Data and Code for:

"Air Pollution and the Labor Market: Evidence from Wildfire Smoke"

By Mark Borgschulte, David Molitor, and Eric Yongchen Zou

July 2022

This document describes how to replicate the results in this paper. Data and scripts are organized into the "1_build" folder (workhorse data creation) and the "2_analysis" folder (tables and figures replication). The following software is required:

- STATA (scripts last run with version MP16.0)
  - User-contributed packages: ftools, gtools, ingap, reghdfe, regsave, sortobs, wyoung
- R (scripts last run with version 4.1.2)
  - Libraries: tidyverse, haven, sf, rmapshaper, tmap
- A system environment variable (or a STATA global macro) "Rep_smokelabor" that defines the location of the replication folder
- Total computation time is about 48 hours on Windows 10, 16 GB, Intel Core i7-8850H @ 2.60Ghz (6 processors).

1. The "1_build" folder

The following schematic summarizes the structure of "1_build". Each subfolder of "1_build" corresponds to a data source.

```
1_build/
├── source1/
├── source2/
│   ├── raw/                    ** raw data
│   ├── scripts/                ** scripts that process the raw data
│   │   ├── build_source2.do
│   │   └── …
│   └── proc/                   ** processed data
└── source3/
```

Below we provide an overview of each data source. Step-by-step instructions are included in the .do files. Several steps may take a long time to run, which are indicated by (🕓) below. Interim output files are provided where possible. The scripts need to be run in the order in which they are listed below.

1.1. qwi

- Earnings and employment data
- Citation: United States Census Bureau. 2007-2019. "Quarterly Workforce Indicators. Collection "sa_f_gc_ns_oslp_u." https://lehd.ces.census.gov/data/qwi/R2020Q4 [accessed November 12, 2020]; "Quarterly Workforce Indicators. Collection sa_f_gc_n3_oslp_u." https://lehd.ces.census.gov/data/qwi/ R2021Q2 [accessed March 29, 2022];
- Collection "sa_f_gc_ns_oslp_u" data contain county-gender-age-naics2-year-quarter level observations; Collection "sa_f_gc_n3_oslp_u" data contain county-gender-age-naics3-year-quarter level observations
- A copy of the raw data is provided in "$Rep_smokelabor/1_build/raw"; decompress before executing "scripts/build_qwi.do"
- "scripts/build_qwi.do":
    - Create county-quarter level earnings and employment variables
    - Create county-quarter-age group level earnings and employment variables
    - Create county-quarter-naics2 industry level earnings and employment variables
    - Create county-quarter-naics3 industry level employment variable (🕓)

1.2. lau

- Labor force participation data
- Citation: U.S. Bureau of Labor Statistics. 2007-2019. "Local Area Unemployment Statistics" https://download.bls.gov/pub/time.series/la [accessed September 16, 2021]
- A copy of the raw data is provided in "$Rep_smokelabor/1_build/lau/raw/la.data.zip"; decompress before executing "scripts/build_lau.do"
- "scripts/build_lau.do":
    - Create county-quarter level labor force participation counts

1.3. seer

- Population counts data
- Citation: U.S. County Population Data by The Surveillance, Epidemiology, and End (SEER) Results Program. 1969-2019. "County-Level Population Files - Single-year Age Groups. All States Combined (adjusted), 1969-2019 White, Black, Other" https://seer.cancer.gov/popdata/download.html [accessed February 26, 2021]
- A copy of the raw data is provided in "$Rep_smokelabor/1_build/seer/raw/us.1969_2019.singleages.adjusted.exe"; decompress before executing "scripts/build_seer.do"
- "scripts/build_seer.do":
    - Create county-year level population counts
    - Create county-year-age level population counts

1.4. irs

- County in- and out-migration data

- Citation: Internal Revenue Service. SOI Tax Stats - Migration Data. 2007-2019. "U.S. Population Migration Data. County-to-County Migration Data" https://www.irs.gov/statistics/soi-tax-stats-migration-data [accessed November 13, 2021]
- A copy of the raw data is provided in "$Rep_smokelabor/1_build/irs/raw"; decompress before executing "scripts/build_irs.do"
- "scripts/build_irs.do":
    - Create county-year level in- and out-migration variables

1.5. hms

- Wildfire smoke exposure data
- Citation: National Oceanic and Atmospheric Administration. Office of Satellite and Product Operations. Hazard Mapping System. 2007-2019. "Hazard Mapping System Fire and Smoke Product." https://satepsanone.nesdis.noaa.gov/pub/volcano/FIRE/HMS_ARCHIVE/ [accessed October 19, 2021]
- Availability: Raw smoke shapefiles can be downloaded from the link above. The following files include processed data:
  - "1_build/hms/raw/county_smoke_quarter.dta": county-quarter smoke exposure days
  - "1_build/regdata/county_day.dta": county-day smoke exposure indicators
  These data are imported from a companion working project: Miller, Nolan, David Molitor, and Eric Zou. 2022. A Causal Concentration-Response Function for Air Pollution: Evidence from Wildfire Smoke.

1.6. aqs

- Air pollution data
- Citation: U.S. Environmental Protection Agency Air Data. 2007-2019. "Daily Summary Data 2007-2019." https://aqs.epa.gov/aqsweb/airdata/download_files.html. [accessed January 13, 2021]
- Availability: Raw data can be downloaded from the link above. The following files include processed data:
  - "1_build/hms/raw/county_smoke_quarter.dta": county-quarter pollution averages
  - "1_build/regdata/county_day.dta": county-day pollution averages
  These data are imported from a companion working project: Miller, Nolan, David Molitor, and Eric Zou. 2022. A Causal Concentration-Response Function for Air Pollution: Evidence from Wildfire Smoke.

1.7. weather

- Temperature, precipitation, and wind direction data
- Weather variables are built from the following two sources:
- Global Historical Climatology Network (GHCN) temperature and precipitation data
  Citation: Menne, Matthew J., Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, Ron Ray, Russell S. Vose, Byron E.Gleason, and Tamara G. Houston. National Oceanic and Atmospheric Administration National Centers for

Environmental Information. 2007-2019. “Global Historical Climatology Network”. ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year [accessed January 13, 2021]

- North American Regional Reanalysis (NARR) wind speed and direction data
  Citation: National Centers for Environmental Prediction. 2007-2019. “North American Regional Reanalysis U-wind at 10 m and V-wind at 10 m.” ftp://ftp.cdc.noaa.gov/Datasets/NARR/Dailies/monolevel. [accessed November 14, 2021]
  Publication citation: Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, and Coauthors, 2006: North American Regional Reanalysis. Bulletin of the American Meteorological Society, 87, 343–360, doi:10.1175/BAMS-87-3-343
- Availability: Raw station and grid level data can be downloaded from the links above. The following files include processed data:
  - “1_build/weather/raw/narr_county_quarterly.dta”: county-quarter wind direction bins
  - “1_build/hms/raw/county_smoke_quarter.dta”: county-quarter temperature and precipitation bins
  These data are imported from a companion working project: Miller, Nolan, David Molitor, and Eric Zou. 2022. A Causal Concentration-Response Function for Air Pollution: Evidence from Wildfire Smoke.

1.8. geo

- This folder contains Census administrative boundary data used in the steps above.
- county_2010: 2010 Census county shapefiles
  - Citation: United States Census Bureau. 2010. “TIGER/Line Geodatabases County TIGER/Line shapefiles tl_2010_us_county10.” https://www2.census.gov/geo/tiger/TIGER2010/COUNTY/2010/tl_2010_us_county10.zip [accessed July 2, 2016]
  - Data file: “/geo/county_2010/cnty_cen2010.shp”
- tiger_simple: Cartographic Boundary Shapefiles (simplified representations of selected geographic areas from the Census Bureau’s MAF/TIGER geographic database; we use these files for plotting maps)
  - Citation: U.S. Census Bureau. 2010. “Cartographic Boundary Files - Shapefile” https://www2.census.gov/geo/tiger/GENZ2010/gz_2010_us_050_00_500k.zip [accessed November 15, 2018]
- ls_county.dta: dta version of county list created from county_2010
- state_to_region_division_cw.dta: State to Census Division and Census Region crosswalk file created from “https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf” [accessed July 2, 2016]

1.9. regdata

- This folder contains scripts that combine individual data sources to create regression-ready data files.
- “scripts/build_regdata.do”:
  - Create county-quarter level main regression file
  - Create county-quarter-age group level regression file
  - Create county-quarter-industry group level regression file
- Copies of the output files are included in the folder

2. The “2_analysis” folder

Scripts in this folder take input data from “1_build”, conduct analysis, and produce tables and figures. Each .do file corresponds to one table or figure of the paper, as indicated by the file name. The scripts can be run in any order. Output tables are stored in subfolder “output_tables”, output figures are stored in subfolder “output_figures”, and logfiles are stored in subfolder “output_logs”. The following schematic summarizes the structure of “2_analysis”. Processing time is long for “figure5.do” which involves the resampling method of Westfall and Young (1993).

```
2_analysis/
├── figure1.do
├── figure2.do              ** scripts that replicate a figure
├── output_figures/
│   ├── figure2.pdf         ** output figure
│   └── …
├── output_logs/
│   ├── figure2.log         ** logfile
│   └── …
└── figure3.do
```