

Grain Today, Gain Tomorrow: Evidence from a Storage Experiment with Savings Clubs in Kenya

Shilpa Aggarwal* Eilin Francis[†] Jonathan Robinson[‡]

February 28, 2018

This document describes the data used to generate the main results in the paper. We include 6 datasets and the 5 main surveys we use in the analysis. All analysis was conducted in Stata, and we include the do file which will replicate all the main results (SEAR_master_results_28feb18.do).

1 Datasets

1.1 Main merged dataset (SEAR_analysis_working_28feb18.dta)

This dataset merges all the key pieces of data and is used for the bulk of the analysis. This includes 6 main surveys/datasets, described below. Variables from each survey are prefixed with a different letter:

1. ROSCA-level baseline is prefixed with rb_
2. Individual baseline is prefixed with b_
3. Individual midline is prefixed with m_
4. Individual endline is prefixed with e_
5. ROSCA-level endline is prefixed with er_
6. Coupon redemption data is prefixed with r_

Variables are named according to the question number in the relevant survey. We have included labels for most variables; for those that are not clear or not labeled, please refer to the surveys.

1.2 Sales data (sales_module_FINAL.dta)

This dataset contains all recorded sales from the endline survey, reshaped at the transaction level. For each transaction, the dataset reports the date, price and quantity (in goro-goros, a volume measure in Western Kenya equivalent to roughly 2.2 kg) of the sale.

*Indian School of Business, email: shilpa_aggarwal@isb.edu

[†]University of California, Santa Cruz, email: elfranci@ucsc.edu

[‡]University of California, Santa Cruz and NBER, email: jmrtwo@ucsc.edu

1.3 Prices data (prices_by_month.dta)

This has data, reshaped by month, for three measures of prices: (1) expectations from the baseline survey; (2) realized prices from the sales records; and (3) recorded prices by local shops.

1.4 Logbook data (logbook_ledger.dta)

This dataset includes ISRA and GSRA transactions, at the transaction level. This includes the values of all deposits and withdrawals, and the date of each transaction.

1.5 ROSCA-level compliance (ROSCA_compliance_FINAL.dta)

This dataset has information on every sampled ROSCA and whether they were traced for the intervention. This is used to calculate ROSCA-level attrition.

1.6 Secondary prices from nearby city of Kisumu (prices_Kisumu.dta)

We compiled secondary price data from online databases compiled by the FAO, the WFP, and RATIN. This dataset shows the prices in each of these datasets, by date, for the city of Kisumu, Kenya (the closest city to our project site with sufficient coverage in those datasets).

2 Surveys

We include soft copies of 5 surveys:

- Individual baseline: the baseline survey administered at the first ROSCA meeting, starting in August 2015
- ROSCA baseline: a ROSCA-level survey conducted concurrently with the individual baseline
- Individual midline: a survey conducted in early 2016
- Individual endline: the final follow-up survey, which was collected starting in August 2016
- ROSCA endline: a ROSCA-level endline collected concurrently with the individual endline