Documentation on data and code

main.xlsx

This Excel file contains the main calculations that lead to the final results of the paper, including all the figures. Each tab is described below.

Tab name	Description
GDP	Present the US GDP up to 2019 that comes from the World Bank.
Population	Present the US population up to 2019 that comes from the World Bank.
Income_pop_share	Present the share of population and aggregate income by urban and rural in selected years.
Income_data_process	Obtain per capita income by urban and rural.
Expected_MD	Present the expected marginal damages in 2008 by urban and rural.
MD_results	Obtain the marginal damages by urban and rural.
Emission_results	Present the emissions by urban and rural.
GED_$2019	Obtain the pollutant damages in 2019 dollar.
Sensitivity	Conduct sensitivity analysis with various values of income elasticity.

Other tabs are self-contained, including data input, calculation, and then result output, while **income_pop_share, Expected_MD** and **Emission_results** contain intermediate results which will be described below.

List of counties designated as urban from the Census Metropolitan Statistical Areas data

In order to partition emissions, population, and aggregate income into urban and rural areas, this paper uses Metropolitan Statistical Areas (MSA) from the US Census to designate urban counties. This work is inside the directory - .**msa**.

- **.\data_input**
  - These are the original data files from the US Census. Each file contains the FIPS for the counties designated as MSA for a given year.
- Code_export_msa_fips.do
  - This Stata script is used to extract MSA FIPS from the original data files.
- .**\result_output**
  - Each file contains the FIPS for the counties designated as MSA for a given year.

Expected marginal damages in 2008 by urban and rural

The expected marginal damages are emission-weighted marginal damages, based on the results from 2008 APEEP. This work is inside the directory – .**expected_marginal_damages**. The APEEP results used for this study is the PM mortality health effects for criteria pollutants, produced by the 2008 version of APEEP. The results are included here, while the APEEP code can be obtained at Prof. Nick Muller's website.

- **.\data_input**
  - .\APEEP_results_output: The results from 2008 APEEP calculation for 3109 counties in the US, including emissions, marginal damages, and damages.
  - AP2_Fips_List_by_source.csv: The FIPS list matching the APEEP results for each row, and to be merged with the APEEP results.
  - data_msa_fips_2010.dta: The Metropolitan Statistical Areas FIPS list from 2010 to be used to designate urban or rural for the 2008 results.
- **.\code**
  - Sort the Stata scripts by name, and execute them in order. Need to set the root directory for each code.
  - The code together obtains the emission-weighted marginal damages by urban and rural areas for each criteria pollutant.
- **.\result_output**
  - results_2008_final_msa.csv: Contains the emission-weighted marginal damages by urban and rural areas for each criteria pollutant.

Urban/rural emissions

The work to obtain emissions by urban and rural areas from 1970 to 2019 takes two steps, which are included in the directory -**.\emissions**.

Step 1

Obtain urban and rural emissions for the selected years in which county-level emissions are available at the National Emission Inventory (NEI).

- **.\data_input**
  - This folder contains original data files on county-level emission from the USEPA NEI. Some are at the tier 1 level, while others are at the tier 2 level.
  - Also included are the MSA FIPS data files, to be used for partition.
- **.\data_intermediate**
  - The csv files are the county-level emission data extracted from the original data files.
- **.\code**
  - The Stata scripts **emission_county_level_[year].do** are used to obtain county-level emissions by pollutant.
  - The Stata script **code_emission_by_urban_rural.do** is used to partition emissions into urban and rural areas using the MSA designation.
- **.\result_output**
  - Each file contains urban and rural emissions by pollutant for a given year.

Step 2

We obtain urban/rural share of emissions for the selected years and then obtain urban/rural share of emission for 1970 to 2019 through interpolation. We then obtain national emissions for 1970 to 2019. Finally, we multiply the urban/rural share with the national emissions to obtain the urban and rural emissions. This work is done in the Excel file – **emission_partition.xlsx**. Each tab is further described below.

- **Emission_partition.xlsx**
  - Urban_rural_emisisons: The result output from Step 1 is gathered here.
  - Emission share: The urban emission share is calculated.
  - NEI_national: The national emissions for the criteria pollutants are gathered here. The original data file is **national_tier1_caps_1970_2019.xlsx**.
  - NH3_T1: Obtain NH3 emissions for 1970 to 1989 through extrapolation by tier 1 emissions. The coefficients for extrapolation are displayed on the figures by tier 1 category.
  - PM25_T1: Obtain PM2.5 emissions for 1970 to 1989 through extrapolation by tier 1 emissions. The coefficients for extrapolation are displayed on the figures by tier 1 category.
  - Emission_results: Obtain the urban and rural emissions for 1970 to 2019.

Urban/rural share of population and aggregate income

We obtain urban/rural share of population and aggregate income for the selected years based on census and American Community Survey (ACS). The work is inside the directory - .**income_pop**.

- **.\data_input**
  - This folder contains the census and ACS 5-year data, obtained from SocialExplorer.
  - Also included are the MSA FIPS data, to be used for partition.
- **.\data_intermediate**
  - This folder contains the datasets for population and aggregate income, extracted from the original data files.
- **.\code**
  - This Stata script obtains datasets containing population and per capita income by urban and rural, for 70, 80, 90, 00, 2010, and 2016.
- **.\result_output**
  - This folder contains the population and per capita income by urban and rural.