

README: Replication files for “The effect of survey mode on data quality: Experimental evidence from Nigeria”

Overview

This repository contains the necessary files for replicating the analysis in the study on “The effect of survey mode on data quality: experimental evidence from Nigeria” by Pauline Castaing, Ivette Contreras, Yannick Markhof, Amparo Palacios-Lopez, Akiko Sagesaka, and Philip Wollburg.

Following the completion of fieldwork, the Raw Data of these surveys were cleaned to improve the quality and also ensure that all Personal Identification Information (PII) were removed from the public version to maintain the confidentiality of the respondents. The code in this replication package prepares the analytical data and generates applicable Tables and Figures included in the manuscript, mostly pulling data files from the public versions of these two datasets.

Requirements

Memory and Runtime Requirements: The Stata analysis code approximately requires 10 minutes to execute completely.

Software Requirements: Stata (Tested on version 18)

Required Stata packages:

- outreg2
- estout
- coefplot
- swindex
- mmerge
- ietoolkit

Instructions for replicators

1. Prepare Environment: Open the master dofile (MasterDofile_ModeEffectsNigeria.do). Ensure user settings, including unique user number and OneDrive paths, are correctly configured.
2. Install Packages: If necessary, install user-written packages by modifying line 72 in the master dofile (change 0 to 1). This is a one-time setup per computer.
3. Set Repository Path: Modify line 57 to reflect the path of the repository on your computer.

4. Execute Analysis: The master dofile (MasterDofile_ModeEffectsNigeria.do) can be used if one wants to run all the dofiles with just a touch of a button.
5. Generate Outputs: Running the master dofile will construct the main datasets for analysis and generate all tables and graphs from the working paper.
6. Output Location: Outputs will be stored in \Outputs\ Figures and in \Outputs\Tables. Main analysis outputs are in the 'Main' folder, and appendix outputs in the 'Appendix' folder.

Datasets: Description and Availability Statement

Summary of availability: Some data restricted.

1. General Household Survey, Panel 2023-2024 – Mixed-Mode Study

Public Availability: Data from this study is yet to be made public. The Public Datafiles will be available in the World Bank Microdata library on July 2026 where users will be able to download and use for any applicable research. While these datasets can be used for reproducibility verification, they cannot be shared publicly on any platform at this stage. The data from the enumerators' questionnaire (\Data\Raw\Mixed mode study\MM5\mm5_interviewers.dta) will not be released due to sensitivity concerns arising from the very small number of observations.

Source: World Bank. The data for this randomized survey experiment were collected by the National Bureau of Statistics in Nigeria as part of the Nigeria's General Household Survey Panel Wave 5 (GHS-Panel W5) 2023/24.

Description: The dataset includes five rounds of data collection, labeled MM1 through MM5, as well as data from enumerator interviews. The households included in this study constitute a subsample of those interviewed in the GHS-Panel Wave 5.

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.

2. Nigeria's General Household Survey Panel Wave 5

Public Availability: Data from this study is publicly available at the following link: <https://microdata.worldbank.org/index.php/catalog/6410/study-description>. It was last accessed in December 2024. The dataset on aggregated consumption expenditure data (Data\Raw\GHS-Panel W5\totcons.dta) and on languages used for the interview (Data\Raw\GHS-Panel W5\secta_harvestw5_lang_data.dta) will be forthcoming in the coming months in the microdata library. The data from the enumerators' questionnaire (\Data\Raw\ Data\Raw\GHS-Panel W5\GHS-Panel interviewers surveys) will not be released due to sensitivity concerns arising from the very small number of observations.

Source: World Bank. The GHS-Panel is the result of a partnership that the Nigerian National Bureau of Statistics has established with the Bill and Melinda Gates Foundation (BMGF) and the World Bank (WB).

Description: The General Household Survey-Panel (GHS-Panel) is implemented in collaboration with the World Bank Living Standards Measurement Study (LSMS) team as part of the Integrated Surveys on Agriculture (ISA) program. The objectives of the GHS-Panel include the development of an innovative model for collecting agricultural data, interinstitutional collaboration, and comprehensive analysis of welfare indicators and socio-economic characteristics. The GHS-Panel is a nationally representative survey of approximately 5,000 households, which are also representative of the six geopolitical zones. The 2023/24 GHS-Panel is the fifth round of the survey with prior rounds conducted in 2010/11, 2012/13, 2015/16 and 2018/19. The GHS-Panel households were visited twice: during post-planting period (July - September 2023) and during post-harvest period (January - March 2024).

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.

3. HFPS harmonized dataset

Public Availability: This dataset is publicly available at the following link: https://github.com/lms-worldbank/HFPS_harmonization_working/blob/main/Final%20datasets/HH_%204%20Dec%202024.zip. It was last accessed in December 2024.

Source: World Bank.

Description: The LSMS has made available an open-access, harmonized dataset on High-Frequency Phone Surveys. The dataset covers all LSMS-HFPS survey rounds and allows linkages at the household and individual levels over time. It can be integrated with the LSMS-Integrated Survey on Agriculture (LSMS-ISA) series of nationally representative face-to-face household surveys in the same countries, which include geovariables and cluster locations.

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.

DoFile Structure and Execution Order

Master Dofile (MasterDofile_ModeEffectsNigeria.do): The primary script for running all other dofiles. Run this first for complete replication following steps in instructions for replicators.

The two main sections of the do files are as follows:

Part 1. Data Import and Cleaning (Cleaning/Part1_MasterDataCleaning.do): While direct access to the raw dataset may not be available, a detailed do-file has been prepared for its cleaning and preparation. This do-file efficiently imports the dataset into Stata and

performs a series of essential tasks to prepare the data for variable construction and subsequent analysis.

Part 2. Data Analysis (/Analysis/Part2_MasterDataAnalysis.do): Performs main and appendix analysis, including descriptive statistics, regression tables, and graphs. All outputs are stored in specified directories (\Outputs\Tables and Outputs\Figures).

List of analysis codes and outputs

All code can be executed from the Master.do script, providing a centralized approach for running the full analysis. Additionally, codes for individual exhibits are designed to be executed independently, with the condition that folder globals and custom programs are correctly set up in both the master dofile and the respective section dofiles. This flexibility ensures the accurate reproduction of all numerical data presented in the paper. For ease of reference, a table is included which specifies the name of each output as mentioned in the paper, its location, the program responsible for its execution, and the individual output's file name in the folder.

Main Analysis:

Folder path for individual dofiles: \$dofiles/Analysis/Sections		
Exhibit	Program	Individual output file
Table 1 Panel A	Section_4_AverageModeEffects.do	Table_1a.csv
Table 1 Panel B	Section_4_AverageModeEffects.do	Table_1b.csv
Table 2	Section_5_Heterogeneity.do	Table_2.rtf
Table 3	Section_5_Heterogeneity.do	Table_3.rtf
Figure 1	No syntax	
Figure 2	No syntax	
Figure 3	Section_4_AverageModeEffects.do	Figure_3.png
Figure 4	Section_5_Heterogeneity.do	Figure_4.png
Figure 5	Section_5_Heterogeneity.do	Figure_5.png
Figure 6	Section_5_Heterogeneity.do	Figure_6.png
Figure 7	Section_5_Heterogeneity.do	Figure_7.png

Appendix Analysis:

Folder path for individual dofiles: \$dofiles/Analysis/Sections		
Exhibit	Program	Individual output file
Table A1	Section_3_Experimental_Design.do	Table_A1.xlsx
Table A2	No syntax	
Table A3	Section_3_Experimental_Design.do	Table_A3.xlsx

Table A4	Section_3_Experimental_Design.do	Table_A4.rtf
Table A5	Section_5_Heterogeneity.do	Table_A5.rtf
Table A6	Section_3_Experimental_Design.do	Table_A6.rtf
Table A7	Section_3_Experimental_Design.do	Table_A7.rtf
Table A8 panel A	Section_3_Experimental_Design.do	Table_A8a.rtf
Table A8 panel B	Section_3_Experimental_Design.do	Table_A8b.rtf
Table A9	Section_5_Heterogeneity.do	Table_A9.xlsx
Table A10	Section_6_SurveyErrors.do	Table_A10.rtf
Table A11	Section_6_SurveyErrors.do	Table_A11.rtf
Figure A1	Section_3_Experimental_Design.do	Figure_A1.png
Figure A2	Section_3_Experimental_Design.do	Figure_A2.png
Figure A3	Section_4_AverageModeEffects.do	Figure_A3.png