Contents

1 Introduction ................................................. 2
   1.1 What is KGFS? ........................................... 2
   1.2 The Intervention ......................................... 2
       1.2.1 Implementing Partner, KGFS ..................... 2
       1.2.2 Study Geography .................................. 2
       1.2.3 Level of Randomization ........................... 3

2 Survey Overview ............................................ 4
   2.1 Survey Data ............................................. 4
   2.2 Timeline ............................................... 5
   2.3 Sampling Methodology ................................. 5

3 Survey Instrument and Content ................................ 6
   3.1 Electronic Survey Forms ............................... 6
   3.2 First Visit Survey ..................................... 6
   3.3 Follow-up Survey ...................................... 7
   3.4 End Call Survey ....................................... 9

4 Data Entry and Cleaning .................................... 9
   4.1 PII Data ................................................ 9

5 File Structure .............................................. 10
   5.1 Identifier variables .................................. 10
   5.2 Non-response codes ................................... 10
   5.3 Codebook ............................................. 10
1 Introduction

This User Reference describes the Daily Consumption Survey (DCS) study, file structure for first visit, follow up call, and end call. This note is intended as an easy reference and introduction to the DCS dataset. **Final cleaned dataset is in a long format, at the household level.**

1.1 What is KGFS?

The KGFS project is - in as few words as possible - a cluster randomized controlled trial seeking to assess the economic impact of expanding financial access in rural India. The base of the intervention, or the “treatment”, is the random expansion of bank infrastructure across rural villages in 3 districts of Tamil Nadu, India. We seek to evaluate the “first-stage” impact of this exposure on households’ loan, savings, and insurance taking behaviour, but also the impact of the intervention on a wide range of outcomes that collectively reflect household well-being (e.g. household income, occupational choice, consumption) as well as individual well-being (e.g. network centrality and connectedness, women empowerment, health).

The acronym we use, KGFS, is the name of the banking model used by our implementing partner (the large financial institution, or LFI, that builds and operates the banks) on this project. It stands for “Kshetriya Gramin Financial Services,” which translates as “Regional Rural Financial Services.” We colloquially refer to this project as a whole as “the KGFS project” or just “KGFS.”

1.2 The Intervention

1.2.1 Implementing Partner, KGFS

KGFS states that its mission is to “maximize the financial wellbeing of every individual and every enterprise in remote rural India by providing complete financial services.” Their model can be viewed as an alternative to the standard microfinance movement in India, which has focused primarily on microcredit, or on urban areas:

- In terms of products, KGFS offers formal financial products such as loans, savings, and insurance.
- In terms of outreach, KGFS offers tailored financial advice provided through local village branches, in order to effectively reach individuals in financially marginalized rural communities.

1.2.2 Study Geography

The study takes place in three districts of Tamil Nadu: Thanjavur, Thiruvarur, and Pudukkotai. Figure 1 shows these districts. The figure below represent the different administrative levels we usually refer to and the corresponding variables in our datasets.
1.2.3 Level of Randomization

KGFS agreed to open branches in eligible areas, i.e. those that satisfied certain conditions, such as the size of the population near the branch site or the saturation of competitors in the area. As a result, in 2009, we started by identifying 101 viable “service areas”, i.e. areas that satisfied KGFS’ criteria, across the three districts.

After KGFS had approved of the identified service areas, we paired comparable service areas together (using a minimum distance matching algorithm). This created 50 pairs.\(^1\) Within each pair, we randomized which service area would receive a KGFS branch first.\(^2\)

In sum, randomization is done at the service area level. Figure 2 shows the resulting treatment and control service areas.

---

\(^1\)One “pair” is in fact a triplet, consisting of one treatment area and two control areas.

\(^2\)The idea was that eventually, KGFS would open branches in all 101 areas (given that they operate across Tamil Nadu) - but by delaying the opening of branches in control areas, we were given a window in which to evaluate the impact of the branch openings in treatment areas versus control areas.
2 Survey Overview

2.1 Survey Data

A huge amount of survey data was collected, in part due to multiple surveys being conducted simultaneously. Figure 3 shows the overlapping timelines of the major surveys. Each survey, which we refer to as “survey components” or simply “components”, focuses on measuring certain aspects of household and individual life.

It is important to note that each survey was different in terms of the type of information being collected: for example, one asks specifically about household expenditures, while another asks about the household’s informal networks. Thus, it sometimes made sense to conduct different surveys with the same household, in order to collect a fuller set of information on the household. As a result, there are cases where the set of households sampled for one survey overlaps with the set of households sampled for another survey.

In this user reference guide, everything that we are referring to is only for KGFS Daily Consumption Survey (DCS) dataset.
2.2 Timeline

The consumption survey aims to capture income and expenditure flows of our respondents. The consumption survey consists of three main components:

- **First Visit Survey**
- **Phone Follow Up Survey**
- **End Call Survey**

It is primarily a phone follow up survey. The survey was conducted in 2016, below is a timeline of the different components and rounds.

<table>
<thead>
<tr>
<th>Survey Rounds</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3 Sampling Methodology

DCS was conducted in Social Network Mapping (SNM) villages across 32 pairs. The survey rolled out in 34 pairs, but was later restricted to 32 pairs. Pair 40 and 37 were dropped in Round/Month 2. Location Id 21143407 was also

---

3 The SNM survey is a part of a broader effort to collect data on households’ individual social networks in the village they live. In each service area, one village was randomly picked with the following criteria: (1) The village population is comprised of between 50 to 170 households using data on panchayat officials’ best estimate of the village population as population measure; (2) the village lies between 1 to 3 kilometers from the designated branch (or center) village.

4 KGFS required us to release the pairs immediately and the households belonging to these pairs were dropped. The consumption survey has been completed for the households in Round/Month 1 for these pairs.
dropped during Round/Month 3 since the village did not have electricity (only 7 households out of 120 had electricity which is a requirement for phone follow up surveys. No household in this location id was part of the survey). On average, there are 4 gramums per pair (or 2 gramums per service area).

The household head and the spouse of the household head were surveyed every 48 hours. Preparation for the survey began in August 2015 and the survey rolled out in February 2016. The total sample size is 1224 households. This works out to be roughly 38 households per pair, 18 households per service areas or 9 households per gramum. Due to operational limitations, only 200 - 250 households can be followed in a month so the sample was split into 5 sets (Round or Month 1-5) This also enables us to capture seasonal variations across months. The households were first sampled at the gramum level depending on the number of households per gramum. The sampled households were then randomly assigned over 5 months.

Households were administered the first visit survey. Once consent is obtained from the head of the household and the spouse of the head of the household, they are assigned to the follow up survey. 15 follow up surveys are administered over 30 days after which the end call survey is administered.

Resampling was carried out if:

- the head of the household was older than 60 years.
- migration due to pregnancy or health reasons.
- the respondent had an hearing issue.

3 Survey Instrument and Content

3.1 Electronic Survey Forms

Each month had a different form since it was easier to keep track. The syntax was CS_First_Visit_Survey_Month\'x\' for the first visit forms. For the final dataset, all the rounds are combined.

3.2 First Visit Survey

The First Visit Survey included the following modules: household roster, farming, existing self-employment, marine and livestock, labor, shocks, loans and lenders, savings and insurance, assets, consumption, household decision making.

1. Introduction: The introduction section asked the basic information from the respondent including district, block, panchayat, street, door number. All these information is personal identifiable and are not accessible in the public dataset. This section also take the consent from the respondent, and only if the respondent agree to consent, the survey will continue.

2. Household Roster: The household roster module profiles each household. It surveys the gender and age of household members, identifies the head of household, and provides information on the education level, religion, and caste of each household members.

3. Farming: The farming module provides information on the number of crops, primary crop, total cultivated land area, and farming household members for those households that engage in farming.

4. Existing Self Employment: The existing self employment module provides information on non-farming, self-employed households such as the type of self-employment business and the household members involved in it.
5. **Marine and Livestock**: The marine and livestock module surveys the income generated from households engaged in fishing and livestock ownership, as well as the household members involved in the business.

6. **Labor**: The labor module identifies household members engaging in non-household employment (labor or service). It details the type of labor, hours worked, and total income.

7. **Shocks**: The shocks module identifies the monetary cost of various shocks (death of money earner, floods, droughts, pest infestations, serious injury, and unexpected unemployment).

8. **Loans and Lenders**: The loans and lenders module characterizes any financial loaning or lending activity. The amount, source (bank, NGO, close friend or relative, etc), purpose, repayment structure (principal, interest, and collateral), and originator of the loan/lend are identified.

9. **Savings and Insurance**: The savings and insurance module details the amount of and location of any savings, defined as money stored outside of the household. It also characterizes the type of and provider of insurance, such as life insurance, health insurance, crop insurance, accident insurance, etc.

10. **Assets**: The assets module provides the value of all assets in rupees. These assets may include farm equipment, jewelry, electronics, vehicles, livestock, house, or land.

11. **Consumption**: The consumption module identifies spending and consumption patterns of rice, oil, pulses, and other consumption items.

12. **Household Decision Making**: The household decision making module identifies which household member makes the final decision on consumption spending, healthcare spending, education spending, savings, and borrowing.

13. **Contact Details**: The contact detail section is the final section in the first call survey. It takes the consent from the participants for the follow up call (calling every 48 hours to ask questions related to household income and expenditure), and also asked them the time preferred for the call, and take their phone numbers (phone number is personal identifiable information, thus not available in the public data).

### 3.3 Follow-up Survey

The Phone Follow-up Survey included the following modules: food expenditure, non-food expenditure, monetary transfers, in-kind transfers, monetary income, in-kind payments, loans and savings, shocks, other income, other received transfers, other expenditure, other transfers.

The phone follow up survey is ideally administered to the head of the household and the spouse of the head of the household. The second best scenario is to have at least one person in the household as part of our phone follow up survey. Each respondent is administered 15 calls over 30 days. If a respondent does not answer, a no-call form is submitted.

1. **Introduction**: The introduction section asked the basic information from the respondent including district, block, panchayat, street, door number. All these information is personal identifiable and are not accessible in the public dataset. This section also take the consent from the respondent, and only if the respondent agree to consent, the survey will continue.

2. **Food Expenditure**: The food expenditure module specifies the quantity and price of purchased food items in a 48 hour period. The listed food items are comprehensive and include vegetables, meat/fish, eggs, sweets, rice, tobacco, oils, etc.
3. **Non-Food Expenditure**: The non-food expenditure module specifies the quantity and price of purchased non-food items in a 48 hour period. The listed non-food items are comprehensive and include health, housing, education, agriculture, festivals/functions, clothing, etc.

4. **Monetary Transfers**: The monetary transfers module details all the monetary transfers made in a 48 hour period. The recipients of the transfers and the repayment structures are recorded.

5. **In-Kind Transfers**: The in-kind transfers module details all the in-kind transfers made in a 48 hour period. The recipients of the transfers and the good(s) transferred are recorded.

6. **Monetary Income**: The monetary income module specifies all monetary income received in a 48 hour period. The source of the income is specified. These include wage labor, rental income, salary, asset sale, loan repayments, etc.

7. **In-Kind Payments**: The in-kind payments module specifies all in-kind payments received in a 48 hour period. The monetary value and source of the received goods are specified. The sources include wage labor, rental income, salary, asset sale, loan repayments, etc.

8. **Monetary Transfers**: The monetary transfers module details all the monetary transfers received in a 48 hour period that do not need to be repaid. The sources of the transfers are recorded.

9. **In-Kind Transfers**: The in-kind transfers module details all the monetary transfers received in a 48 hour period. The sources of the transfers and the good(s) transferred are recorded.

10. **Loans and Savings**: The loans and savings module characterizes the nature of received financial loans and savings. The amount, source (bank, NGO, close friend or relative, etc), purpose, repayment structure (principal, interest, and collateral), and repayments in the past 48 hours are recorded. The amount and location of savings (post office account, government cooperative, private bank, NGO, etc) are also specified.

11. **Shocks**: The shocks module identifies the monetary cost of various shocks (death of money earner, floods, droughts, pest infestations, serious injury, and unexpected unemployment) that took place in the past 48 hours.

12. **Other Income**: The other income module identifies other household members who earned income (and the amount of income) in the past 48 hours.

13. **Other Received Transfers**: The other received transfers module identifies other household members who received transfers (and the amount transferred) in the past 48 hours.

14. **Other Expenditure**: The other expenditure module identifies other household members who spent money in the past 48 hours and the nature of the expenditure (health, housing, education, agriculture, festivals/functions, clothing, etc).

15. **Other Transfers**: The other transfers module identifies other household members who transferred money in the past 48 hours as well as the amount transferred and the recipient of the transfer.

16. **Work Expectations**: The work expectations module identifies how much money the subject expects to earn in the next 48 hours.
3.4 End Call Survey

The End Call survey included the following modules: farming, existing self-employment, marine and livestock, labor, shocks, loans and lenders, savings and insurance.

1. **Introduction:** The introduction section asked the basic information from the respondent including district, block, panchayat, street, door number. All these information is personal identifiable and are not accessible in the public dataset. This section also take the consent from the respondent, and only if the respondent agree to consent, the survey will continue.

2. **Define Household Member:** The define household member module identifies any members that have exited or entered the household in the past 30 days. Household is defined as those whom the subject ‘shares a roof’ with.

3. **Farming:** The farming module provides information on the number of crops, primary crop, total cultivated land area, and farming household members for those households that engaged in farming in the past 30 days.

4. **Existing Self-Employment:** The existing self employment module provides information on non-farming, self-employed households such as the type of self-employment business and the household members involved in it for those households that engaged in self-employment in the past 30 days.

5. **Marine and Livestock:** The marine and livestock module surveys the income generated from households engaged in fishing and livestock ownership, as well as the household members involved in the business for those households that engaged in marine and livestock in the past 30 days.

6. **Labor:** The labor module identifies household members engaging in non-household employment (labor or service) in the past 30 days. It details the type of labor, hours worked, and total income.

7. **Shocks:** The shocks module identifies the monetary cost of various shocks (death of money earner, floods, droughts, pest infestations, serious injury, and unexpected unemployment) for those households that incurred shocks in the past 30 days.

8. **Loans and Lenders:** The loans and lenders module characterizes any financial loaning or lending activity in the past 30 days. The amount, source (bank, NGO, close friend or relative, etc), purpose, repayment structure (principal, interest, and collateral), and originator of the loan/lend are identified.

9. **Savings and Insurance:** The savings and insurance module details the amount of and location of any savings, defined as money stored outside of the household, in the past 30 days. It also characterizes the type of and provider of insurance, such as life insurance, health insurance, crop insurance, accident insurance, etc.

4 Data Entry and Cleaning

The survey was collected using SurveyCTO platform, and data was downloaded from SurveyCTO. High Frequency Checks were done everyday when collecting the data, and data entry errors are extremely rare. The raw data were corrected if there are errors, and cleaned to prepare for a public use version of the data.

4.1 PII Data

Personal Identifiable Data (PII): You will not see any PII data in the cleaned dataset. All PII variables were either removed from the dataset, or values being replaced to ”PII Data”.
5 File Structure

5.1 Identifier variables

The two main identifier variables in the DCS data are:

- **hhid**: Main household identifier
- **memid**: Person number from the household roster

Other important variables:

- **cont.s.id**: this is a combination of location id which can help to identify which district, block, gram panchayat and village the respondent is located at. It can also help to identify if a respondent located in treated / control village.

5.2 Non-response codes

Non-response codes are straightforward and are present in all cases where the respondent did not provide a response (for example, ”do not know” or ”refuse to answer”). These codes have been labelled in the dataset with the following conventions:

- **-111**: Parents do not cooperate (.a)
- **-333**: Skipped (.b)
- **-555**: Child not able to cooperate (.c)
- **-666**: Not applicable (.d)
- **-777**: Refused to answer (.e)
- **-888**: Other (.f)
- **-999**: Do not know (.g)

5.3 Codebook

An codebook was prepared using Stata and output all the variable names and value numbers from the cleaned data. You can find the codebook with name **DCS Codebook.xlsx**