Replication Materials for "Does Environmental Regulation Matter for Income Inequality? New Evidence from Chinese Communities"

Bihong Huang
International Monetary Fund

Ying Yao
Asian Growth Research Institute

January 23, 2023

This compressed file folder contains Stata and R codes for data construction and analysis. The codes were written in Stata version 17.0 and R version 4.1.0. Note that this folder only includes datasets for generating tables and figures, as the original datasets either require registration or are very large. This document describes how to replicate Tables 1–6 and Figures 1–7 from the paper.

1 Data Source and Construction

The paper uses data from two sources: the China Health and Nutrition Survey (CHNS) and the Emission Database for Global Atmospheric Research (EDGAR). It also uses data from CEIC to demonstrate the background information.

1.1 CHNS Data

(a) Survey Data
Fill out a registration form at the following link to obtain the original survey datasets.
https://www.cpc.unc.edu/projects/china/data/datasets/data-downloads-registration

(b) Location of communities
The Stata file entitled "community_location.dta" contains community locations extracted from Ge (1998), coordinates generated by reverse geocoding, a list of TCZ cities compiled from the government documents, and a list of focal cities from the

China Environment Statistics Yearbook (1993–2015).

(c) Run "chns_construct.do" to obtain individual- and community-level datasets.

The following table lists the output from running the do file.

Data Level	Output	Description
Community	urban_index.dta	Community urbanicity index
	gini_hh.dta	Community income distribution: gini
	incpct_hh.dta	Community income distribution: percentiles
	incdistribution.dta	Merge the above three datasets
Household	hhinc_source.dta	Income sources
Individual	indinc_analysis.dta	Individual income with other socioeconomic factors

1.2 EDGAR v5.0 Global Air Pollutant Emissions

The emission data can be retrieved from

Crippa, Monica; Guizzardi, Diego; Muntean, Marilena; Schaaf, Edwin; Oreggioni, Gabriel (2019): EDGAR v5.0 Global Air Pollutant Emissions. European Commission, Joint Research Centre (JRC) [Dataset]
PID: http://data.europa.eu/89h/377801af-b094-4943-8fdc-f79a7c0c2d19

You can also find the data at https://edgar.jrc.ec.europa.eu/dataset_ap50.

The database was recently updated to EDGAR v6.1 (https://edgar.jrc.ec.europa.eu/dataset_ap61), covering the years from 1970–2018.

(a) Download SO_2 emission time series (1970–2015) by country.
https://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/v50_AP/SO2/v50_SO2_1970_2015.zip

(b) Download annual total SO_2 emissions gridmaps (1970–2015).
https://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/v50_AP/SO2/TOTALS/TOTALS_nc.zip

(c) Download sectoral SO_2 emissions gridmaps (1970–2015): manufacturing sector.
https://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/v50_AP/SO2/IND/IND_nc.zip

(d) Download sectoral SO_2 emissions gridmaps (1970–2015): power sector.
https://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/v50_AP/SO2/ENE/ENE_nc.zip

(e) Unzip the downloaded files to the local machine and run "netCDF_Convert.R" to convert .nc files to .dta files. The generated files only contain China's grid.

(f) Run "edgar_chns_so2.do" to construct SO_2 emissions in China (1970–2015) and match EDGAR emissions data to CHNS communities.

The following table lists the output from running the do file.

Data Source	Output	Description
Emission time series	so2_country_sector.dta	Country sectoral SO_2 emissions
	so2_country_total.dta	Country total SO_2 emissions
	so2_country_combine.dta	Append the above two datasets
Gridmaps	so2_total_wide.dta	Annual total SO_2 emissions
	so2_power_wide.dta	Sectoral SO_2 emissions: power sector
	so2_manu_wide.dta	Sectoral SO_2 emissions: manufacturing sector
	so2_edgar_combine_wide.dta	Merge the above three datasets
	so2_match_wide.dta	SO_2 Emissions for CHNS communities
	so2_match_long.dta	SO_2 Emissions for CHNS communities (long format)

1.3 CEIC Data

This database requires a subscription, so the authors removed the data from the .dta files used to generate the figures. However, we list the names of these series here for your reference.

(a) CN: City Labor Market: Demand: Person in Charge

(b) CN: City Labor Market: Demand: Technical Professional

(c) CN: City Labor Market: Demand: Clerical Staff & Related

(d) CN: City Labor Market: Demand: Service & Attendant

(e) CN: City Labor Market: Demand: Agricultural Personnel

(f) CN: City Labor Market: Demand: Equipment Operator

(g) CN: City Labor Market: Demand: Manufacturing

(h) CN: City Labor Market: Demand: Electricity, Gas & Water Production & Supply

(i) CN: No of Employee: State Owned: Manufacturing

(j) CN: No of Employee: State Owned: Electricity, Gas & Water Production & Supply

(k) CN: No of Employee: Private & Individual: Manufacturing

- Series (a)–(f) are used to construct "laborbyoccupation.dta", which shows labor market demand by occupation.
- Series (g)–(h) are used to construct "laborbysector.dta", which shows labor market demand by sector.
- Series (i)–(k) are used to construct "laborbysector_number.dta", which shows the number of employees by sector.

2 Replication

Run a do file named "ExcecuteAll.do" which compiles the individual do files used to generate the tables and figures. Before running this compilation file, you need to set up your working directory.

This compilation file includes the following individual do files:

Do file	Description
main_data.do	Obtain the primary data set for analysis
fig1_edgar_chns.do	Plot trends in SO_2 emission in China using two data sources
fig2_ginitrend.do	Plot trends in income inequality in China using CHNS data
fig3_eventgini.do	Visualize the coefficients generated in the event study
fig4_spillover.do	Explore the presence of spillover effects
fig5_percentiles.do	Explore the policy effect on different income percentiles
fig6_wage.do	Investigate wage differences and variations by occupation
fig7_labor_market.do	Demonstrate the labor demand in the city labor market
tab1_summary.do	Generate a summary of statistics for communities
tab2_baseline.do	Obtain baseline results
tab3_robustness.do	Obtain results of robustness checks
tab4_heterogeneity.do	Investigate the heterogeneity of policy effect
tab5_decomposition.do	Identity sources of income (in)equality
tab6_gini_commtype.do	Identify the role of manufacturing in community income distribution

3 Stata Packages

In addition to built-in Stata commands, the authors employ user-written packages to create indices, figures, and tables. The packages used are listed as follows:

(a) Income distribution indices are generated by `ineqdeco` (Jenkins, 1999) and `ineqdec0` (Jenkins, 1999)
Jenkins, Stephen P. 1999. INEQDECO: Stata module to calculate inequality indices

with decomposition by subgroup, Statistical Software Components S366002, Boston College Department of Economics, revised 15 Feb 2021.

Jenkins, Stephen P. 1999. INEQDEC0: Stata module to calculate inequality indices with decomposition by subgroup. Statistical Software Components S366007, Boston College Department of Economics, revised 15 Feb 2021.

(b) Tables are produced by `estout` (Jann 2005, 2007).
Jann, Ben. 2005. Making regression tables from stored estimates. Stata Journal, 5(3): 288-308.

Jann, Ben. 2007. Making regression tables simplified, Stata Journal, 7(2): 227-244.

(c) Coefficient plots are produced by `coefplot` (Jann, 2014).
Jann, Ben. 2014. Plotting Regression Coefficients and other Estimates. Stata Journal, 14(4): 708–737.

(d) Event study plots are produced by `eventdd` (Clarke and Tapia-Schythe, 2021).
Clarke, Damian, and Kathya Tapia-Schythe (2021). Implementing the panel event study. The Stata Journal, 21(4), 853–884.