

README for the Reproducibility Package for “Firms’ Preferences For Emission Reduction Policies And Measures: A Quantitative Analysis For India”

1. Overview

This package reproduces the results in “Analysing Firms’ Preferences for GHG Reduction Policies and Measures in India.” It contains a refined (de-identified) survey dataset and one (“Command do file.do”) producing the final results reported in the paper. Running the final do-file executes the full analysis and writes all regression tables to the `outputs` folder in under 10 minutes.

This study aims to understand Indian firms’ likely behavior in response to climate change obligations and government policies that enforce these obligations. Understanding the behavior or preferences of firms helps to design effective policies to reduce GHG emissions and help meet India’s climate change mitigation commitments.

Data Availability Statement

There are three data files. One Excel and two Stata files. The Excel file is Raw survey data, and the second Stata file, `Data_statafile.dta`, is the survey data used for this analysis. The third file `Pop new.dta` file, contains secondary data related to key variables of population (all firms) collected from the Prowess database by the Centre for Monitoring Indian Economy (CMIE).

The Stata main data file provided has been refined from the raw survey data. Firm names and other information that could identify individual respondents have been removed, as they contain confidential firm information. Where data were missing from survey responses, values were, where possible, supplemented from the individual firm’s publicly available website (not from confidential sources).

There is another stata `Pop new.dta` file which has data related to key variables of population (all firms) collected from the Prowess database by the Center for Monitoring Indian Economy (CMIE). This file (`Pop new`) is used to check Heckman two-step correction and to check Sample selection bias.

Data Sources

- **Filename:** [`Data_statafile.dta`]
- **Source:** Primary survey of firms in India, conducted by [Dr. Gobvinda R. Timilsina, World Bank]
- **URL:** Not publicly downloadable — confidential firm-level survey data
- **Access year:** [2023-24]
- **Access instructions for restricted data:** [“Not shared with third parties; available for verification under NDA via reproducibility@worldbank.org (mailto:reproducibility@worldbank.org)” or similar, per your own data-sharing agreement.]
- **Filename:** [`Pop new.dta`]

- **Source:** Secondary data collected from the Prowess database by the Centre for Monitoring Indian Economy [CMIE]
- **URL:** <https://prowess.cmie.com/> (<https://prowess.cmie.com/>) [It is a paid database and not freely available.]
- **Access year:** [2023-24]

Instructions for Replicators

1. Open the provided `Data_statafile.dta` file in Stata (refined/de-identified version of the raw survey data — see *Data Availability Statement* above).
2. Install the required Stata packages:

```
ssc install asdoc, replace
ssc install estout, replace // provides -eststo-
ssc install outreg2, replace
```

3. **Final results do-file** (`Command do file.do`) — contains the code that produces the final results reported in the paper. Run this file to reproduce the key analysis.
4. Running the final results do-file executes the key analysis and saves all regression tables to the `outputs` folder.
5. The script runs in under 10 minutes.
6. Figure data are provided directly in `Figure.xlsx` (originally extracted from 'Raw data' file, where question-wise responses are given. Responses have been refined and segregated under different categories - sectorwise, emission type and size level and underlying data using Excel's table functions, and do not require running Stata code.). The Figures Excel file contains question-wise data extracted from the raw survey data. The firms' responses are subsequently aggregated by sector, firm size, and emission intensity. The firm size is classified based on total revenue, following the size categories specified by the Government of India. The sector classification is based on the sector information provided in the survey raw data.

The Excel sheets titled "Sector Segregation" and "Size Segregation" contain the underlying response-level data for each respective survey question. For each question, the relevant filters were applied to identify responses corresponding to the respective sector or firm-size category. At the end of each sector/size group, the Subtotal function was used to calculate the total number of responses for that category. These calculated values were then copied and pasted as values into the respective figure sheets.

The sector classification reported in Table 1 of the manuscript comprises the following seven categories: Energy and Electricity (ENEL), Non-energy Mining and Processing (NEMP), Emission-intensive Manufacturing and Construction (EIMC), Low-emission-intensive Manufacturing (LEIM), Transportation (TRAN), Services (SERV) & Agricultural, Food, and Others (AGRI)

For the analysis, firms belonging to ENEL, NEMP, EIMC, and TRAN are classified as energy/emission-intensive firms, while firms belonging to LEIM, SERV, and AGRI are classified as non-emission-intensive firms. This classification is used to distinguish between emission-intensive and non-emission-intensive firms in the analysis.

The firm-size classification is based on total revenue, following the categories specified by the Government of India. The three categories used are Micro, Small, and Medium & Large firms.

7. Table 2 consists of (firms) respondents' sectors and size (on the basis of revenue) segregation, which is based on the raw data (responses) of questions 2 and 3 & 4.

Question 2 asks about the sectors and provides 10 choices. The sector information is then manually converted into numerical values as given in their respective sector columns, and these are further compiled into 7 sections by combining a few sectors. The Excel files show each formula and its calculations. The firm sizes are calculated on the basis of the responses given in Questions 3 and 4. The attached Excel file shows the raw responses and is segregated into three categories using an Excel formula. For some firms that did not report the revenue, we tried to obtain it from public sources. Wherever firms reported the exact revenue (Question 3) or the range of revenue streams (Question 4, which gives five ranges), the response to Q3 is also converted into the five ranges given in Q4, and these five ranges are then combined into three columns as given below. The response codes are: 1 for less than ₹10 crore, 2 for ₹10 crore to ₹100 crore, 3 for ₹100 crore to ₹500 crore, 4 for ₹500 crore to ₹1,000 crore, and 5 for more than ₹1,000 crore. These are then combined into the following three categories: 1 = Micro- Annual Revenue < 10 crore; 2 = Small, 10 to 100 crore; and 3 + 4 + 5 = Medium and Large, More than ₹100 crore annual revenue. The Excel files show each formula and its calculations.

8. Appendix tables B1 to B7 capture the quantitative information of sector-wise analysis of firms' preferences for different choices. These are the quantitative representation of the results presented in Figures 1–Therefore, all these details are available in Figures excel. The data used is provided directly in Appendix.xlsx (originally extracted from the 'Raw data' file) and uses Excel's table functions, so it does not require running Stata code.
9. Scale Descriptive Excel file contains the information of the questions 13, 15 and 17, which measure each response on a scale of 1-5. These are helpful to make table B3, B4 and B7 and their respective figures.

Software Requirements

- **Stata**, version 16 or later (code has been tested for compatibility with Stata 16)
- **Microsoft Excel** (used to generate the figure data — see *List of Exhibits / Code Description* below)

Memory and Storage Requirements

The full analysis (final results do-file) runs in under 10 minutes on a standard desktop/laptop; no unusual memory or storage requirements beyond a standard Stata installation.

Code Description

- Stata “do” files is provided:

Command do file.do (final results): loads the refined dataset, runs the SUR/IMR and multivariate probit models, and exports all regression tables (via `asdoc`, `eststo`, `outreg2`) to the `Outputs` folder. This is the file a replicator should run to reproduce the paper's results. - Code is written for and tested on **Stata version 16**. - Required packages for exporting results: **asdoc**, **eststo** (from `estout`), and **outreg2**. - The full script runs in **under 10 minutes**. - Data underlying the paper's figures are provided in `Figure.xlsx`. These values were extracted from `Figure.doc`, which itself can be regenerated from the same data using Excel's table functions. - `Tables.xls` contains the full set of marginal effects results; only the marginal effects reported in **Tables 4 and 5** of the paper are drawn from this file.

List of Exhibits

Appendix A Survey Instrument

Appendix B (B1–B7) Variations of firms' preferences/behaviour by type or sector

Appendix C - Table C1. Regression Results of Various SUR Models With IMR. Standard errors clustered at the individual respondent level. - **Table C2.** Robustness analysis — Multivariate Probit: Firms' Responses to Carbon Pricing. Standard errors clustered at the individual respondent level.

Folder Structure

```

Project/
├── Data/
│   ├── Raw/
│   │   └── Raw data.xlsx                (original, unedited survey data)
│   └── Processed/
│       ├── Data_statafile.dta          (refined/de-identified main analysis sample)
│       └── Pop new.dta                 (population dataset used for probit/Heckman se
lection step)
├── Code/
│   └── Command do file.do              (main / final-results script)
├── Outputs/
│   ├── Tables/
│   │   ├── Table_4.rtf
│   │   ├── Table_5.rtf and Table2.xlsx
│   │   ├── Table_C1.rtf
│   │   ├── Table_C2_mvprobit_main.rtf
│   │   └── Appendix B.xls              (Appendix B1-B7: firm preferences/behaviour by
sector)
│   ├── Tables.xlsx                    (full marginal effects; Tables 4 & 5 sourced f
rom here)
│   └── results.xlsx
│   └── Figures/
│       ├── All figures-Data for Detail Descriptive Analysis....xlsx
│       └── Figures.docx
└── README.md

```

Notes on this structure: - Raw data.xlsx is the untouched original file; it is never edited directly - Data_statafile.dta is the refined file and is used for analysis. Data_statafile.dta is the main analytical sample used in the second-stage/outcome regressions. - Pop new.dta is the population-level dataset used for the first-stage probit/selection model (Inverse Mills Ratio); - All table outputs (Table_4.rtf, Table_5.rtf, Table_C1.rtf, Table_C2_mvprobit_main.rtf, Appendix B.xls, Tables.xlsx, results.xlsx) are generated by Command do file.do and should not be edited by hand — rerunning the script will overwrite them. - Figure-related files (All figures-Data for Detail Descriptive Analysis....xlsx, Figures.docx) are Excel/Word-based and are not produced by the Stata script — see Instructions for Replicators, step 5.

—Table2.xlsx is generated by using raw data.

Contact - Dr Monika Gupta, monika.gupta@spjimr.org (<mailto:monika.gupta@spjimr.org>)

Paper: “Firms’ Preferences For Emission Reduction Policies And Measures: A Quantitative Analysis For India” **Last updated:** September 2, 2026