humidity, precipitation, pressure - atmospheric, solar radiation - solar, and wind - resultant (mph).
Step 2: Choose last day of month for "end date"
Step 3: Choose a location (can not be revealed)
Step 4: Choose "hourly data"
Click 'retrieve data'

To download data in bulk, perform the following steps:
Step 1: Click 'Pick Sites/Dates for Download' at bottom of display
Step 2: Get 'all data', change date range accordingly (choose a month or a quarter for a range, otherwise an error message may occur in Step 4 below), check sites of interest (can not be revealed)
Step 3: Use data for "only if checked" - this will generate CSV
Step 4: Click 'Complete data record' for data format
Step 5: Click "get data," which will download data in CSV format

B. Process data

i. Once part A is complete for all time periods and parameters, run pollYY.do and weatYY.do, where YY = {2009,2010}. This do-files reads in the CSV files for pollution and weather obtained above, and combines into one file for each year. These do-files will need to be modified to match the names of the CSV files.

ii. Run pollYYsum.do and weatYYsum.do to keep only valid measures and prepare the data as described in the manuscript (daily measures from 6am-3pm).

2. Productivity data

The productivity data from the farm was provided as a series of .xls files over irregular time intervals. Convert each .xls file into a .dta file using stat-transfer. Then run 'chkmMMDDYY.do', where the date corresponds to when files were received from the farm. These do-files perform basic cleaning of the series of files for each date, and then combines them into one usable file. Enclosed is a representative do-file (chkm070210.do), which would need to be slightly altered depending on the exact names of the files obtained from the farms. After each 'chkm' do-file is run, 'chkm_combine.do' combines each of the separate files.

3. Merging data

Both the productivity and environment data sets are at the daily level, so the merge was straightforward using the 'mergepw.do.' This gives the final data set for analysis: 'mergepw.dta.'

4. Analysis

i. Prior to running the analysis, 'prep4reg.do' does the final setup for regressions by creating dummy variables and final variable definitions. Monitor locations are replaced with XXXX.

ii. Summary statistics for Table 1 and data for Figure 4 are in 'sumstats.do,' labeled within the do-file. Figure 4 was created within Excel.

iii. 'figures1_3.do' creates graphs for Figures 1-3 and figure5.do for Figure 5.

iv. The do-files for the regression analyses are named with the same name as the corresponding table in the manuscript. E.g., table2.do produces the results for Table 2. Letters following the table number indicate they are for different columns within the Table, with the first line of the do-file indicating which columns. Some of the programs call ado files, which are also enclosed; these were created by other individuals, with details given at the top of each file. Users must install 'outreg2' before running these do-files (type "ssc install outreg2" at command line in stata).