

1 Introduction

The following document accompanies the replication files for Greenstone and Hanna (2014). All of the files needed to replicate the paper's results are to be found in "Greenstone and Hanna (2014) Replication Files - 01102014.zip." This package includes raw data, data cleaning and analysis .do files (for use with Stata), as well as copies of the tables and figures as they appear in the paper.

The following document consists of two sections. The first section will provide important instructions for working with the enclosed files. The second section will summarize the purpose of each .do file contained within.

2 Replication Instructions

Begin by decompressing (unzipping) "Greenstone and Hanna (2014) Replication Files - 01102014.zip." Most recent operating systems have a native utility for this; for older operating systems, you may need to download a decompression utility. Save the unzipped folder to a location that is convenient to you. You may rename the directory anything you like, but do not rename any of the subfolders, and do not rearrange the file structure. The folder is structured to enable replication with the click of a button.

The remaining steps will allow you to run the data cleaning and analysis .do files. From the root directory ("Greenstone and Hanna (2014) Replication Files - 01102014," unless you have renamed it), navigate to **/Greenstone and Hanna (2014) Replication Files - 01102014/Do Files/india_pollution_master_01102014.do**. Open this file with a text editor. Line 31 reads:

```
global mypath "/Users/jbpetkun/Dropbox/India Pollution/Greenstone and Hanna (2014)
Replication Files - 01102014/"
```

This line uses a global macro to set the working directory. Replace everything between the quotation marks with the location of your root directory.

Lastly, one of the .do files uses two user-generated commands, so you will need to download the accompanying .ado file. In order to do so, navigate to [textbf/Greenstone and Hanna \(2014\) Replication Files - 01102014/Save to Ado Folder](http://www.stata.com/support/faqs/programming/personal-ado-directory/). You should see that this folder contains two .ado files: **qlrmean.ado** and **qlrtrend.ado**. Copy both files to your clipboard. Navigate to the .ado directory for your installation of Stata. If you have trouble locating, you can find additional information here: <http://www.stata.com/support/faqs/programming/personal-ado-directory/>. Save both of the provided .ado files to this directory. If you find that you are unable to run "Table5_Figure7_01102014.do," then you probably do not have the .ado files saved properly.

After setting the working directory and saving the .ado files, you will be ready to run the files. The "master" program is **india_pollution_master_01102014.do**. Simply run this program, and it will automatically execute all of the data cleaning and analysis .do files. Should you wish to run another program independently, simply set the working directory (using the "mypath" macro) within the desired program.

3 Catalog of .do Files

The remaining section will briefly summarize each of the .do files. The .do files fall into two categories: data cleaning and data analysis.

3.1 Data Cleaning

1. **airpol04_in_01102014.do**

The objective of this program is to insheet and clean raw 2004 air pollution data so it can be added to the complete air pollution dataset.

2. **airpol0507_in_01102014.do**

The objective of this program is to insheet monitor-year raw pollution data from 2005-2007, taken from the CPCB website, and clean it in preparation for merging with the complete air pollution dataset.

3. **ap_clean_01102014.do**

This file creates both city-year and monitor-year databases with air pollution data from 1987-2003, having variables such as mean, max, min for three pollutants: SPM, SO₂, and NO₂. These data are cleaned in preparation for merging with the complete air pollution dataset.

4. **census_in_01102014.do**

This file merges the University of Maryland's Indian District Database, which contains census data from 1981 and 1991, with the 2001 Census. These census data are later used as controls in the regression analysis.

5. **clean_districts_01102014.do**

The objective of this program is to standardize Indian district names in accordance with the 1981 district names and borders.

6. **clean_districts_1961_01102014.do**

This file cleans district names from the raw air temperature/precipitation dataset (**temp'prcp.dta**).

7. **clean_names_01102014.do**

This file cleans location names: states, districts, cities, location descriptions.

8. **clean_waternames_01102014.do**

This file cleans and reconciles conflicting river names and corrects misspelt city and river names.

9. **controls_merge_01102014.do**

This program's objective is to merge the final India pollution/mortality data with additional additional state*year controls, including newspaper references to pollution and corruption, industry compliance, and corruption index scores. These data were prepared in newspaper'prep.do, industry'compliance'prep.do, and corruption'prep.do, respectively.

10. **corruption_prep_01102014.do**

This program's objective is to prepare state corruption scores from Transparency International, and to merge those new data with the existing India Pollution datasets.

11. **im_addpolicy_01102014.do**

The objective of this program is to merge relevant air and water policy variables with the infant mortality data.

12. **im_city_data_01102014.do**

This file creates a 1987-2004 month-by-city database of mortality variables including birth, death, infantdeath, maternaldeath, and stillborn. As underlying data sources it uses 1995-2004 data Rema and RAs collected from each state of India; and from several year-specific files obtained from books in the Berkeley library (the years are 87, 89, 90, 92, 94, 95).

13. **industry_compliance_prep_01102014.do**

This program's objective is to prepare data representing industrial compliance, in Indian states, and to merge those new data with the existing India Pollution datasets.

14. **merge_ap_policy_01102014.do**

This file merges air pollution data and policy data and adds income controls.

15. newspaper_prep_01102014.do

This program's objective is to prepare data representing Times of India keyword search results, and to merge those new data with the existing India Pollution datasets.

16. policy_clean_01102014.do

This program cleans the policy dataset in preparation for merging it with the raw pollution and infant mortality data.

17. pov_merge_01102014.do

This program cleans the income dataset and prepares it for merging with the pollution data. The income data are used as a control in the regression analysis.

18. probareas_01102014.do

This file insheets a table of Problem Area Action Plan location and year of declaration, and then adds specific information about pollutant focus in each Problem Area.

19. temp_prpc_01102014.do

This file prepares the air temperature and precipitation data for merging with main water and air datasets. The air and temperature data were provided by Olivier Deschenes.

20. waterclean-apr09_01102014.do

The objective of this program is to compile and clean the raw water pollution data, and to prepare them for merging with control variables.

21. waterpol_2005_01102014.do

This file insheets raw data from the CPCB (by river basin) and cleans it in preparation for merging with the main water dataset.

22. waterpolicy_sam_01102014.do

The objective of this program is to perform manual cleaning of place names and policy enactment variables in the raw water pollution data.

3.2 Data Analysis

1. AppFigure1_01102014.do

The objective of this program is to generate Appendix Figure 1 in the paper. Appendix Figure 1 gives kernel density graphs of ambient air quality, surface water quality, and infant mortality.

2. AppTable1_01102014.do

The objective of this program is to generate Appendix Table 1 in the paper. Appendix Table 1 provides summary statistics, by year, of air quality, water quality, and infant mortality.

3. AppTable2_01102014.do

The objective of this program is to generate Appendix Table 2 in the paper. Appendix Table 2 provides the distribution of air and water pollution policies across all Indian cities, by year.

4. AppTable3_01102014.do

The objective of this program is to generate Appendix Table 3 in the paper. Appendix Table 3 provides the distribution of air and water pollution policies across all Indian cities, by RELATIVE policy year.

5. AppTable4_01102014.do

The objective of this program is to generate Appendix Table 4 in the paper. Appendix Table 4 provides a robustness check on the two-stage regression analysis by varying the range of event years used in both regression stages.

6. Figure1_01102014.do

This program generates Figure 1 for the academic paper. Figure 1 compares existing pollution levels in India, the US, and China.

7. Figure4.01102014.do

The objective of this program is to generate Figure 4 for the paper. Figure 4 plots trends in Air Quality, Water Quality, and Infant Mortality in India during over the course of the sampler period.

8. Figure5.01102014.do

The objective of this program is to generate Figure 5 in the paper. Figure 5 plots the air policy event-study estimates developed in the first stage of the two-stage regression analysis.

9. Figure6.01102014.do

The objective of this program is to generate Figure 6 in the paper. Figure 6 plots the water policy event-study estimates developed in the first stage of the two-stage regression analysis.

10. Figure8.01102014.do

The objective of this program is to generate Figure 8 in the paper. Figure 8 plots an event study of the effect of catalytic converters on infant mortality. The event year estimates are derived from the first stage of the two-stage regression approach.

11. Figure9.01102014.do

The objective of this program is to generate Figure 9 in the paper. Figure 9 graphs total nationwide references to air and water pollution in the Times of India newspaper over the course of the sample period.

12. Table1.01102014.do

This program tabulates the distribution of air and water policies across years, resulting in Table 1 for the paper.

13. Table2.01102014.do

Objective: The objective of this program is to generate air pollution, water pollution, and infant mortality summary stats. This program generates Table 2 for the paper.

14. Table3.01102014.do

The objective of this program is to generate Table 3 of the paper. This program performs mean shift and trend break analysis of air pollution policies using both a single-stage standard DiD approach as well as a two-stage event study and trend break approach.

15. Table4.01102014.do

The objective of this program is to generate Table 4 of the paper. This program performs mean shift and trend break analysis of water pollution policies using both a single-stage standard DiD approach as well as a two-stage event study and trend break approach.

16. Table5.Figure7.01102014.do

This program performs QLR Breakpoint tests on the air and water pollution policies. This file generates both Table 5 and Figure 7 for the academic paper. The breakpoint tests utilize user-generated commands (**qlrmean.ado** and **qlrtrend.ado**).

17. Table6.01102014.do

The objective of this program is to generate Table 6 for the paper. This program performs two-stage regression analysis, resulting in mean-shift and trend-break estimates using infant mortality as the outcome.

18. Table7.01102014.do

The objective of this program is to generate Table 7 for the paper. This program performs the standard two-stage analysis from Greenstone and Hanna, however in this case, the event years have been interacted with "high" and "low" categories formed on the basis of two control variables: urban literacy rate and annual number of references to air pollution in the Times of India.