Documentation of data programs to reproduce the results of

Money (Not) to Burn: Payments for Ecosystem Services to Reduce Crop Residue Burning

American Economic Review: Insights

B. Kelsey Jack, Seema Jayachandran, Namrata Kala and Rohini Pande

Overview

This repository contains the (raw and cleaned) survey data and code necessary to reproduce the main dataset and results in the main text and online appendix using Stata.

One file (0_master.do) runs all the code to generate to build the dataset and produce the tables and figures referenced in the paper text and appendix. The code is commented accordingly to make it easy to correlate the output files with the corresponding figures or tables in the manuscript. The replicator should expect the code to run for about 15 minutes.

This README file has six sections. Section 1 "Data Availability Statements" explains how each dataset can be obtained. Section 2 "Dataset List" summarizes the data files. Section 3 "Computational Requirements" explains the setting under which the codes were run and documents requirements. Section 4 "Description of Programs" gives a high-level overview of the program files and their purpose. Section 5 "Instructions to Replicators" gives instructions to conduct the replication. Section 6 "List of Tables and Programs" provides information on which tables and figures are produced by which program files.

Data Availability and Provenance Statements

<u>Summary</u>

The data used in our analysis is publicly available. We include in the replication package all the raw data files used in our study from which all PII was removed.

Dataset list

Dataset name
completion_marked.xlsx
hfc_replacements.xlsx
Listing_raw.dta
preloads_v2.dta
vil_data.dta
intervention_raw.dta
imported_raw.dta
SBR_Baseline_revisit_v1.dta
SBR_Baseline_v1.dta
SBR_Baseline_v2.dta
SBR_Baseline_v3.dta
SBR_Baseline_v4.dta
SBR_Baseline_v5.dta
SBR_Baseline_v6.dta
SBR_Baseline_v7.dta
SBR_Baseline_v8.dta
SBR_Baseline_v9.dta
SBR_Baseline_v10.dta
SBR_Baseline_revisit_v1.dta
02_postcons.dta
vil_status.dta
9_villages_re-randomization_re_prioritize.dta
3_villages_assigned.dta
9_villages_stratification_var
replacement_106000691.dta
01_hfc_replacements_main.xlsx
02_hfc_replacements_postcons.xlsx
spotcheck_survey_wide_raw.dta
preloads_all.dta
mon_sc_discrepancies.dta
el_preload_v1.dta
endline_cons_obs.dta
endline_raw.dta
hfc_replacements.xlsx
int_preloads_1-7.dta
RS_field_LOOCV.dta
RS_iter_LOOCV.dta
vegInd2018_2019.csv
SBR_Census_v1_All.dta

Satellite data

We use satellite data from PlanetScope (Planet Labs PBC, 2018) and from Copernicus Sentinel-2 (2021), a product of the European Space Agency.

Data, code, and instructions for processing satellite data and the random forest model are available in Zenodo repositories [available here] (Walker, 2024a) and [available here] (Walker, 2024b). The methods were published as arXiv:2209.10148, "Detecting Crop Burning in India using Satellite Data" (2022), by Kendra Walker, Ben Moscona, Kelsey Jack, Seema Jayachandran, Namrata Kala, Rohini Pande, Jiani Xue, and Marshall Burke [available here].

The Planet imagery used in our analysis is governed by a university license granted to UCSB, which limits publication rights to: *(i) use, access, and view Content through the Platform; (ii) to reproduce, store, display, and print Content; (iii) create Derivative Products; and (iv) publish articles related to or based upon the Content; all for non-commercial purposes, and solely in education and research journals. Licensee agrees that all such publications must include an attribution that clearly and conspicuously identifies Planet Labs Inc. as the source of the Content on which the publication was based.* Note that Content refers to *data generated from satellite imagery made available to Licensee pursuant to the terms of this Agreement and as set forth in detail in an applicable Order Schedule*.

The Sentinel-2 imagery from ESA is publicly available.

Detailed instructions on accessing and processing Planet data for users with their own license, and on accessing and processing Sentinel-2 data for all users, are available at the above Zenodo repositories.

Computational requirements

- Stata (code was last run with version 17)

Software Requirements

The following Stata packages are required to run the code. The ado files are provided in the "Ado" folder, and the Stata ado search path is appropriately adjusted in "Do/0_master.do". External ado files can be redownloaded (not recommended) by running Do/_adosetup.do.

Package	Version	Author	Notes
_gwtmean	2.0.0	David Kantor	
catplot	2.0.2	Nick J. Cox	
elasticregress	1.2	Wilbur Townsend	
estout	3.31	Ben Jann	
grc1leg	1.05	Vince Wiggins	
icw_index	-	Adrien Bouguen	
ihstrans	1.0	Jan Helmdag	
lassoShooting	12	Christian Hansen	
mdesc	2.1	Dan Blanchette and Rose Anne Medeiros	
readreplace	2.0.0	Ryan Knight and Matthew White	Modified by _adosetup.do
rmiss2	2.0.0	Richard Goldstein	
texdoc	2.4.0	Ben Jann	
unique	1.2.4	Michael Hills and Tony Brady	
winsor2	1.1	Lian Yujun	
xfill	1.0.0	Tony Brady	
add_lines_stat_decile		(authors)	Custom-written
add_lines_stat_mean_sd		(authors)	Custom-written
statdesc_table_decile		(authors)	Custom-written
statdesc_table_mean_sd		(authors)	Custom-written

Memory and Runtime Requirements

There are no specific computational requirements other than those necessary to run STATA. The authors were able to run all scripts in the following computing environment: 11th Gen Intel(R) Core(TM) i7-1185G7 @ 3.00GHz (4 cores, 8 thread); STATA/MP 17.

Total computation of data construction and analysis takes less than 15 minutes.

Description of programs/code

"Do/0_master.do" sets directory, ado files location, and runs all dofiles needed for this replication package. Below, we provide an overview of the program files and their purpose.

- "Do/cleaning_listing.do": cleans listing data
- "Do/cleaning_intervention.do": cleans intervention data
- "Do/cleaning_baseline.do": cleans baseline data
- "Do/cleaning_spotchecks.do": cleans spotchecks data
- "Do/cleaning_monitoring.do": cleans monitoring data
- "Do/cleaning_endline.do": cleans endline data
- "Do/cleaning_ml.do": cleans machine learning based burning outcome data
- "Do/cleaning_census.do": cleans census data
- "Do/merge_and_create_variables.do": merges all the datasets and creates the main datasets used in the analysis.
- "Do/1_analysis.do": runs the main analysis and exports tables and figures
- "Do/2_cost_effectiveness.do": runs the cost-effectiveness analysis
- "Do/_adosetup.do": downloads external ado files (not run by default)

Instructions to Replicators

1. Please construct one master folder with the following structure of subfolders, include the corresponding files in each folder:

   - Data

     *** Create the following nine folders and store the listed files in each folder ***

     1. Listing
        - completion_marked.xlsx
        - hfc_replacements.xlsx
        - Listing_raw.dta
        - preloads_v2.dta
        - vil_data.dta
     2. Intervention
        - intervention_raw.dta
     3. Baseline
        - imported_raw.dta
        - SBR_Baseline_revisit_v1.dta
        - SBR_Baseline_v1.dta
        - SBR_Baseline_v2.dta
        - SBR_Baseline_v3.dta

- SBR_Baseline_v4.dta
- SBR_Baseline_v5.dta
- SBR_Baseline_v6.dta
- SBR_Baseline_v7.dta
- SBR_Baseline_v8.dta
- SBR_Baseline_v9.dta
- SBR_Baseline_v10.dta
- SBR_Baseline_revisit_v1.dta
- 02_postcons.dta
- vil_status.dta
- 9_villages_re-randomization_re_prioritize.dta
- 3_villages_assigned.dta
- 9_villages_stratification_var
- replacement_106000691.dta
- 01_hfc_replacements_main.xlsx
- 02_hfc_replacements_postcons.xlsx
- replacement_106000691.dta

4. Spotchecks
   - spotcheck_survey_wide_raw.dta
   - preloads_all.dta
   - mon_sc_discrepancies.dta
5. Monitoring
   - monitoring_main_raw.dta
   - monitoring_plot_raw.dta
   - 01_hfc_replacements_main.xlsx
   - 02_hfc_replacements_plot.xlsx
   - Compliant Farmers_ongoing list_from RA track.xlsx
   - Monitoring_track_RA.xlsx
   - monitoring_qual.dta
   - mon_sc_discrepancies.dta
6. Endline
   - el_preload_v1.dta
   - endline_cons_obs.dta
   - endline_raw.dta
   - hfc_replacements.xlsx
   - int_preloads_1-7.dta
7. ML
   - RS_field_LOOCV.dta
   - RS_iter_LOOCV.dta
   - vegInd2018_2019.csv
8. Census
   - SBR_Census_v1_All.dta

- Do

  *** Store the following dofiles (12) here ***
  - cleaning_listing.do

- cleaning_intervention.do
- cleaning_baseline.do
- cleaning_spotchecks.do
- cleaning_monitoring.do
- cleaning_endline.do
- cleaning_ml.do
- cleaning_census.do
- merge_and_create_variables.do
- 1_analysis.do
- 0_master.do
- _adosetup.do

- Output
  - Tables
  - Figures
- Ado

  *** Store Ado files provided here ***

2. Edit line 24 of "Do/0_master.do" to set your directory to this master folder.

3. Compiling "Do/0_master.do" will build the dataset and populate the Tables and Figures folders.

   *** Please note that the preamble of "Do/0_master.do" must be run always first if any of the dofiles is ran independently.

   The code is commented accordingly to make it easy to correlate the output files with the corresponding figures or tables in the manuscript.

List of tables and programs

Figure/Table #	Program	Line #	Output file
Figure 1	1_Analysis.do	23	contractelegibility_takeup_nostar.png
Figure 2		316	cost_effectiveness_graph.pdf
Table 1		1006	teffects_compliance_notburn_crm_nostar.tex
Table 2		1058	hetero_fin_distrust_compliance_with_trustpay_cashshortage_takers_part_a_nostar.tex hetero_fin_distrust_compliance_with_trustpay_cashshortage_takers_part_b_nostar.tex
Figure A.1	N/A	N/A	design.tikz
Figure A.2	N/A	N/A	timeline.tikz
Figure A.3	N/A	N/A	RSimage1.jpg RSimage2.jpg RSimage3.jpg
Figure A.4	1_Analysis.do	1191	rs_burn_measure_robust.png
Table A.1		1472	samples_comparison_nostar.tex
Table A.2		1654	balance_test_nostar.tex
Table A.3		1970	attrition_nostar.tex
Table A.4		2029	hetero_attrition_reg_nostar_v2.tex
Table A.5		2217	tablea9_teffects_notburningmeasures_nostar.tex
Table A.6		2330	treatment_effects_takeup_compliance_notburning_nostar.tex
Table A.7		2447	crm_leebounds_nostar.tex
Table A.8		2896	yield_sowingdelay_endline_nostar.tex
Table A.9		3350	hetero_selection_takeup_nostar.tex
Table A.10		3414	hetero_crm_compliance_nostar.tex
Table A.11		3477	confusion_nostar.tex
Figure A.5		3578	rs_balance.png

References

Copernicus Sentinel-2 (processed by ESA). 2021. MSI Level-1C TOA Reflectance Product. Collection 0. European Space Agency. https://doi.org/10.5270/S2_-d8we2fl

Jack, B.K., Jayachandran, S., Kala, N., and Pande, R. 2023. "Data and Code for: Money (Not) to Burn: Payments for Ecosystem Services to Reduce Crop Residue Burning. *American Economic Association* [publisher], Inter-university Consortium for Political and Social Research [distributor]. http://doi.org/10.3886/E1955621V2

Planet Labs PBC. 2018. Planet Application Program Interface: In Space for Life on Earth. Planet. https://api.planet.com

Walker, Kendra. 2024a. klwalker-sb/burntfields_punjab: initial release (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.11225292

Walker, Kendra. 2024b. Extra data to accompany code in GitHub burntfields_punjab, both used in Walker et. al. 2022, Detecting crop burning in India using satellite data (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10987987