Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes

Data publication – ReadMe

Authors: Gharad Bryan, James J. Choi, Dean Karlan

This folder contains all the data files, code and documents that are necessary to replicate the analysis presented in the paper. Below is a description of each one of the subfolders and its content.

Software needs:

Data, code and user-written commands are compatible with Stata 15 or higher.

1. documents

This folder contains all the documents that will help with the replication of the results presented in the paper. These include the survey instrument, a data map, the pre-analysis plan etc.

2. data

This folder contains the raw data for the study, stripped of PII. This is also the directory where new versions of the data will be saved (clean data and data with outcome variables).

3. ado

This folder contains all the user written commands that are necessary for running the cleaning and analysis.

- fastcd
- labrec
- outreg2
- rmfiles
- sxpose
- winsor
- orthout
- xml_tab
- estrat
- tabmiss
- fsum
- qqvalue
- dm31

4. do

This folder contains all the do files that are needed to replicate the results in the paper – main text and appendix. They are listed in the do-file "0_master" in the order in which they should be run. Each one has a description of what the code does, and what the inputs and outputs are. See data map at "documents\Data map" for a description of the data flow, inputs and outputs of each do files.

The folder structure is set up so all the code can run, once the user sets their path in the do-file "0_master". The do files are:

- "6ms_globals"

Sets globals used for file paths and model specifications.

- "programs"

Defines short programs for tasks that will be repeated throughout the analysis – these are all described in the code.

- "6ms_clean"

Cleans raw data, previously stripped from PII. The specific cleaning tasks performed in this code are described at the beginning of the do file. They include correcting some errors in the data, labeling variables, recoding values, etc. The cleaning done is not exhaustive and does not include all the variables in the data that are not used for the analysis.

- "6ms_outcome_variables"

Generates and groups outcome variables and indices. Most of the outcomes analyzed are grouped in indices according to the methodology described in the paper, which aggregates and standardizes variables according to the mean and sd of the control group. See "Appendix Table 5: Outcome variable construction" in the paper for a full description of how each outcome is constructed.

- "6ms_orthogonality"

Generates and exports orthogonality table for pre-treatment characteristics, collected during the 6^{th} month survey, presented in appendix table 1.

- "6ms_joint_tests"

This do file tests 3 hypothesis across all outcomes to decide if treatment groups can be pooled in order to obtain greater statistical power. The table that is created in this do file are not presented in the paper, but the results are mentioned in in the text.

- "6ms_tables_all"

Calculates treatment effects for primary outcomes, mechanisms and secondary outcomes, presented in tables 1, 2 and 3 the main text of the paper. These results are exported in excel, tables – one for each group of outcomes and for each regression specification. The exported tables also contain the treatment effects for the components of each index or outcome of interest, which appear in the appendix of the paper. Finally the do file save all these results in a dta and an excel file in the form of a matrix. These files are easier to visualize and compare.

- "6ms_qvalues"

Calculates q values - adjusted p values for multiple hypothesis testing - for primary outcomes. The procedure is described in the do file.

- "6ms_qvalues_religion_intrinsic"

Calculates q values - adjusted p values for multiple hypothesis testing - for primary different version of the religion intrinsic index. The procedure is described in do file 9.

- "6ms_tables_switchedcom"

Calculates treatment effects for primary outcomes, mechanisms and secondary outcomes and exports tables for main indices, including communities that switched treatment status (which were deleted from main analysis). These results are presented in the appendix tables 3-5.

- "6ms_qvalues_switchedcom"

Calculates q values - adjusted p values for multiple hypothesis testing - for primary outcomes including communities that switched treatment status. The procedure is described in do file 9.

- "6ms_tables_temptation"

Calculates treatment effects for consumption of temptation goods, which include cigarettes and alcoholic beverages, presented in the appendix table 41. The results are saved in one excel file for each regression specification

- "6ms_tables_income_robusteness"

Calculates treatment effects and exports tables for income robustness checks, presented in appendix table 42.

- "6ms_qvalues_income_robusteness"

Calculates q values - adjusted p values for multiple hypothesis testing - for measures of income. The procedure is described in do file 9.

- "6ms_stats_religion_groups"

Calculates summary stats for religious affiliation by treatment group, and also for the total sample. These are presented in the appendix table 44.

- "6ms_treatment_catholic_protestant"

Calculates the effect of treatment on the probability of being catholic vs protestant, after the program. This is calculated for all the regression specifications. Results are presented in table 7 in the main text of the paper.

- "6ms_tables_nocontrols"

Calculates treatment effects for primary outcomes, mechanisms and secondary outcomes and exports tables for main indices, without controls for respondent characteristics. These results are presented in the appendix tables 9-11.

- "6ms_qvalues_nocontrols"

Calculates q values - adjusted p values for multiple hypothesis testing - for primary outcomes without controls for respondent characteristics. The procedure is described in do file 9.

- "6ms_IV_specification"

Calculates treatment effects and exports tables for primary economic outcomes for OLS and IV regressions of composite religiosity index. These results are presented in appendix table 45.

- "6ms_tables_baselinecontrols"

Calculates treatment effects for primary outcomes, mechanisms and secondary outcomes and exports tables for main indices, controlling for baseline survey characteristics where available. These results are presented in the appendix tables 6-8.

- "6ms_qvalues_baselinecontrols"

Calculates q values - adjusted p values for multiple hypothesis testing - for primary outcomes controlling for baseline survey characteristics where available. The procedure is described in do file 9.

- "6ms_attrition_analysis"

Calculates treatment effects for primary under various missing data assumptions, and predicts atrition at 6 months from baseline characteristics. These results are presented in appendix table 2 and 43.

- "6ms_endogenous_stratification"

Calculates treatments effects by terciles of predicted religiosity for primary outcomes, where terciles are predicted by endogenous stratification. Results are presented in table 4 in the main text of the paper.

- "6ms_social desirability"

Tests for presence of social desirability bias by applying the technique described in Coffman et al (2017) that compares list randomized responses to direct elicitation. Results are presented in table 5 in the main text of the paper.

The do files run in stata 15 and higher. For older versions of stata, the last part of do file #7 needs to be commented out (this is indicated in the code). The do files for the Endline analysis are a subset of the 6-month do-files listed above, except one:

- "EL_pvalues_6mo_v_EL"

Calculates the p-values testing whether the treatment effects are the same at 6 months and 30 months for V and Any V.

5. Results

NOTE: There may be a possibility of finding some minor discrepancies between the replication output and the publication results due to versioning issues, however, the statistical significance of the results remain unchanged.

For the regressions that use the outcome "no domestic violence, list-rand." or its components, the user will have to add 1 to the control mean outputted by the code to get the control mean reported in the paper (in Tables III and XIX, and Appendix Tables 5, 8, 11, 39, 50, 53, 80). This is because the sign of this

outcome was reversed so that it measures the absence of domestic violence. Therefore, the control mean outputted by the regressions is the negative of the prevalence of domestic violence, so adding 1 to the control mean gives the share of respondents for which there is no domestic violence in the household.

The "6 months" and "Endline" results subfolders consist of:

- 1. Balance&Orthogonality

This folder contains the results for appendix table 1 (6 months) and 46 (Endline)

- 2. Pooling tests

This folder contains he results for the joint tests that helped decide to pool treatment groups together during analysis at 6 months (folder not present in endline analysis).

- 3. Main results

This folder contains the results for all main outcomes of interest, including index components. The subfolder "Tables" contains results in the form of separate tables for each regression specification. The folder "Matrices" contains results exported as matrices, with one file for each regression specifications. These include all treatment effects presented in tables 1-3 and appendix tables 12-40 for 6 months analysis and tables 6, 8, 9 and appendix tables 54-81 for endline analysis.

- 4. Additional stats

This folder contains the additional statistics needed for the tables (p values and q values). The folder "q values" contains the adjusted q values for different tables (main text and appendix tables). The folder "tests" contains the p values that result from the additional tests for the equality of treatment effects.

- 5. Appendix

This folder contains regression results for the appendix. These are classified into

  - Income: contains results presented in appendix table 42 (6 months) and 83 (Endline) with income robustness checks;
  - Main outcomes - No drop: Treatment effects for main outcomes without dropping communities that switched treatment, presented in appendix tables 3-5 (6 months) and 48-50 (Endline);
  - Temptation goods: Contains treatment effects for consumption of temptation goods (cigarettes and alcoholic beverages), presented in appendix table 41 (6 months) and 82 (Endline);
  - No Controls: Treatment effects for main outcomes without controls for respondent characteristics, presented in appendix tables 9-11 (6 months) and 51-53 (Endline);
  - Attrition analysis: Contains two items. (1) Prediction of attrition at 6 months and 30 months using predictors from baseline and 6 months respectively, presented in appendix tables 2 (6 months) and 47 (Endline), and (2) treatment effect bounds for primary outcomes calculated under varying assumptions for missing data, presented in appendix tables 43 (6 months) and 84 (Endline);
  - Baseline controls: Treatment effects for main outcomes controlling for baseline survey characteristics where available, presented in appendix tables 6-8 (6 months only);

- Religion: contains summary statistics by religion affiliation and treatment status presented in appendix table 44 (6 months only);
- IV spec: Contains OLS and IV regressions of composite religiosity index on primary economic outcomes, presented in appendix table 45 (6 months only).