Documentation for Published Datasets

Paying Farmers Not to Burn: A Randomized Trial of Payments for Ecosystem Services in India

Aug 2021

Seema Jayachandran (Northwestern University), Namrata Kala (MIT Sloan School of Management), Kelsey Jack(UC Santa Barbara), Rohini Pande (Yale University)

---

1. Folder structure
    0. Readme Document
    1. Questionnaires
        - baseline_questionnaire_v1
        - baseline_questionnaire_v10
        - intervention_questionnaire
        - monitoring_questionnaire
        - Spotchecks_questionnaire
    2. Datasets
        - baseline_raw
        - intervention_raw
        - monitoring_raw
        - Spotchecks_raw

---

2. Questionnaires
    a. Baseline
        i. Enumeration dates: 2 October – 13 October, 2019
        ii. Sample: The respondents in the sample were selected among a list of farmers who are members of a co-operative society in the districts of Faridkot and Bathinda in the State of Punjab in India (1,668 respondents). The criteria by which the farmers were deemed eligible for the baseline survey were the following:
            1. They cultivated land in Kharif 2019
            2. They cultivated paddy on land ranging between 0 and 13 acres
            3. They plan to grow any crop in Rabi 2019
            4. They plan to harvest in the 3rd week of October or later
        iii. There are two versions of the questionnaire: Version 1 of the questionnaire was administered on the first day of the survey, Version 10

on all the following days; the versions are identified by variable *survey_version*. The differences between version 1 and 10 are:
    1. Section K was dropped in Version 10
    2. The variable *crop_burning* had the wrong label: *D8* instead of *7a* (this was corrected while creating the final dataset)
    3. The variable *land_plot_seccrop* had the wrong choices: *crop* instead of *crop_alt* (this was corrected while creating the final dataset)

b. Intervention
    i. Enumeration dates: October 10 – October 17, 2019
    ii. Sample: Only respondents in the treatment arms were surveyed (1,182 respondents)
    iii. The variables *landsize_inconsistent* and *selfreported_acreage* were added to the questionnaire on day 5 of the survey after some farmers had disagreed with the recorded land size in the first days of the intervention.

c. Monitoring
    i. Enumeration dates: 18th October – 03rd December, 2019
    ii. Sample: Only respondents in the treatment arms who called JPAL to request monitoring of their plots were surveyed (189 respondents)

d. Spotcheck
    i. Enumeration dates: 18th October – 03rd December, 2019
    ii. Sample: 1 plot each of 50% of the farmers in every village in the sample (including control villages) were randomly selected for spot checks (755 respondents)

Note that not all questions from the questionnaires are present as variables in the datasets due to containing personal identifiers or duplicate information across data sets.

---

3. Datasets
    a. Baseline
        i. The baseline dataset has variables on the demographics of the respondents, their land and agricultural outcomes, their crop residue management techniques, their ownership and renting of machines, their household assets, their children's health, their sources of income, their access to credit, and their trust in groups of people and institutions.
    b. Intervention
        i. The intervention dataset has variables on the treatment assignment, the amount to be paid to the farmer upfront and after monitoring, whether the respondent's household could be reached during intervention, the respondent's program participation decisions, and reasons why they decided not to participate.
    c. Monitoring
        i. The monitoring dataset has variables for each monitored plot on whether different signs of burning are detectable.

d. Spotcheck
    i. The spot checks dataset has variables for each plot that was randomly selected for checking on whether different signs of burning are detectable.

All datasets merge on the variable *resp_id* (the respondent ID). Note that the *monitoring* dataset has more than one observation for some respondents in case there were several visits by a monitor to these respondents. The plot level variables in the baseline dataset determine which number (position) the plot was assigned (i.e. first plot, second plot etc.). The plot numbers in the monitoring and spot checks data correspond to the plot numbers in the baseline data.

---

4. Notes on Variables
    a. Baseline
        i. The answers to variables *crm_aware* (D1: Knowledge of crop residue management techniques) and *crm_try* (D2: Tried crop residue management technique before) are not consistent e.g. respondents who answered that they are not aware of crop residue management techniques, also answered that they have tried some before. This might be due to a misunderstanding of the questions.
        ii. Some of the measured plot area variables (*field_area_* *) equal zero, which is a measurement error.
        iii. The following variables are preloaded from the listing exercise of the co-operative members: *check_cultivation, confirm_cultivation, cultivation_confirmed, land_cult_list, check_landsize, confirm_landsize, landsize_confirmed, check_rabi confirm_rabi, rabi_confirmed, harvest_plan, check_harvest, confirm_harvest, harvest_plan_confirmed*
            1. where the variables *land_cult_list* and *harvest_plan* contain data obtained during the listing exercise
            2. where the variables *check_* * state whether during the baseline survey the respondent deemed the data obtained during the listing exercise correct or incorrect
            3. where the variables *confirm_* * contain the true values in case the respondent deemed the listing data as incorrect
            4. where the variables **_confirmed* contain the true values for all respondents (combining the correct data from the listing exercise and the corrections during the baseline survey)
    b. Intervention
        i. The four treatment arms are indicated by the variable *treatment*. The respondents who are present in the baseline dataset, but not in the intervention dataset are in the control group. The four treatment arms are (where PES stands for Payment for Ecosystem Services):

1. PES 800/acre: respondents would receive 800 rupees per acre if monitoring determined that they had not burned their paddy stubble
2. PES 1600/acre: respondents would receive 1600 rupees per acre if monitoring determined that they had not burned their paddy stubble
3. PES 800/acre with 25\% upfront payment: respondents receive 200 rupees per acre at the time of signing the contract and would receive an additional 600 rupees per acre if monitoring determined that they had not burned their paddy stubble
4. PES 800/acre with 50\% upfront payment: respondents receive 400 rupees per acre at the time of signing the contract and would receive an additional 400 rupees per acre if monitoring determined that they had not burned their paddy stubble

ii. *participate_levelone* is a variable indicating whether the respondent decides to continue participating after hearing a general introduction of the program offered. This implies that the respondent wants to hear the wording of the agreement.

iii. *participate_leveltwo* is a variable indicating whether the respondent decides to continue participating after hearing the wording of the potential agreement between them and JPAL. This implies that the respondent wants to hear the terms and conditions connected to the agreement.

c. Monitoring
d. Spot checks
e. Plot level variables
   i. In the baseline dataset, the plot number is given by the variables called *land_plot_id_**. There are several plot level variables in wide format with the plot number indicated in the suffix of the variable (e.g. _1, _2, etc.) and also in the label of the variable (Plot 1, Plot 2, etc.).
   ii. In the monitoring dataset, the plot number is given by the variables *plot_pos_**. The plot numbers correspond with the plot numbers in the baseline dataset. There are several plot level variables in wide format with the plot number indicated in the suffix of the variable (e.g. _1, _2, etc.) and also in the label of the variable (Plot 1, Plot 2, etc.).
   iii. In the spot checks dataset, the plot number is given by the variable *plot_pos*. The plot numbers correspond with the plot numbers in the baseline dataset. Only one plot per farmer was checked, so each observation corresponds to one plot of a farmer.