

Replication manual: “Improving Last-Mile Service Delivery using Phone-Based Monitoring”

1 Overview

1.1 Description of content

The replication archive contains a directory with 3 sub-directories, a master analysis file, and this Readme document. This manual describes in detail how one can replicate all analysis-based tables and figures in the published version of “Improving Last-Mile Service Delivery using Phone-Based Monitoring”. To begin with, we briefly describe each sub-directories’ content (for full details, see Section 4):

- *Code* contains one Stata do-file for each table and figure. For instance, the table on the effects on other MAO outcomes (Table 5) is generated using “*Table5.do*”. Note that those do-files will not run without first setting the globals in the master do-file “*Master.do*”.
- *Input* stores all data files used in the analysis. All files are in Stata 13 format (and as such are forward compatible). Details on the data are provided below in Section 2.
- *Output* contains all .tex and .pdf files generated by the analysis code. Every time one runs a code file, the respective output file is overwritten. Note that this folder is empty at the outset and can be populated by running the analysis code.

Additionally, the base directory contains the master analysis file which creates the computing environment in Stata to execute the analysis

- *Master.do* This is the key-file which one should open in order to execute the analysis. The user has to set a global macro pointing to this replication directory (see next subsection). Then additional paths are automatically set (to data and output directories). After this, one can run each do-file separately.

1.2 Example

Suppose one downloaded the zip-archive containing the content described above and unpacked it to “C:/Users/testUser/ReplicationDirectory/”. In order to replicate the finding of - say Tables 3 and 4, the key encashment tables - one would open Stata and open the “*Master.do*” file. The key step is to enter the base directory in line 9 to point to the directory where the extracted contents of the zip-file are. In this example, line 9 should become (note the absence of a trailing slash):

```
global Directory "C:/Users/testUser/ReplicationDirectory"
```

This path becomes the main directory, and separate paths for data, output, and code are set (lines 11-15). At this point, one can open the “*Table3and4.do*” file and execute the do-file or take a closer look at the analysis. Alternatively, one can follow the content of “*Master.do*”.

2 Data

2.1 Description

The sub-directory “*Input*” contains several data files. We provide the administrative datasets that were used in the primary analysis as well as all the variables from the phone survey that were used in secondary analyses. These are discussed in detail in section 4, when we also describe the variables.

2.2 Identifiers

The data sets used in the paper contained information such as administrative details of mandals and land holdings of farmers that could have potentially been used to identify outcomes for particular individuals. We have removed all the original identifiers from the data that would identification of individuals or small groups of individuals.

Given these concerns, certain data files cannot be merged with each other. Specifically, the variables used in the balance tests are split into three different files. While the uniqueness of their MAO and strata identifiers is retained, the actual IDs are different and cannot be matched between the three files. Similarly, none of these files have identifiers that allow them to be merged with any of the famer-level administrative data sets.

There are some key ID-variables which are used in almost all of the analysis:

- “*dist_uid*” is the district ID.
- “*dist_mandal_uid*” is the mandal ID.
- “*mao_id*” is the MAO ID, which is used in most specifications to cluster the standard errors. Note that certain MAOs are responsible for multiple mandals, so in some cases there are multiple *dist_mandal_uid* values matched to the *mao_id*.
- “*strata*” is the Strata ID that is used to implement randomization strata fixed effects in most specifications.

3 Code

3.1 Structure

The directory *Code* contains an individual code-file for each table or figure (except for tables 3-4 and A5-A6, which are rolled into common code-files respectively)¹. The name of the file corresponds to a table or figure in the paper. All do-files are written such that once the environment is set - as described in section 1.2 - one can work with them independently. Each code file saves a table or figure file in the *Output* directory.

Note that most of the code is not very computationally intensive and will run in a matter of minutes. However, Figure A2 and Table A7 rely on Monte Carlo simulations and will typically take

¹ The replication directory does not contain the code and data set up to directly reproduce Table 2 (the balance tests) as it is in its current form in the paper. In order to prevent the identifiability of mandals, we have split the vector of variables on which we test balance into three categories. Thus, there are three smaller data sets (summarized in section 4), and a singular .do file “BalanceTests.do” that loops over all three. The existing code generates three different .tex files with balance tests. In addition, the replication directory does not include code and data to reproduce the randomization map.

6 hours and 6 days to run depending on the computer. They can be run more quickly with fewer iterations.

3.2 External packages used

Throughout the analysis, we used a set of external packages which can be downloaded from *SSC* in Stata. We give a brief overview over the packages we use here:

- *reghdfe*: For all our regression specifications, we use the *-reghdfe-* command to efficiently implement strata fixed effects on a large number of observations
- *_gwtmean*: We use this command to calculate weighted means within groups of variables.
- *texsave*, *sxpose*: To generate tables 3, 4, and A7, we use this user-written command that outputs the stata dataset itself as a .tex file with a large number of formatting and customization options.
- *esttab*: Throughout the analysis we're using the combination of *-eststo-* and *-esttab-* to create publication-style tables in .tex format.

4 Detailed list of content

We describe the contents of every sub-directory contained in the replication zip-folder:

- *Code*: “*Table3.do*” through “*TableA8.do*” create the tables in Stata. “*Figure2.do*” through “*FigureA2.do*” create figures in Stata.
- *Output* : Contains the tables and figures; can be populated by running the do-files
- *Master.do*: see Sections 1.1 and 1.2
- *Input*: this folder contains all data files used in the analysis. In alphabetical order:
 - *AdminData.dta*: contains variables from the land registry of all farmers in the state, the administrative dataset for Rythu Bandhu with check encashment outcomes, and some variables from the phone survey that are used in the primary analysis. A full list of the variables is provided below. The data set is at the plot level, as in the land registry.
 - *a1*: Dummy for whether the farmer reported receiving a check receipt in the phone survey. Note that this is only for farmers in the administrative dataset with whom a phone survey was conducted.
 - *admin_dist*: Dummy variable for whether the MAO reported distributing the check, based on MAO database
 - *admin_encash*: Dummy for check encashment, based on administrative bank records of check encashment.
 - *adj_encash*: Dummy for check encashment for farmers at the time that the phone survey was conducted in their district. This is based on the bank administrative records, but adjusted for call date. This variable reflects the encashment status as of the date of the phone survey for farmers sampled for the phone survey and the median date of calls made to the district in the case of farmers who were not sampled for the phone survey.
 - *admin_encash_ontime*: Dummy for whether check was encashed by June 8th, based on administrative bank records of check encashment.
 - *b3*: Dummy for check encashment, as reported in the phone survey (only for farmers in the administrative dataset matched with respondents from the phone survey)
 - *dist_mandal_uid*: Mandal identifier

- *dist_uid*: District identifier
 - *encash_date*: Date of encashment, for those who encashed their checks. Based on administrative bank records of check encashment.
 - *encash_days*: Days taken to encash the check (difference between the dates of encashment and distribution). Based on administrative bank records of check encashment.
 - *encash_days_report*: Running variable used in survival analysis to test the direct effect of report cards sent to MAOs and DAOs. Based on administrative bank records of check encashment.
 - *farmer_uid*: Farmer identifier
 - *hasmobile*: Dummy for whether farmer has a registered phone number in the land registry
 - *land_q*: Land size quartiles. This is based on the land registry, and splits farmers into five groups based on the size of their land holdings (0-20th percentile statewide, 20-40th, etc)
 - *landsize_q25*: Land size bins (of 2.5% each). This is based on the land registry, and divides farmers into 40 equally sized bins based on the percentile ranking of the size of their land holdings relative to the state (0-2.5th percentile of holdings, 2.5-5th percentile, etc.)
 - *mao_id*: MAO identifier
 - *mao_sampled*: Dummy for whether the MAO was sampled for the MAO survey
 - *mao_surveyed*: Dummy for whether the MAO was sampled for the MAO survey, and the survey was completed
 - *report*: Dummy for the sending of DAO and MAO reports (used in survival analysis)
 - *sampled*: Dummy for whether farmer was sampled for phone survey
 - *sampled_completed*: Dummy for whether farmer was sampled for phone survey, and the survey was completed
 - *strata*: Randomization strata identifier
 - *treatment*: Treatment dummy
 - *unique_farmer_tag*: Tag for farmer uniqueness since some farmers received multiple checks for different landholdings. This can be used to collapse the data to farmer level instead of the current form, which is at the check level.
- BalanceTests1.dta: contains the full set of variables from the land registry that are used for balance tests. The strata and MAO identifiers in this data set cannot be used to link to any of the other files.
- *hasmobile*: Proportion of farmers in the mandal who have a registered phone number in the land registry
 - *landsize*: Mean land holdings among farmers in the mandal
 - *mao_id_bl*: MAO identifier. Note that this is different from the MAO identifier in the Admin data and other data sets, and so cannot be matched.

- *num_farmers*: Number of farmers in the mandal
 - *p25_landsize*: 25th percentile of land holdings in the mandal
 - *p50_landsize*: Median land holdings in the mandal
 - *p75_landsize*: 75th percentile of land holdings in the mandal
 - *strata_b1*: Strata identifier. Note that this is different from the strata identifier in the Admin data and other data sets, and so cannot be matched.
 - *treatment*: Treatment dummy
- BalanceTests2.dta: contains the full set of variables from the 2011 Census and other federal datasets that are used for balance tests. The strata and MAO identifiers in this data set cannot be used to link to any of the other files.
- *allbanks_total*: Number of banks in mandal
 - *atm_dist_cont*: Average distance to nearest ATM in the mandal
 - *female_pop*: Share of female population
 - *hhshare_banking*: Share of households using banking services
 - *kuchha_length*: Total length of gravel road in mandal (km)
 - *literacy*: Literacy rate in mandal
 - *mao_id_b2*: MAO identifier. Note that this is different from the MAO identifier in the Admin data and other data sets, and so cannot be matched.
 - *nearest_bank*: Average distance to nearest bank in the mandal
 - *pucca_length*: Total length of all-weather road in mandal (km)
 - *pucca_road*: Share of villages with all-weather road(s)
 - *sc_pop*: Share of SC population
 - *st_pop*: Share of ST population
 - *strata_b2*: Strata identifier. Note that this is different from the strata identifier in the Admin data and other data sets, and so cannot be matched.
 - *tot_pop*: Total mandal population
 - *treatment*: Treatment dummy
- BalanceTests3.dta: contains the full set of variables from administrative records housed with the state government that are used for balance tests. The strata and MAO identifiers in this data set cannot be used to link to any of the other files.
- *rainfall*: Average rainfall in mandal 2013-17 (mm)
 - *paddy*: Share of villages with paddy as main crop
 - *power_all*: Share of electrified villages
 - *dist_hyd*: Average village distance from Hyderabad
 - *irr_area*: Share of irrigated land

- *hhshare_mobile*: Share of households owning mobile phones
 - *work_pop*: Share of working population
 - *mao_age*: Age of MAO
 - *mao_gender*: Gender of MAO (Female = 1)
 - *num_samples_17*: Number of SHC samples (2017)
 - *shc_farmers_17*: Number of farmers covered by SHCs (2017)
 - *tests_entered_17*: Number of SHC tests conducted (2017)
 - *shc_available_17*: Number of SHC tests available on portal (2017)
 - *num_samples_16*: Number of SHC samples (2016)
 - *shc_farmers_16*: Number of farmers covered by SHCs (2016)
 - *tests_entered_16*: Number of SHC tests conducted (2016)
 - *shc_available_16*: Number of SHC tests available on portal (2016)
 - *mao_id_b3*: MAO identifier. Note that this is different from the MAO identifier in the Admin data and other data sets, and so cannot be matched.
 - *strata_b3*: Strata identifier. Note that this is different from the strata identifier in the Admin data and other data sets, and so cannot be matched.
- CostEffectiveness.dta: contains the daily total value of checks encashed in treatment and control mandals, and the total value of checks that are eligible for encashment in treatment and control mandals. This data set is used to run sensitivity analysis on the cost-effectiveness estimates presented in Figure 4.
- *value_*: Variables beginning in “value” contain the daily value (in rupees) of checks encashed in each mandal daily
 - *land_sum*: The total land size owned by farmers in the treatment and control mandals
 - *treatment*: Treatment dummy
- LandsizeVector.dta: contains the exact value of land size for each of the farmers in the dataset. This comes from the government land registry, but does not contain identifiers and so cannot be linked to other data sets.
- *landsize*: The exact number of acres owned by a farmer, in acres.
- MCDData.dta: contains a limited set of variables, including adjusted encashment outcomes and village, mandal and district identifiers. These cannot be linked to any other data set in the package to protect identifiability. This file is used to process the Monte Carlo specifications used in Table A7 and Figure A2.
- *dist_mandal_uid*: Mandal identifier. Note that is different from the mandal identifier used in the Admin data and other data sets, and so cannot be matched.
 - *land_m_q*: Mandal level land size quintiles (for control mandals)

- *land_y_q*: Village level land size quintiles (for treatment mandals)
 - *mao_id*: MAO identifier. Note that this is different from the MAO identifier in the Admin data and other data sets, and so cannot be matched.
 - *num_of_villages*: Number of villages in treatment mandals
 - *farmer_pop*: Farmer population in the treatment mandals
 - *strata*: Strata identifier. Note that this is different from the strata identifier in the Admin data and other data sets, and so cannot be matched.
 - *treatment*: Treatment dummy
 - *village_uid*: Village identifier
- MiscData.dta: data at the mandal level, with a limited number of variables that are used to test for spillovers across mandals and whether the treatment affected other MAO activities.
- *ada_code*: Code for the revenue division
 - *dist_mandal_uid*: Mandal identifier. Note that is different from the mandal identifier used in the Admin data and other data sets, and so cannot be matched.
 - *mandals_ada*: Number of mandals in the revenue division
 - *mao_id*: MAO identifier. Note that this is different from the MAO identifier in the Admin data and other data sets, and so cannot be matched.
 - *num_samples_18*: Number of SHC (soil health card) samples sent for testing in the mandal in 2018
 - *shc_available_18*: Number of SHCs available on online portal in the mandal in 2018
 - *shc_farmers_18*: Number of farmers covered by SHCs in the mandal in 2018
 - *tests_entered_18*: Number of SHC tests conducted in the mandal in 2018
 - *trt_ada*: Number of treatment mandals in the revenue division
 - *strata*: Strata identifier. Note that this is different from the strata identifier in the Admin data and other data sets, and so cannot be matched.
 - *treatment*: Treatment dummy
- PhoneData.dta: contains the set of variables from the phone survey (apart from check receipt and encashment) that are used in the final analysis.
- *a6_chq_where_y*: Dummy for whether check was received at the Gram Sabha
 - *a8_bribe_y*: Dummy for whether farmer was asked to pay a bribe at the time of check collection
 - *b7_satisfied_y*: Dummy for whether farmer was satisfied with the program
 - *mao_id*: MAO identifier. Note that this is different from the MAO identifier in the Admin data and other data sets, and so cannot be matched.
 - *strata*: Strata identifier. Note that this is different from the strata identifier in the Admin data and other data sets, and so cannot be matched.
 - *treatment*: Treatment dummy

- *weight*: (Inverse probability of sampling) weight