**ReadMe File for: Inefficient water pricing and incentives for conservation**

Ujjayant Chakravorty[1], Manzoor H. Dar[2], and Kyle Emerick[3]

## Overview

The code in this replication package constructs the analysis from twelve data sources using Stata. The main do-file "replicate_results.do" runs all the code to generate the data for the 9 figures and 50 tables in the paper. After running the main do-file, the auxiliary do-file "replicate_results_A15.do" provides the code and instructions to produce table A16. Using Stata 16, the replicator should expect the code to run for about 10 minutes. The do file "setup_file.do" should be run first. It will install all necessary packages and programs.

Figure 1 and 2 are maps made using GIS software. The folder GIS maps contains a text file named "map instructions.txt". That text file contains instructions for recreating Figure 1 (with associated data contained in the folder Figure 1) and an explanation of how Figure 2 was created.

Any questions about these replications materials should be directed to the corresponding author at kyle.emerick@tufts.edu.

## Data Availability and Provenance Statements

All the data (primary and secondary) used to support the findings of this study have been deposited in the following repository: openicpsr-147781.

The data were collected by the authors, and are available under *a* Creative Commons Non-Commercial license.

### Statement about Rights

☐ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

### Summary of Availability

All data provided can be made publicly available.

---

[1] Department of Economics, Tufts University and Toulouse School of Economics, Medford, MA 02155-6722
[2] International Crops Research Institute for the Semi-arid Tropics, Hyderabad, India.
[3] Corresponding author. Email: kyle.emerick@tufts.edu. Department of Economics, Tufts University and CEPR, Medford, MA 02155-6722.

## Details on each data source

> ### Primary Data

There are eight datasets that were generated from surveys conducted by our teams in the field. The questionnaires for these surveys, as well as protocols, are provided as part of the replication package (survey_instruments folder). The table below lists these datasets (input_primary folder), along with the date the surveys were conducted, the sample selection criteria, and the corresponding questionnaire and protocol files.

| Datafile | Date the survey was administered | Questionnaire | Interviewer Instructions | Sample Selection Criteria |
|---|---|---|---|---|
| rct1_baseline.dta | Nov-Dec 2016; Jan-March 2017 | Survey_instruments/awdbase_v7.xls | Survey_instruments/Baseline survey protocol.pdf | 10 farmers in each village identified by the tube-well operator. Must be cultivating land in command area of the tube-well |
| rct1_boro2018 pipesales | January 2018 | Survey_instruments/awd_phase2base.xls | | Same as RCT1 baseline survey. |
| rct1_followup.dta | July 2017 | Survey_instruments/awdfollow.xls | | Same as RCT1 baseline survey. |
| rct1_watermeasure.dta | July 2017 | Survey_instruments/waterlevel.xls | Survey_instruments/Water Level Protocol.pdf | Same as RCT1 baseline survey. Two visits for each farmer's study plot, as described in Water Level Protocol.pdf |
| rct2_baseline.dta | December 2017 | Survey_instruments/awd_phase2base.xls | Survey_instruments/Phase 2 Pipe Sales Protocol.pdf | 25 farmers enumerated by the tube-well operator as cultivating at least one rice plot in the command area of the tube well. |
| rct2_pipesales.dta | Jan-Feb 2018 | Survey_instruments/awd_phase2base.xls | Survey_instruments/Phase 2 Pipe Sales Protocol.pdf | Same as RCT2 baseline survey. |
| rct2_usage.dta rct2_dailyusagefarmer.dta | Feb-May 2018 | Survey_instruments/awd_phase2water.xls | Survey_instruments/Water Level Protocol.pdf | Same as RCT2 baseline survey. |

> ### Secondary Data

There are four datasets that use secondary data obtained from external sources, that are described below (input_secondary folder).

Datafile: rct2_cardreload_paba.dta

The Barind Multipurpose Development Authority provided data on loading activity for the prepaid cards that were distributed as part of RCT 2. For each card, the BMDA provided each time it was re-loaded and the amount loaded onto the card by the farmer. The data were provided directly to the researchers by the BMDA.

Datafile: `AgronomicLiterature.dta`

The paper compares its results to those published in other 17 agronomic literature papers. The list of these papers, with their complete reference, is included in the secondary dataset "Agronomic Literature."

Datafile: `distance_upazila_phase2.dta`

This dataset was produced in ArcGIS. It was created by calculating the distance between each farmer's house and the four major cities / towns in the area (Mohanpur, Paba, Tanore, and Rajshahi). The distances between each farmer's house and each of the four cities / towns were calculated using the "Generate Near Table" tool in ArcGIS.

Datafile: `GIS_covariates.dta`

Using confidential identifiers (latitude and longitude) we obtain individual geospatial characteristics using publicly available data on Google Earth engine: soil levels of water, carbon, sand, clay, and elevation. These five raster images were downloaded from Google Earth Engine. They were imported into ArcGIS, where the "Extract Multi Values to Points" tool was used to extract the five characteristics for each of the plots in RCT 1.


**Dataset list**

| Data file | Source | Notes | Provided |
|---|---|---|---|
| data/input_primary/rct1_baseline.dta | Primary survey | Baseline survey for RCT 1 | Yes |
| data/input_primary/rct1_boro2018pipesales | Primary survey | Pipesales for RCT 1 villages | Yes |
| data/input_primary/rct1_followup.dta | Primary survey | Follow-up survey for RCT 1 | Yes |
| data/input_primary/rct1_watermeasure.dta | Primary survey | Results from in-season water measurements for RCT 1 | Yes |
| data/input_primary/rct2_baseline.dta | Primary survey | Baseline survey for RCT 2 hourly irrigation card experiment | Yes |
| data/rct2_dailyusagefarmer.dta | Primary survey | Daily panel of rct2_usage.dta | Yes |
| data/input_primary/rct2_pipesales.dta | Primary survey | Data on AWD pipe sales for hourly irrigation card experiment | Yes |
| data/input_primary/rct2_usage.dta | Primary survey | Data on AWD pipe usage for hourly irrigation card experiment | Yes |
| data/input_secondary/AgronomicLiterature.dta | Various Agronomic Literatures (see .dta file for complete References) | Dataset containing results from agronomic trials on AWD. These results are compared to the RCT impact estimates to generate Figure 5 | Yes |

| | | in the paper. | |
|---|---|---|---|
| data/input_secondary/ distance_upazila_phase2. dta | ArcGIS | Using ArcGIS it estimates distance between each farmer's house and four major cities | Yes |
| data/input_secondary/ GIS_covariates.dta | Google Earth Engine | Farmer-level soil characteristics | Yes |
| data/input_secondary/rct2 _cardreload_paba.dta | Barind Multipurpose Development Authority (BMDA) | Data on hourly irrigation card reloading behavior for treatment farmers in Paba upazila (one of the three upazilas in the second RCT) | Yes |

# Computational requirements

## Software Requirements

The software used to run the code is Stata (last run with version 17). The following programs are required and available at scc install:

– estout
– mhtexp
– lassopack
– rforest
– boottest
– regsave

In addition, the user-written programs sg97_5.ado *(Stata Journal, volume 12, number 4)* and gr0034.ado *(Stata Journal, volume 8, number 2)* are also required to run the code. These are readily available to download.

The required program fanreg.ado is provided in the do-file folder.

All of these programs are installed by running the do file "setup_file.do"

## Controlled Randomness

The following random seeds are set in the program "replicate_results.do:"

– Line 228, seed 76561
– Line 1094, seed 591852
– Line 2588, seed 67184
– Line 2909, seed 567123

## Memory and Runtime Requirements

The code was last run on a 7-core Intel-based laptop with Windows 10 Home version 10.0.19043.

## Description of programs/code

Program "replicate_results.do" does all the analysis and produces all tables and figures of the paper, except for Table A15, which is reproduced using program "replicate_results_tableA15.do"

# Instructions to Replicators

1.  Make sure you have a folder structure like the following in your working directory:

    –   *data*
        *   *input_primary*
        *   *imput_secondary*
        *   *derived*
    –   *do_file*
    –   *results*

2.  Run the program "setup_file.do" to install all packages / programs that are used in the main replication file.

3.  Run program "replicate_results.do" to run all steps in sequence.

4.  Run program "replicate_results_A15.do".

    –   First run up to line 136
    –   At this point, from the browse data window, copy in the clipboard the p-values generated (column 2)
    –   Continue to run the program from line 137 onwards.
    –   When the program pauses, go into edit form (edit command), and paste the copied p-values in the corresponding column.
    –   Press q (as prompted in the command window), and the program will finish running.


# List of tables and programs

Tables 1-7, A1-A14, A16-A19, C1-C6, D1, as well as figures 3-5, A2-A3, B1, C1-C3, are generated in the do-file "replicate_results.do"

Table A15 is generated in the do-file "replicate_results_A15.do"

The file "map instructions.txt" explains how Figures 1 and 2 were created.