DOCUMENTATION FOR REPLICATION PACKAGE FOR 'MOBILE'IZING AGRICULTURAL EXTENSION'
SHAWN COLE AND A. NILESH FERNANDO (5/30/2020)

A. DIRECTORY STRUCTURE

~/do/ --- contains do files
~/dta/ --- contains final datasets
~/dta/ascii/ --- contains ASCII versions of final datasets
~/raw/ --- contains raw data
~/raw/ascii/ -- contains ASCII versions of raw dataset
~/out/ --- contains output of dofiles
~/ado/ --- contains ado files needed to run dofiles
~/res/ --- contains questionnaires and codebook

B. VERSION COMPABILITY AND EXTENSIONS

1. Dofiles written for Stata 15/MP using 1.6 Ghz Intel Core i5 (MacOS Mojave)

2. User must install ado files below to run code:
a. lazystar.ado included in ~/ado/ and produces tables
b. _gweightave2.ado included in ~/ado/ and computes Anderson indices

3. Additional STATA packages needed to run dofiles (type SSC install):
a. pdslasso - needed to compute double-machine learning estimates
b. lassopack - needed to compute double-machine learning estimates
c. wyoung - needed to adjust for multiple hypothesis testing (Appendix A9)
d. winsor – needed to produce winsorized variables (multiple tables)

C. DATA FILES

1. ~/raw/bas-mid-end-append.dta – raw data household survey data files from the baseline, midline, and endline surveys. Note, the sheets A-C3 in the codebook (~/res/codebook_ej.xlsx) describe these variables.

2. ~/dta/bas-mid-end-append_v2.dta -- this is the main data file and contains the stacked baseline, midline, and endline data. Note the sheets A-C3 in the codebook (~/res/codebook_ej.xlsx) describe these variables.

3. ~/dta/baseline_lassocleaned.dta – this file contains the variables used in executing the double machine learning/LASSO algorithms. The variables are specified in Appendix A18. The variables themselves are the same as those in the codebook, except they are prefix by 'ls_'.

3. ~/dta/ao_content_data.dta – this data file contains information on the content of both questions asked on the ao platform and push calls delivered. See 'AO Content' in the codebook (~/res/codebook_ej.xlsx) for details.

4. ~/dta/incoming_monthly_data.dta -- this data file contains monthly information incoming AO usage. See 'AO Usage' in the codebook (~/res/codebook_ej.xlsx) for details.

5. ~/dta/ push_call_monthly_data.dta -- this data file contains monthly information on push calls. See 'AO Usage' in the codebook (~/res/codebook_ej.xlsx) for details.

6. ~/dta/wtp_panel_a.dta includes data on WTP experiments (BDM & TIOLI). See WTP in the codebook (~/res/codebook_ej.xlsx) for details.

7. ~/dta/wtp_panel_b.dta -- includes data on WTP experiments (BDM & TIOLI). Similar to 6. (above) but reshaped. See WTP in the codebook (~/res/codebook_ej.xlsx) for details.

8. ~/dta/peer_survey.dta - includes data on the survey administered to peers (both those in the study as well as non-study peers). See Peer in the codebook (~/res/codebook_ej.xlsx) for details.

9. ~/dta/village_census_info.dta -- publicly available Indian census data (2011) for the villages in the study. The data was obtained from SHRUG (2010):
http://www.devdatalab.org/shrug_download/
See 'Census' in the codebook (~/res/codebook_ej.xlsx) for details.

D. REPLICATION INSTRUCTIONS

First, the raw data from all household survey rounds is contained in~/raw/bas-mid-end-append.dta. The dofile ~/do/data_prep.do takes this raw data and produces various variables required for the analysis.

For each of the figures, tables, and appendices cited below, the 'input(s)' describe the data file (stored in ~/dta) needed to produce the output. The 'dofile' specifies the name of the dofile (stored in ~/do) that produces the figures/tables/appendices and 'output(s)' describe the name of the output(s) produced by the code (stored in ~/out)

To run the dofiles, first ensure compliance with section B, and then change the directory path to ~/do/ so that relative references in the dofiles work.

Figure 3 - Usage of Mobile Phone-Based Information by Month

input 1: ~/dta/incoming_monthly_data.dta
input 2: ~/dta/push_call_monthly_data.dta
dofile: ~/do/f3_ej.do
output 1: ~/out/f3_panel_a.pdf
output 1: ~/out/f3_panel_b.pdf

Computational Time: < 1 min

Table 1 – Summary Statistics and Balance

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/t1_ej.do
output: ~/out/t1_ej.xls

Computational Time: < 1 min

Table 2 – Usage of Avaaj Otalo

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/t2_ej.do
output: ~/out/t2_ej.xls

Computational Time: < 1 min

Table 3 – Effects on Sources of Agricultural Information and Knowledge

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/t3_ej.do
output: ~/out/t3_ej.xls

Computational Time: < 1 min

Table 4 - Effects on Summary Indices of Input Adoption

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/t4_ej.do
output: ~/out/t4_ej.xls

Computational Time: < 1 min

Table 5 - Effects on Yield, Demand, and Profit

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/t5_ej.do
output: ~/out/t5_ej.xls

Computational Time: < 1 min

Table 6 - Spillover and Peer Effects

input 1: ~/dta/bas-mid-end-append_v2.dta (study respondents)
input 2: ~/dta/peer_survey.dta (non-study respondents)
dofile: ~/do/t6_ej.do

output 1: ~/out/t6_1_ej.xls (study respondents)
output 2: ~/out/t6_2_ej.xls (non-study respondents)

Computational Time: < 1 min

Appendix A1 - Topics of Questions Asked and Push Calls

Input: ~/dta/ao_content_data.dta
dofile: ~/do/a1_ej.do
output: ~/out/a1_ej.csv

Note: In the 'varnames' for the output files, the prefix 'p' refers to push calls, while 'q' refers to questions. The suffix E refers to endline while M refers to Midline. the prefix 'pct' denotes a percentage. For example, 'pct_p_cotton_E' is the percent of push calls about cotton at endline.
The orientation of the output was changed manually.

Computational Time: < 1 min

Appendix A2 - Randomisation Check

input: ~/dta/bas-mid-end-append_v2.dta (study respondents)
dofile: ~/do/a2_ej.do
output 1: ~/out/a2_baseline_ej.csv
output 2: ~/out/a2_midline_ej.csv
output 2: ~/out/a2_endline_ej.csv

Note: For each csv, the total number of variables is the number of rows minus one (AO variable).
After deleting the last row (AO variable), sum the columns in sig1_1, s1_5, and sig1_10 to

compute the number of variables with imbalances at the 1%, 5%, and 10% level, respectively. The fraction of imbalanced variables to total gives the percentages reported in the appendix.

Computational Time: < 3 mins

Appendix A3 - Effects on Sources of Information by Source and Decision Type

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a3_ej.do
output: ~/out/a3_ej.xls

Computational Time: < 1 min

Appendix A4 - Components of Profit and Input Expenditure

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a4_ej.do
output: ~/out/a4_ej.xls

Computational Time: < 1 min

Appendix A7 - Balance for Peer Regressions

input 1: ~/dta/bas-mid-end-append_v2.dta (study respondents)
input 2: ~/dta/peer_survey.dta (non-study respondents)

dofile: ~/do/a7_ej.do

output 1: ~/out/a7_1_ej.xls (study respondents)
output 2: ~/out/a7_2_ej.xls (non-study respondents)

Computational Time: < 1 min

Appendix A8 - Characteristics of Attritors by Treatment Status

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a8_ej.do
output: ~/out/a8_ej.xls

Computational Time: < 1 min

Appendix A9 - Adjusting P-Values for Family-wise Error Rate

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a9_ej.do
output 1: ~/out/a9_panel_A.csv
output 2: ~/out/a9_panel_B.csv

Computational Time: Each panel can take up to 6 hours to run.

Appendix A10 - Effects on Agricultural Knowledge

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a10_ej.do
output: ~/out/a10_ej.xls

Computational Time: < 1 min

Appendix A11 - Heterogeneous Effects by Education and Income

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a11_ej.do
output: ~/out/a11_ej.xls

Computational Time: < 1 min

Appendix A12 - Main Outcomes by Sub-Treatment Arms

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a12_ej.do
output: ~/out/a12_ej.xls

Computational Time: < 1 min

Appendix A13 - Main Results with Unweighted Indices

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a13_ej.do
output: ~/out/a13_ej.xls

Computational Time: < 1 min

Appendix A14 - Effects on Components of Cotton Index

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a14_ej.do
output: ~/out/a14_ej.xls

Computational Time: < 1 min

Appendix A15 - Heterogeneity with Respect to Village Size

input 1: ~/dta/bas-mid-end-append_v2.dta
input 2: ~/dta/village_census_info.dta
dofile: ~/do/a15_ej.do
output: ~/out/a15_ej.xls

Computational Time: < 1 min

Appendix A16 - Results from WTP Experiments

input 1: ~/dta/wtp_panel_a.dta
input 2: ~/dta/wtp_panel_b.dta

dofile: ~/do/a16_ej.do

output 1: ~/out/a16_panel_a.xls
output 1: ~/out/a16_panel_b.xls

Note: The first row of output 1 produces the average bid. The third row is the number of participants. The 6th row is a count of the number who bought an AO subscription. The table orientation was changed to portrait manually.

Computational Time: < 1 min

Appendix A17 - Yield Results by Survey Round

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a17_ej.do
output: ~/out/a17_ej.xls

Computational Time: < 1 min

Appendix A19 - Effects on Sowing by Survey Round

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a19_ej.do
output: ~/out/a19_ej.xls

Computational Time: < 1 min

Appendix A21 - Main Results by Survey Round

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a21_ej.do
output: ~/out/a21_ej.xls

Computational Time: < 1 min

Appendix A24 - Spillover Effects

input: ~/dta/bas-mid-end-append_v2.dta
dofile: ~/do/a24_ej.do
output: ~/out/a24_ej.xls

Computational Time: < 1 min

Bibliography

Asher, Sam and Lunt, Tobias and Matsuura, Ryu and Novosad, Paul (2019): "The The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG)", Mimeo, available from: http://www.devdatalab.org/shrug_download/