NATIONAL BUREAU OF STATISTICS

# Basic Information Document

National Panel Survey (NPS 2010-2011)

2010-2011

[UNITED REPUBLIC OF TANZANIA]

**ACRONYMS**

| | | |
|---|---|---|
| BMGF | - | Bill & Melinda Gates Foundation |
| DECRG | - | Development Economics Research Group |
| DFID | - | United Kingdom Department for International Development |
| EA | - | Enumeration Area |
| HH | - | Household |
| HHID | - | Household identification variable |
| JICA | - | Japan International Cooperation Agency |
| LSMS-ISA | - | Living Standards Measurement Study-Integrated Surveys on Agriculture |
| MDG | - | Millennium Development Goal |
| MKUKUTA | - | National Strategy for Growth and Reduction of Poverty |
| NBS | - | Tanzania National Bureau of Statistics |
| THBS | - | Tanzania Household Budget Survey |
| TZNPS | - | Tanzania National Panel Survey |
| UNFPA | - | The United Nations Population Fund |
| UNICEF | - | The United Nations Children's Fund |

# Table of Contents

## Contents

## Background

The 2010/2011 Tanzania National Panel Survey (TZNPS) is the second in a series of nationally representative household panel surveys that collect information on a wide range of topics including agricultural production, non-farm income generating activities, consumption expenditures, and a wealth of other socio-economic characteristics. Both the first and second rounds of the TZNPS were implemented by the Tanzania National Bureau of Statistics (NBS). The first round of the survey was conducted over twelve months from October 2008 to September 2009. The main fieldwork of the second round of the TZNPS started in October 2010 and completed in September 2011, with specialized tracking teams remaining in the field until November 2011. The third round of the TZNPS is scheduled to start in the fall of 2012.

The main objective of the TZNPS is to provide high-quality household-level data to the Tanzanian government and other stakeholders for monitoring poverty dynamics, tracking the progress of the Mkukuta poverty reduction strategy[1], and to evaluate the impact of other major, national-level government policy initiatives. As an integrated survey covering a number of different socioeconomic factors, it compliments other more narrowly focused survey efforts, such as the Demographic and Health Survey on health, the Integrated Labour Force Survey on labour markets, the Household Budget Survey on expenditure, and the National Sample Census of Agriculture. Secondly, as a panel household survey in which the same households are revisited over time, the TZNPS allows for the study of poverty and welfare transitions and the determinants of living standard changes

NBS was advised on technical issues related to survey design and implementation by the TZNPS Technical Committee, which included representatives from line ministries, government agencies and development partners, such as the Ministry of Agriculture, Food Security and Cooperatives, Ministry of Finance, Millennium Challenge Account - Tanzania, World Bank, DFID, UNICEF, UNFPA, and JICA. The majority of funding for the second wave of the National Panel Survey has been provided by a grant from the European Commission. Additional complementary funding for targeted activities (including data entry, supervision and tracking) and technical assistance has been provided by the World Bank through the Living Standards Measurement Study - Integrated Surveys on Agriculture [LSMS-ISA[2]] program.

This document describes all aspects of the TZNPS 2010/11, including the set of survey instruments, sample design, survey implementation, and the resulting data sets.

[1] MKUKUTA is a Kiswahili acronym for the National Strategy for Growth and Reduction of Poverty. The current phase began in 2005 and will finish in 2010, and is the government strategy to meet the MGD and other national development goals. The focus of this round of MKUKUTA includes growth and the reduction of poverty, improved quality of life and social well-being, and governance and accountability.

[2] The Living Standards Measurement Study is an ongoing research initiative within the Development Economics Research Group of the World Bank with the goal of promoting and improving the collection of household level data in developing countries around the world. Further information can be found at www.worldbank.org/lsms. The LSMS-ISA project supports governments in seven Sub-Saharan African countries to generate nationally representative, household panel data with a strong focus on agriculture and rural development. Further information can be found at www.worldbank.org/lsms-isa.

## Survey Instruments

The second round of the TZNPS consists of four survey instruments: a Household Questionnaire, Agriculture Questionnaire, Fishery Questionnaire, and a Community Questionnaire.

The Household Questionnaire is comprised of thematic sections. A detailed description of the contents of the questionnaire can be found in Table 1. This comprehensive questionnaire allows for the construction of a full consumption-based welfare measure, permitting distributional and incidence analysis. This project also recognizes the imperative to look beyond the household as a unit of analysis in order to improve the quality, relevance and sustainability of agricultural data systems. Although data collection is structured around a household panel survey, the data on labor, education, and health status were collected at the individual level. Moreover, in some household activities (like non-farm enterprise), the questionnaire records which specific members are engaged in the activity.

The Agricultural Questionnaire collects information relative to a household's agricultural activities. Information is collected at both the plot and crop level on inputs, production and sales. Table 2 provides a detailed description of the contents of the questionnaire. This questionnaire was administered to any household that engaged in any farming or livestock holding.

The Fisheries Questionnaire was developed in partnership with the World Fish Program to collect data on household fishery activities, fish processing, and fish trading. This includes data on the inputs, outputs, labour, and sales. All this data is divided into two reference periods, the high and low season. This data is collected at the household level. Table 3 provides a more comprehensive list of the sections found within the Fishery Questionnaire.

The Community Questionnaire collects information on physical and economic infrastructure and events in surveyed communities[3], as described in Table 4. In each selected survey community, key informants are interviewed by the field team supervisors. Information about the respondents for the community questionnaire is collected individually in section CI of community questionnaire.

The questionnaires were developed in collaboration with line ministries and donor partners, including the Technical Committee, over a period of several months. The NBS solicited feedback from various stakeholders in regards to survey content and design. The round two questionnaires were piloted in the Morogoro region in June 2010, in conjunction with supervisor training. After piloting, the questionnaires were further revised and finalized by August 2010. Questionnaire manuals were developed with detailed instructions for field staff during training and as the main survey reference guide over the course of the field work.

[3] Note that this is not a "community" in the sociological sense, but rather a mechanism to collect information about the areas where the households selected for the survey are located. In most rural areas, EAs are defined by village boundaries and therefore community refers to the village. In urban areas the distinction is less clear, and occasionally single community questionnaires were administered to adjoining EAs. Therefore, the number of EAs and community questionnaires will not identical.

## Table 1: Household Questionnaire

| Section | Name | Level of observation | Data files | Unique Identifier | Description |
|---|---|---|---|---|---|
| A | Household Identification | Household | HH_SEC_A | y2_hhid | household identifier variables, weights, cluster identification, strata identification, 2008/2009 household id, enumerator, supervisor, data entry clerk identifiers, and date and time of interview |
| B | Household Member Roster | Individual | HH_SEC_B | y2_hhid indidy2 | roster of individuals living in the household, relationship to the household, gender, year of birth, age, variable to link individuals between survey rounds, marital status, spouse identification, parental status, and place of birth. |
| C | Education[4] | Individual | HH_SEC_C | y2_hhid indidy2 | educational attainment, school characteristics, and expenditures |
| D | Health | Individual | HH_SEC_D | y2_hhid indidy2 | general health status and on utilization of health services |
| E | Labour[5] | Individual | HH_SEC_E1 | y2_hhid indidy2 | labor market participation during the last seven days, wage work, non-farm enterprise activity, and domestic activities within the home |
| | | | HH_SEC_E2 | y2_hhid indidy2 | secondary non-farm enterprise activity |
| F | Food Outside the Household | Individual | HH_SEC_F | y2_hhid indidy2 | TZ shilling value of food consumed outside the home during the last seven days |
| G | Subjective Welfare | Individual | HH_SEC_G | y2_hhid indidy2 | subjective welfare assessment of standard of living |
| H | Governance | Single Individual per Household | HH_SEC_H1 | y2_hhid itemcode | knowledge of and interaction with public officials |
| | | | HH_SEC_H2 | y2_hhid itemcode | participation in village meetings |
| | | | HH_SEC_H3 | y2_hhid itemcode | media usage |
| I | Food Security | Household | HH_SEC_II | y2_hhid | information on the household's diet and food intake |
| | | | HH_SEC_I2 | y2_hhid | months during with food insecurity was experienced |
| J | Housing, Water and Sanitation | Household | HH_SEC_J1 | y2_hhid | dwelling characteristics, water source, sanitation facilities, and use of fuel and electricity |
| | | | HH_SEC_J2 | y2_hhid itemcode | details on water source usage |

[4] For a more detailed description about the Tanzanian Educational System, please reference Appendix E.
[5] For a detailed explanation of the TASCO and ISIC codes and their use within the data, please reference Appendix F and G.

| Section | Name | Level of observation | Data files | Unique Identifier | Description |
|---|---|---|---|---|---|
| K | Food Consumption | Household | HH_SEC_K1 | y2_hhid itemcode | quantity and value of food consumed within the household during the last seven days by source. |
| | | | HH_SEC_K2 | y2_hhid itemcode | frequency of consumption of food items |
| | | | HH_SEC_K3 | y2_hhid | shared meals with non-household members |
| L | Non-Food Expenditures (past one week and one month) | Household | HH_SEC_L | y2_hhid itemcode | non-food expenditure during the last week or last month |
| M | Non-Food Expenditures (past twelve months) | Household | HH_SEC_M | y2_hhid itemcode | non-food expenditure during the past 12 months |
| N | Household Assets | Household | HH_SEC_N | y2_hhid itemcode | assets |
| O | Assistance and Groups | Household | HH_SEC_O1 | y2_hhid itemcode | government and non-governmental assistance received |
| | | | HH_SEC_O2 | y2_hhid indidy2 | credit / savings groups |
| P | Credit | Household | HH_SEC_P | y2_hhid loancode | loans or credit received by the household during the last twelve months, conditions of the transaction, and intended use |
| Q | Finance | Household | HH_SEC_Q | y2_hhid | details on the households financial assets and financial availability |
| R | Recent Shocks to Household Welfare | Household | HH_SEC_R | y2_hhid shockid | economics shocks experienced by the household during the last five years |
| S | Deaths in the Household | Individual | HH_SEC_S | y2_hhid hh_s02_1 | deaths of household members in the last two years |
| U | Anthropometry | Individual | HH_SEC_U | y2_hhid indidy2 | anthropometric information on household members aged over seven months |
| V-2 | Filter Questions for Agricultural & Fishery Questionnaire | Household | HH_SEC_V2 | y2_hhid | identification questions on households who farm or have livestock or fisheries |
| FILTERS | Screening questions | Household | HH_SEC_FILTERS | y2_hhid | module screening questions for modules H, K, O, P, and S |
| | | | | | |
| GIS | Household Geovariables | Household | TZY2.HH.Geovariables | y2_hhid | compiled geospatial information for the household location (see table in appendix A) |
| CONS1 | Household Consumption Year 1 | Household | TZY1.HH.Consumption | hhid | household consumption aggregates from year 1 (see appendix C) |
| CONS2 | Household Consumption Year 2 | Household | TZY2.HH.Consumption | Y2_hhid | household consumption aggregates from year 2 (see appendix C) |

## Table 2: Agriculture Questionnaire

| Section | Name | Level of observation | Data files | Unique Identifier | Description |
|---|---|---|---|---|---|
| 1 | Household Roster | Individual | AG_SEC01 | y2_hhid indidy2 | age, and sex of the household members, copied from the household questionnaire |
| 2 | Plot Roster | Plot | AG_SEC2A | y2_hhid plotnum | list of all plots cultivated or owned by the household during the last completed long and short rainy seasons |
| | | | AG_SEC2B | y2_hhid plotnum | |
| 3 | Plot Details | Plot | AG_SEC3A | y2_hhid plotnum | detailed plot information (agricultural practices, ownership status of the land, use of fertilizers or pesticides, inputs received on credit, and labor inputs) for long and short rainy season |
| | | | AG_SEC3B | y2_hhid plotnum | |
| 4 | Annual Crops by Plot | Plot-crop | AG_SEC4A | y2_hhid plotnum zaocode | crops planted and harvested, seeds used, and any unexpected losses experienced for long and rainy season |
| | | | AG_SEC4B | y2_hhid plotnum zaocode | |
| 5 | Annual Crop Production and Sales | Crop | AG_SEC5A | y2_hhid zaocode | quantity and value of crops sold, post-harvest losses, and storage |
| | | | AG_SEC5B | y2_hhid zaocode | |
| 6 | Permanent Crops by Plot | Plot-crop | AG_SEC6A | y2_hhid plotnum zaocode | age of plants, agricultural practices, quantity harvested, and unexpected losses |
| | | | AG_SEC6B | y2_hhid plotnum zaocode | |
| 7 | Permanent Crops – Production and Sales | Crop | AG_SEC7A | y2_hhid zaocode | quantity and value of crop sold, post-production losses, and storage |
| | | | AG_SEC7B | y2_hhid zaocode | |
| 8 | Outgrower Schemes and Contract Farming | Crop | AG_SEC8A | y2_hhid zaocode | outgrower or contract farming agreements for annual crops in long rainy and short rainy seasons, and permanent crops |
| | | | AG_SEC8B | y2_hhid zaocode | |
| | | | AG_SEC8C | y2_hhid zaocode | |
| 9 | Processed Agricultural Products and Agricultural Bi-Products | Crop-product | AG_SEC09 | y2_hhid zaocode ag09_02_3 ag09_03 | agricultural products that were processed during the last twelve months, including the cost and value at sale |
| 10A | Livestock | Animal | AG_SEC10A | y2_hhid lvstkcode | livestock owned by the household during the last twelve months, caretaking practices, and unexpected losses |
| 10B | Livestock By-Products | Animal by-product | AG_SEC10B | y2_hhid itemcode | quantity and value of livestock byproducts produced by the household during the last twelve months |

| Section | Name | Level of observation | Data files | Unique Identifier | Description |
|---|---|---|---|---|---|
| 10B | Livestock By-Products | Agricultural Service | AG_SEC10C | y2_hhid itemcode | sale of agricultural services |
| 11 | Farm Implements and Machinery | Implement | AG_SEC11 | y2_hhid itemcode | farm equipment owned or used by the household during the last twelve months |
| 12 | Extension | Extension source | AG_SEC12A | y2_hhid sourceid | interaction with governmental and non-governmental extension agents |
| | | | AG_SEC12B | y2_hhid ag12b_02 | |
| NETWORK | Network Roster | Entity | AG_NETWORK | y2_hhid ag_nid | network roster |
| FILTERS | Screening questions | Household | AG_FILTERS | y2_hhid | Screening questions for sections 2, 8, 9, 10 and 12 |
| | | | | | |
| GIS | Plot Geovariables | Plot | Plot.Geovariables | y2_hhid plotnum | distance of agricultural plot to the household dwelling |

## Table 3: Fishery Questionnaire

| Section | Name | Level of Observation | Data files | Unique Identifier | Description |
|---|---|---|---|---|---|
| A | Survey Information | Survey information | | | Information concerning the Fishery Questionnaire |
| B | Fisheries Calendar | Household | FS_B1 | y2_hhid | Indicates which months of the year fishing activities take place, with a distinction between high and low season |
| **High Season** | | | | | |
| C | Household Labour | Household | FS_C1 | y2_hhid<br>fs_c00 | Lists every member of the household who was engaged in fishing activities during the high season |
| D | Fishing Labour | Household | FS_D1 | y2_hhid | Details on hired labour related to fishing activities, including number of workers and their wages |
| | | | FS_D2 | y2_hhid | |
| | | | FS_D3 | y2_hhid | |
| E | Input | Household | FS_E1 | y2_hhid<br>gearid | Details on the kind of equipment used by households engaged in fishing activities, including fishing gear, boats/engines, and other inputs |
| | | | FS_E2 | y2_hhid<br>boatengine_id | |
| | | | FS_E3 | y2_hhid<br>inputid | |
| F | Output | Household | FS_F | y2_hhid<br>fishid | Details on the kind of fish caught and what is done with the catch by the household |
| G | Gear Rented Out | Household | FS_G | y2_hhid<br>gearid | Details on fishing gear rented out during the high season |
| H | Trading | Household | FS_H1 | y2_hhid<br>fishcode | Details of fish trading practices engaged in by households during the high season |
| | | | FS_H2 | y2_hhid<br>costid | |
| | | | FS_H3 | y2_hhid | |

| Low Season | | | | | |
|---|---|---|---|---|---|
| **Section** | **Name** | **Level of Observation** | **Data files** | **Unique Identifier** | **Description** |
| I | Household Labour | Household | FS_I | y2_hhid fs_i00 | Same as Section C but during the low season |
| J | Fishing Labour | Household | FS_J1 | y2_hhid | Same as Section D but during the low season |
| | | | FS_J2 | y2_hhid | |
| | | | FS_J3 | y2_hhid | |
| K | Input | Household | FS_K1 | y2_hhid gearid | Same as Section E but during the low season |
| | | | FS_K2 | y2_hhid boat_engineid | |
| | | | FS_K3 | y2_hhid inputid | |
| L | Output | Household | FS_L | y2_hhid fishid | Same as Section F but during the low season |
| M | Gear Rented Out | Household | FS_M | y2_hhid gearid | Same as Section G but during the low season |
| N | Trading | Household | FS_N1 | y2_hhid fishcode | Same as Section H but during the low season |
| | | | FS_N2 | y2_hhid costid | |
| FILTER | Screening questions | Household | FS_FILTERS | y2_hhid | Screening quetsions for sections F, H, L, and N |

## Table 4: Community Questionnaire

| Section | Name | Level of observation of the data file | Data files | Unique Identifier | Description |
|---|---|---|---|---|---|
| A-1 | Community Identification | Community | Not provided | | Community identifier variables and GPS coordinates |
| A-2 | Survey Staff Details | Community | COMSEC_CA | id_01 id_02 id_03 id_04 | Directly observed community characteristics |
| CB | Access to Basic Services | Service | COMSEC_CB | id_01 id_02 id_03 id_04 | Availability of basic services, the name of the nearest provider, and the distance to their location. |
| CC | Investment Projects | Investment Project | COMSEC_CC | id_01 id_02 id_03 id_04 | Presence of local investment projects for basic services in 2007/8, including sources and amount of funding |
| CD | Land Use | Community | COMSEC_CD | id_01 id_02 id_03 id_04 | Land-use practices in the village and any notable changes in land ownership/appropriation |
| CE | Agriculture | Community | COMSEC_CE | id_01 id_02 id_03 id_04 | Presence of and participation rate of cooperatives in the community, the availability of improved maize inputs, and an assessment of the quantity of local rainfall in the previous rainy seasons |
| CF | Demography and Family Issues | Community | COMSEC_CF | id_01 id_02 id_03 id_04 | Origins of the village, predominant religions and ethnic groups, and marriage and inheritance practices |
| CG | Governance | Community | COMSEC_CG | id_01 id_02 id_03 id_04 | Village assemblies and tribunals during the last twelve months, and their basic characteristics |
| CH | Water and Sanitation | Community | COMSEC_CH | id_01 id_02 id_03 id_04 | Details about the communities access and use of water and sanitation services |
| CI | Roster of Community Leaders | Community Leader | COMSEC_CI | id_01 id_02 id_03 id_04 cm_i01 | Community leaders including their occupation, political party, and other socio-demographic characteristics |
| CJ | Market Prices | Item | COMSEC_CJ | id_01 id_02 id_03 id_04 itemid | Prices for basic foodstuffs and nonfood items |
| | | | | | |
| GIS | EA Offset | EA | TZY2.ED.Offsets | clusterid | Offset GPS coordinates for Enumeration Area Center |

## Sample Design

The sample design for the second round of the NPS revisits all the households interviewed in the first round of the panel, as well as tracking adult split-off household members. The original sample size of 3,265 households was designed to representative at the national, urban/rural, and major agro-ecological zones. The total sample size was 3,265 households in 409 Enumeration Areas (2,063 households in rural areas and 1,202 urban areas). It is also be possible in the final analysis to produce disaggregated poverty rates for 4 different strata: Dar es Salaam, other urban areas on mainland Tanzania, rural mainland Tanzania, and Zanzibar.[6]

Since the TZNPS is a panel survey, the second round of the fieldwork revisits all households originally interviewed during round one. If a household has moved from its original location, the members were interviewed in their new location. If that location was within one hour of the original location, the field team did the interview at the time of their visit to the enumeration area. If the household had located more than an hour from the original location, details of the new location were recorded on specialized forms, and the information passed to a dedicated tracking team for follow-up.

If a member of the original household had split from their original location to form or join a new household, information was recorded on the current whereabouts of this member. All adult former household members (those over the age of 15) were tracked to their new location. Similar to the protocol for the re-located households, if the new household is within one hour of the original location, the new household was interviewed by the main field team at the time of the visit to the enumeration area. For those that have moved more than one hour away, their information was passed to the dedicated tracking team for follow-up. Once the tracking targets have been found, teams are required to interview them and any new members of the household.

The total sample size for the second round of the NPS has a total sample size of 3924 households. This represents 3168 round-one households, a re-interview rate of over 97 percent. In addition, of the 10,420 eligible adults (over age 15 in 2010), 9,338 were re-interviewed, a re-interview rate of approximately 90 percent.

## Implementation

The field staff was trained in Morogoro in August 2010 over four weeks. Enumerator and data entry training was done concurrently. During a standard training week, four days were spent in classroom, and one day in field training. On each Saturday of the training month, the field staff was debriefed on the previous day's field exercise and what they had learned over the previous week. Over the four week training period, the field staff spent one week on the Household questionnaire, Agricultural questionnaire, Fishery questionnaire and tracking, and field practice respectively. Over the training period, the field teams were administered two tests. The goal was to gain feedback from the training sessions and to select the enumerators. Overall, there

[6] For a more detailed description of the sample design, please reference the NPS 2008/2009 BID.

were 52 enumerator candidates, with 48 being selected. At the end of the training the enumerators were proved Swahili field manuals.

Prior to the training period, two pilots were conducted. In early July 2010, the new questionnaire instruments were piloted in Morogoro. In mid-July 2010, the tracking pilot was conducted in Tanga and Handeni. Households from year one were revisited to provide the field staff practice in conducting tracking fieldwork. After the pilots, extensive discussion and revisions were conducted with the participation of all team supervisors.

The main data collection began in October 2010 and finished in September 2011, with tracking fieldwork continuing until November 2011. The survey was implemented by eight mobile field teams, each composed of: one supervisor, four enumerators, one data entry technician, and one driver.

The teams visited each enumeration area for between 4-5 days. The questionnaires were administered to the selected households over the course of that time. This allowed the field team to make return visits to the household to complete the entire Household questionnaire and, for farm households, Agriculture questionnaires, and for Fishery questionnaires. To ensure the depth and quality of each section of the survey, the questionnaire was administered across multiple respondents to the most knowledgeable about each topic. For all of the sampled households, areas of all owned and/or cultivated agricultural plots were measured via GPS unless the household refused, the terrain was too difficult, or if the plot was more than 1 hour from the location of the household. Anthropometric measurements were taken for all individuals that were at home, not too ill, and willing to participate.

If the field teams enter an enumeration area and find that the entire household or a member(s) of the household has moved, they are required to follow the tracking protocol. If the entire household has moved from the original residence, teams are required to fill a T-1 form. The T-1 form contains information on the new location of the household, allowing for the teams to locate and interview the household members. If a member or members of the household have split from the original household, a T-2 form is filled out by the teams. Similar to the T-1, a T-2 form contains information on the location of the member(s) who have split from the household. Once the tracking targets have been found, teams are required to interview them and any new additions to the household. Out of the tracking individuals/households, only those over 15 years of age are included in the tracking protocol unless an individual under 15 years of age moved with another individual over 15 years of age, and both were part of the round one data collection.

Within the tracking protocol, there are local and distance cases. Local and distance tracking applies to both T-1 and T-2 forms. Local tracking occurred when the tracking target is within one hour traveling distance from the original EA and at least one tracking member from the household is over 15 years of age. If that is the case, the teams are required to interview the tracking target before leaving the original EA. Distance tracking occurs when the tracking target is not within one hour traveling distance from the original EA. In this case, the teams fill out the appropriate tracking form and send the information NBS headquarters. Once at NBS, the distance tracking case is given to the tracking team, who is then responsible for locating that household and conducting the interview.

The mobile tracking team consisted of one supervisor, two interviewers, one data entry technician, and one driver. In addition, there were two dedicated tracking enumerators that remained in Dar es Salaam. The tracking team began interviews three months after the beginning of fieldwork to allow enough time to accumulate a sufficient number of tracking targets. Tracking targets were grouped into geographic regions, and the team would visit the regions approximately every 2-3 months. Any tracking target not located was remained in the pool to be visited during the next trip, in addition to any new tracking cases that had accumulated in the intervening months. In addition, the regular field teams also sporadically would perform tracking within their interview regions if there was a backlog of cases. Finally, following the completion of the main fieldwork activities, four supervisors led dedicated tracking teams to interview the remaining cases.

Data entry was done concurrently with data collection by the data entry technician, using a laptop, known as first data entry. The data entry program was a CSPro-based system, developed by NBS with support from the World Bank. This facilitated the performance of internal crosschecks prior to departure from the enumeration area, allowing enumerators to return to households and clarify inconsistent information on the questionnaires. Data files from completed EAs were then e-mailed to headquarters using 3G modems. These files were concatenated and periodic checks were done to ensure the fieldwork was proceeding according to the calendar. The field teams also send the paper questionnaires back to the headquarters on a monthly basis.

Once the paper questionnaires and data files for completed EAs were received at NBS headquarters, a double entry procedure was implemented. Eight data entrants were hired by NBS to re-enter the data from the paper questionnaires into the CSPro-based data entry system for all households and questionnaires administered. A cross comparison between the entered values in the field based data entry and double entry was conducted and any differences in values between the two were flagged for manual inspection of the physical questionnaire. Corrections based on this inspection exercise were ultimately encoded in the dataset.

Additionally, an extensive review of data files was conducted, including interviewer errors such as missing values, ranges and outliers. Observations were returned for manual inspection of the physical questionnaires if continuous values fell outside five standard deviations of the mean, categorical values were not eligible responses, or there were internal inconsistencies within the dataset (for example, the age of an individual was not consistent with their educational status, there was more than one head of household listed, an individual was engaged in multiple primary activities, the quantity of crops and their byproducts produced, harvested, and sold not listed, the distance from the market and an individual's plot was not listed, the number of weeks, days per week, and hours per day an individual engaged in fishery activity was not recorded, the species and quantity of fish caught, bought, sold, or traded was not listed, etc).

When it was determined that these values were the result of data-entry error, the values were corrected. In addition, cases deemed to reflect obvious enumerator error were also corrected in this cleaning process. The majority of such cases involved the use of incorrect measurement units, e.g. recording grams as kilograms or vice versa.

## Data Set

The TZNPS consists of several data files. Each data file pertains to a section of the questionnaire or a set of sections that are for the same level of observation. The complete lists of data files as well as the unique identification variables are listed in Tables 1-4; there are between one and four unique identification variables in each data set.

Households are identified by a sixteen-digit number. The first fourteen of these digits are constructed from the region, district, ward, locality, enumeration area, and household numeric identifiers from the first round of the panel. These should not be considered to be the location identifiers for the second round data. The last two digits is the identification number for the tracking target with the lowest individual id from round one. For example, let us assume that a household in round one consisted of 6 members with individual id numbers from 1-6. If in round two, members 1, 3 and 6 remained in the original household while members 2, 4, and 5 split into a different household together. Then, in round two, the first household receives the 14 digit number as explained above and the minimum of the individual ids from round one of the individuals that are living together. In this case it will be 01. The second household receives the 14 digit number as described above, and 02, since it is the smallest value of the three individual IDs from round one that split together.

*Merging between Round 2 data sets*
The household identification variable is "y2_hhid" in the data files. When merging or linking individual data files, it is necessary to use at minimum the household identification variable ("y2_hhid") as well as the other unique identification variables pertinent to the data files being merged.

The required variables to merge data sections depend on which two sections are being merged. For example, merging sections A and J (basic identification and household characteristics) require only the "y2_hhid" variable as it is the single unique identifier in both datasets. Merging sections B and C (roster information with education levels) requires merging on both the y2_hhid and the individual's id ("indidy2") from round two[7]. This combination of variables will be unique in the roster datafile, as only one person can have a particular roster number. Similar patterns will be found in other combinations of data files, such as the plot number "plotnum" and crop id "zaocode" variables in the agricultural datasets.

*Merging round 1 and round 2*
Additionally, it is possible to link the data from round two with those from round one. To merge household level datasets, it is necessary to use the "hhid_2008" id variable (found in Section A of the household data) and the "hhid" variable in the year 1 data. Similar principles apply when merging individuals between years. Again the "hhid_2008" / "hhid" provide the link at the household level. To match the individuals, it is necessary to use the combination of the "hhid" and "sbmemno" variables from year 1 and the "hhid_2008" and "hh_b06" variables from year 2. The "hh_b06" variable contains the individual id number (the "sbmemno" variable) from the

[7] It is important to note that while the y2_hhid variable is constructed using the year 1 individual ids, a separate individual identification number for each individual (old and new) is provided in round two.

first year. This combination should only be used to match between years. For all merges within year 2, the "indidy2" variable should be used.

## Weighting / Expansion Factors

In order to produce nationally representative statistics with the NPS data, it is necessary to apply weighting or expansion factors. The panel survey weights adjust for differences in the probability of selection into the NPS round 1 sample for observations in various strata, 2008/2009 households splitting into multiple households in 2010/2011, and attrition between rounds of the survey.[8]

The first round of the NPS sample was a multi-stage clustered sample design. First stage sampling involved the selection of survey clusters with the probability of selection proportional to cluster size within a stratum. The sampling of these clusters was stratified along two dimensions: (i) eight administrative zones (seven on Mainland Tanzania plus Zanzibar as an eighth zone), and (ii) rural versus urban clusters within each administrative zone. The combination of these two dimensions yields 16 strata. In rural areas a cluster is defined as an entire village. In urban areas a cluster is defined as a census enumeration area. As a general rule, the probability of selection was higher for clusters within strata where existing data sources showed that the variance of key variables of interest for the NPS (e.g., household consumption and maize production) were likely to be very high – implying the need for more observations to produce reliable estimates.

The expansion factors can be found in section A of the household dataset Section A also contains unique identifiers for the first-stage sampling units ("clusterid") and for the sampling strata ("strataid").

## Obtaining Data

Data is available on the Tanzania National Bureau of Statistics website (www.nbs.go.tz), and may be downloaded free of charge. Inquiries sent to the LSMS team at lsms@worldbank.org will be forwarded to NBS for processing.

[8] The details of the sample design – including the sampling strata and the use of multiple sampling frames – are discussed in a separate document, *Sample Design for the National Panel Survey*, April 2009, available from NBS upon request.

# Section Notes

## *HOUSEHOLD QUESTIONNAIRE*

SECTION A

Section A is the cover page that comprises of two parts. The first part includes basic information for household identification such as - the region, district, ward and village within which the household is located. The second part provides information on the staff that conducted the survey such as name of the enumerator, supervisor, data entry clerk, date and time of interview. All identifying variables have been removed from the dissemination version of this dataset to preserve the confidentiality of the respondent. This includes the name of the of the village, the name of the household head, the name of the tracking target, the description of the household location, and the names of the interviewer, supervisor, and data entry operators. The file also contains the weight and stratification variables.

SECTION B

Section B is the household information roster that comprises of basic information on each member of the household such as name, gender, month and year of birth, relationship to head of the household, occupation, marital status etc.

The unique identifiers in this section are the round two household id and indidy2, which is the individual id for each member of the household in round two. Note that indidy2 is different from hh_b06 since the latter is the individual id from round one and is coded as 99 if the individual is a new member in round two and did not exist in round one.

A link between round one and round two data at the individual level is facilitated by using a combination of the household id from round one and individual id from round one. In order to merge the two rounds of data, the "hhid" variable from Section B of round one should be renamed to "hhid_2008" while the "sbmemno" variable from Section B round one should be renamed to "hh_b06." After the renaming of these variables in round 1, a merge can be created using "hhid_2008" and "hh_b06" together as unique identifiers. It is important to note that in the round 2 data, "hh_b06" also includes 99s, which are new members that did not exist previously in round 1. Links will not be created for these cases between the two rounds of data.

It is possible to extend the link between round 1 and round 2 of section B to other sections at the individual level between the two years – C, D, E, F and U. This is done by merging Section B with Section C in round 2 using unique identifiers "y2_hhid" and "indidy2."

In this section, three variables were removed due to confidentiality – the respondent's name, the name of district of previous residence, and the name of the district of birth.

SECTION C

Section C is the module on education and covers questions related to educational attainment, type of school attended, amenities provided at school, performance on national level exams, education related expenditure etc.

Appendix F lists some basic information on the Tanzanian education system. Descriptions for some additional terminologies used in this section are provided below.

PSLE refers to the Primary School leaving exam, organized by the National Examinations Council of Tanzania. It is held during the last week of September every year for all primary school students in D7. A passing grade is required to move to secondary school.

Form 4 and Form 6 are secondary school level exams, organized by the National Examinations Council of Tanzania. A passing grade in Form 4 is required to move to Form 5 while a passing grade in Form 6 is required to seek admission into higher education.

A merge between Sections B and C in Round 2 is possible by using a combination if "y2_hhid" and "indidy2". No variables have been excluded from this section.

SECTION D

This module asks questions on the health of the household member with particular portions focused at women and children. Questions on health provider, expenditures on health services, type and incidence of illness or injury and use of bednets are covered in this section. Additional questions on disability measures have been incorporated in this module using inputs from the Washington Group on Disability Statistics within the United Nations Statistics Division[9].

A merge between Sections B and D in Round 2 is possible by using a combination if "y2_hhid" and "indidy2".

SECTION E:

Module E is divided into two data files. Section E1 includes the primary activity for all individuals aged 5 years and above, covers questions related to employment activities for wage, apprenticeships, self- employment and other general categories. The included topics cover the type of occupation, hours worked, income earned, profits, business operating costs etc. Note that the questions are at the individual level for questions 1 to 53 and 74 to 81, and at the level of the non-farm enterprise for questions 53 to 73. In cases where more than one household member operated a non-farm enterprise, the information is entered only once (see the interviewer instruction in the margin of questionnaire page 20). Question 56 uses alphabetic codes to indicate which household members worked in which non-farm enterprise.

[9] See http://unstats.un.org/unsd/methods/citygroup/washington.htm for further information.

In the event that the same household member operated more than one non-farm enterprise, the information is included in the rows below the table – starting on page 20. To merge the primary activity (section E1) with the roster (or other individual level data file), it is necessary to use "y2_hhid" and the "indidy2" variables. To have a complete list of the activities, it is necessary to append the section E1 and section E2 data, but the resulting file will not be a one-to-one merge with other individual sections due to the presence of secondary activities.

Appendix G includes codes for different occupations based on TASCO. These are specifically for Questions 16, 31 and 45. Additionally, Appendix H includes ISIC codes for questions 17, 32, 46, and 53. These text fields are removed for confidentiality purposed, as well as the text names of the employing companies.

SECTION F:

Module F includes questions on meals, snacks, or drinks consumed outside the household in the past seven days by each household member and associated values for their consumption.

A merge between Sections B and F in Round 2 is possible by using a combination of the "y2_hhid" and "indidy2" variables.

SECTION G:

This module on subjective welfare asks questions about levels of satisfaction with various components such as health, finances, housing, job, health care, education and safety. It also has questions on an individual's perception of welfare in comparison to 3 years and 10 years ago. Respondents are limited to those 15 years and older, and proxy respondents are excluded.

A merge between Sections B and G in round 2 is possible by using a combination of the "y2_hhid" and "indidy2" variables.

SECTION H:

This module is designed to measure a respondent's interaction with government and local leaders. In the first round of the survey, a household member was randomly selected as the respondent. The interviewers were instructed to ask the governance questions to the same respondent in the next round of the survey. If the original household member was unavailable, if for example they were temporarily away from the household, or if the selected household member was not longer part of the household, a replacement respondent was selected using a Kish table. Question H2 indicates that 87 percent of round 1 respondents were re-interviewed in round 2.

For the purposes of confidentiality, the names of the officials are removed from the public access dataset. Should the analysis want to pursue research as to whether the respondent was able to

correctly identify the individual officials, it will be necessary to file a special request with the NBS.

This file is an individual level file that contains only one observation per household. Therefore to merge this information with the individual characteristics, it would be necessary to create a merge variable using the information in section H questions 1 and 3, and merge this created variable to the "indidy2" variable (using also the "y2_hhid") in other individual level datasets. Note that this will not be a one-to-one merge.

SECTION I

This module deals with questions related to food security, and asks the head of the household questions on food consumption patterns in the past seven days with a particular focus on preferences, portions, variety and number of meals consumed. The majority of the data in section I is contained in the data file I1. Question 9 is recorded in a separate data file, I2. During the field work there was a lack of understanding during the initial stages regarding the recall period for question 9. Therefore during the data cleaning process all out-of-range values were removed.

It was also noticed during the data cleaning process that there were a large number of "other" responses in urban areas. Using the text fields, it was determined that the main reason was "no money." Therefore a code was added for "no money," and the questionnaires and data labels were updated. The "no money" field contains all text responses related to "no money," "business was bad," "no / low income," etc. The full text fields (in Swahili) are available upon request.

SECTION J

This module deals with questions on housing, water and sanitation. For housing, questions on tenure status, rental value, housing characteristics such as number of rooms, roofing, wall, and floor material are included. For sanitation, questions on household garbage disposal, toilet facilities, sewage connection are included.

For water, questions on sources of water for drinking, cooking, washing and gardening are included along with time taken to get water from the source. Additionally questions on type of container used to collect water and price of water are asked for each source from which water is collected. For areas with connection to a local water utility, questions on quality of water, hours of supply, and willingness to pay are included. The major source for the questions on water was the Millennium Challenge Account – Tanzania that intends to use these questions for an impact evaluation of their water sector projects in the country.

The section is divided into two data files. The first, J1, contains the majority of the data from the section. This module is administered at the household level, and the identifying variable is "y2_hhid." In addition, section J2 has the water source roster information. Data from this section requires both "y2_hhid" and "itemcode" to uniquely identify observations.

SECTION K

This module focuses on food consumption over the past week within the household and is divided into four parts:

The first part, in data file K1, contains a list of 46 food items under the broad category of cereals, sugar, pulses, nuts and seeds, vegetables, fruits, meat and fish, milk, oil and fats, spices, and beverages. Questions on consumption amounts, source of consumption (purchase, production or other sources) are asked for each food item while the number of days of consumption is asked at the general category level. Following the completion of fieldwork, data from this section was cleaned to correct misalignment of rows, incorrect units, missing decimal points, and other inconsistencies. The main tool for these corrections was the second data entry, though other logic checks were also employed.

An iodine test is also performed to detect its presence in salt and corresponding parts per million is recorded. This information is included in the "Filters" data file.

The final two sections address food security issues within the household. Section K2 contains information on the number of days that particular food categories were consumed, and section K3 asks questions regarding meal sharing. The screening question for section K3 is included in the "Filters" data file.

The unique identifiers for sections K1, K2, and K3 are "y2_hhid" and "itemcode," though it should be noted that "itemcode" represents different things in different sections, and cannot be used to merge between the sections K.

SECTION L

Module L is consists of expenditure on non-food items in the past week or the past month. A total of 3 items fall under the 1 week recall section, and 24 items are listed in the one month recall section.

Similarly to section K, the unique identifier variables in this section are "y2_hhid" and "itemcode," but these variables cannot be used to merge between consumption sections.

SECTION M

This module is similar to section L but has a 12 month recall period for the 19 items. The final two items in the section "wood poles, bamboo" and "grass for thatching roof or other use" are commonly consumed by households but not always purchased. These questions therefore also include the estimated total value in addition to the purchased amount.

Similarly to sections K and L, the unique identifier variables in this section are "y2_hhid" and "itemcode," but these variables cannot be used to merge between consumption sections.

SECTION N

This module on household assets asks questions on the number of items owned by the household for 52 categories. As with the previous three sections, the unique identifier variables in this section are "y2_hhid" and "itemcode" but these variables cannot be used to merge between consumption sections.

SECTION O

This module is divided into two categories, with the first focusing on assistance from government/ non – government institutions (excluding SACCOS) on free food/maize distribution, food for work or cash for work programmes, inputs for work programme, scholarships or bursaries for primary and secondary school and associated amounts received in cash and kind. The name of the organization is excluded due to confidentiality purposes. The unique identifier variables in this section are "y2_hhid" and "itemcode" but these variables cannot be used to merge between sections. Individual characteristics of participating members can be linked using the roster id numbers in question 6, which correspond to the "indidy2" variables in the individual data sets.

The second part focuses primarily on SACCOs, a leading player in cooperative microfinance activities in the country, and asks questions on balance within the group, contribution frequency and amounts and withdrawal amounts. In this section the individual linking variable is included directly ("indidy2"), and the "y2_hhid" and "indidy2" variables can be used to merge between sections.

SECTION P

This module is only administered to those households that borrow from any persons or institutions outside the household and receive cash, goods or services. The screening question for this module is included in the "Filters" section. Questions on the household member seeking loan, amount, repayment amount including interest and reason for loans are included in this module. The name of the lending institution is removed due to confidentiality purposes.

The variables which uniquely identify observations in this section are "y2_hhid" and "loancode." In order to merge loans with individual characteristics, question hh_p04 must be used as the "indidy2" variable.

SECTION Q

Within the Finance module, questions on money transferring services, such as M-PESA, Z-PESA and ZAP, are included along with the reason for using them. The module also asks questions on remittances or financial assistance in the form of cash or in-kind from abroad in the

last 12 months. Finally, questions on banking such as bank account ownership and associated institutions are included.

For the purposes of confidentiality, the names of the institutions at which households maintained accounts (question 19) have been replaced with either "other," or "other SACCOS" if they were clearly identifiable as a SACCOS, if there were less than five total observations. The unaltered dataset can be obtained with special permission from the NBS for researchers that require this information.

SECTION R

This module deals with recent shocks to household welfare. The module first asks if the household has been affected by any of the 18 shocks in the list, and an "other shock" category. Of those shocks the household reported experiencing, the top three in terms of impact are ranked, and follow up questions are asked about those shocks. Note that despite additional training, numerous problems were still observed with the administration of this module by some interviewers, particularly with the ranking of the shocks.

SECTION S

This module records information on deaths within the household, reason for death, duration of illness before death, and associated land or asset losses due to inheritance traditions. The member id is also recorded if the person was present in the round 1. This variable can be linked using question 4 in the same way that question 6 in Section B is used to link living household members. Note that not all deceased household members will have been present in the first round as some may have joined the household in the interim between surveys.

The name of the deceased household member is excluded for confidentiality purposes.

SECTION U

Module U collects anthropometric information for all household members unless there were refusals, illness or other reasons. In the second round of the NPS, 83.81 percent of household members were present for measurement.

SECTION V

Module V1 records GPS coordinates of the household, phone numbers for household members and two reference persons within and outside the community. This information is not publicly available due to confidentiality concerns. A separate GIS dataset has been compiled using household location information. If a researcher needs the specific GIS coordinates beyond what is available in the public datasets, a written research proposal must be submitted to NBS in

application for these data. In the case of university and graduate students, a letter from the department chair in support of the research agenda should accompany the application.

Module V2 contains the screening questions for the agriculture and fisheries modules.

## *AGRICULTURAL QUESTIONNAIRE*

All households that answered Question 206 in Section V2 as yes should appear in this module. This means they could have cultivated plots, owned farms which they did not cultivate, or owned livestock in the last 12 months. Note that the cover page is not included in the dataset because they do not contain any additional information beyond what is included in the household questionnaire cover page.

### SECTION 1

This module is a household member roster and is identical to module B of the household questionnaire. An X mark is attached to those individuals that served as respondents to this questionnaire. The names of the individuals have been removed for confidentiality purposes.

### SECTION 2A

Section 2A is a plot roster of all plots cultivated in the long rainy season (masika) of 2010. Note that in the majority of cases the 2010 long rainy season was the most recent rainy season – though households interviewed towards the end of the fieldwork may have completed the 2011 season as well. The categories include plot name, description, farmer's estimate of area, plot number in 2009, area of plot based on GPS measurement, and weather condition during measurement. Question 5, the plot number in the 2008/2009 survey, can be used to merge plots between survey rounds as it corresponds to the "plotnum" variable in the first round.

The plot areas were measured using eTrex HC series GPS devices. The survey protocols indicate that all plots should be measured as long as they were with one hour's transportation (either on foot, by bicycle / motorbike, or, if possible, by vehicle) from the household. The only other acceptable reason for the plot not to be measure would be if the household refused. This information is recorded in question 8.

The plot names, descriptions, and GPS coordinates are excluded for confidentiality reasons. A separate GIS dataset has been compiled using plot location information. If a researcher needs the specific GIS coordinates beyond what is available in the public datasets, a written research proposal must be submitted to NBS in application for these data. In the case of university and graduate students, a letter from the department chair in support of the research agenda should accompany the application.

SECTION 2B

Section 2B records the same information as section 2A but for the sort rainy season (vuli). There is a screener question on the bottom of page 4 (and included in the Ag Filters dataset) that indicates to which short rainy season the household is referring. Note that the short rainy season occurs only in certain parts of the country.

As with the above, the GPS coordinates have been removed from the public dataset for confidentiality purposes, and the same procedure is in place to obtain their access. In this case, question 16, the plot number in the 2008/2009 survey, can be used to merge plots between survey rounds as it corresponds to the "plotnum" variable in the first round.

SECTION 3A

This module is administered at the plot level for all plots cultivated in the long rainy season of 2010, and includes questions on the usage of plot (rented, cultivated, fallow, given out etc.). Information on the main crop cultivated, decisions on which crop to cultivate, soil type and quality with a focus on erosion, source of irrigation within the plot, ownership status of the plot, rental value, usage patterns for organic and inorganic fertilizers, usage pattern for pesticides and herbicides and inputs on credit. The section also includes questions on household and hired labor on each activity of farming – land preparation and plating, weeding, ridging, fertilizing and harvesting.

SECTION 3B

This module is administered at the plot level for all plots owned or cultivated in the short rainy season of 2009 or 2010 and includes questions on the usage of plot (rented, cultivated, fallow, given out etc.). This would also include plots cultivated in the long rainy season but not in the short if they were owned during the short rains. Information collected is identical to the section above except for the last 10 questions, which relate to plot acquisition.

SECTION 4A

This module is administered for all crops by household and plot in the long rainy season of 2010. Thus the unique identifier is "y2_hhid," "plotnum," and "zaocode," and the plot names are dropped for confidentiality purposes. This section asks questions on crop planting patterns, intercropping, area and quantity harvested, associated losses, crop seeds purchased along with associated values, source and type of seed.

SECTION 4B

This module is administered for all crops by household and plot in the appropriate short rainy season. This includes all crops listed in section 2B and all crops for which with a "yes" to question 36 in section 3B. This section otherwise is identical to section 4A.

SECTION 5A

This module is administered for all crops at the household level grown in the long rainy season of 2010. Thus, the unique identifiers are "y2_hhid" and "zaocode." The reason for the differential disaggregation than Section 4 is the operational assumption of similar crops from different plots of a household sold together and not as differentiated goods. This section includes questions on quantity of crops sold, value of sales, customers crops sold to, average distance that crops were transported to for sale, post-harvest losses, how crop residue was handled, method and duration for which crop was stored. Note that this section should only include annual crops, such as maize, paddy, beans, etc, but due to interviewer errors, there are some permanent and fruit crops also captured.

In some cases more than one crop produced by the household was not represented in the list of response codes. In these cases multiple crops were classified as an "other." Because the specifics of the "other" responses in this section were captured additional codes for others, including 997, where added post collection in order to uniquely identify the other crop within the household.

SECTION 5B

This module is administered for all crops at the household level grown in the appropriate short rainy season. Similarly to section 5A, the identifying variables are "y2_hhid" and "zaocode."

SECTION 6A

This module focuses on fruit crops at the household and plot level for all plots. Questions related to the number of trees/plats planted on the crop, when most of these were planted, how many planted the past 12 months, presence of intercropping, quantity produced, losses before and after harvest, quantity sold, associated value and location sold, method and quantity of crop stored.
During the data cleaning, it was noticed that a large number of "other" codes were identified as soursop fruits. Therefore a new code was created and included in the data for soursop fruit. Also, in the English version on the questionnaire, both peaches and plums appear twice in the crop listing. This is due to the fact that there are multiple fruits common in Tanzania that have the same English translation, and the names are different in the Swahili version.

Similarly to section 4, this dataset is at the plot/crop level, and therefore three variables are necessary to uniquely identify observations, "y2_hhid," "plotnum," and "zaocode." Note that

this section should include only fruits but due to a small number of interviewer errors, some fruits are included in section 4 instead.

As in section 5A, cases with multiple "other" codes where identified and given unique other codes within the case.

SECTION 6B

This module focuses on permanent crops at the household and plot level for all plots. Questions related to the number of trees/plats planted on the crop, when most of these were planted, how many planted the past 12 months, presence of intercropping, quantity produced, losses before and after harvest, quantity sold, associated value and location sold, method and quantity of crop stored. Similarly to section 6A, this dataset is at the plot/crop level, and therefore three variables are necessary to uniquely identify observations, "y2_hhid," "plotnum," and "zaocode."

In this section, there are a numbers of crops within the plot that share the same crop category code (zaocode). This is due to the respondent's differentiation of specifically "Firewood/fodder", "Timber", "Medicinal Plants", "Fence Tree" and "Other" types on the plot. Do to the large number of these cases the dataset is preserved and the option to collapse on specific variables in left to the end user.

SECTION 7A

This module focuses on fruit crops at the household and plot level for all plots. Questions related to quantity of crop sold, associated value and location sold, post production losses and method and quantity of crop stored. Similarly to section 5, this data is recorded at the crop level so the unique identifiers are "y2_hhid" and "zaocode." Also, similarly to section 6A, an additional code was added for soursop fruit.

SECTION 7B

This module focuses on permanent crops at the household and plot level for all plots. Questions related to quantity of crop sold, associated value and location sold, post production losses and method and quantity of crop stored. Similarly to section 5, this data is recorded at the crop level so the unique identifiers are "y2_hhid" and "zaocode" with the addition of the crop name for multiple reported crop category of different variety. Also, similarly to section 6A, an additional code was added for soursop fruit.

Although 7B was intended to be a direct continuation of section 6B, listing all crops present in 6B, interviewers often failed to report crops that were used specifically for own consumption. Consequently, Section 7B has fewer observations than that reported in Section 6B.

SECTION 8

This module focuses on outgrower schemes and contract farming and is divided into three parts. The first part focuses on crops grown as part of the long rainy season 2010, the second on the last completed short rainy season and the third on fruit trees and permanent crops. Before any of the parts are administered, a filter question is set in place that eliminates any households that did not cultivate any crops, permanent crops or fruit trees as part of an outgrower scheme or contract farming system at any time over the past 12 months. For those farmers that did engage in such schemes, the crop name involved, the company from which scheme is obtain, agreement on items decided in advance before planting along with form of commitment, buyer compliance and related issues are all recorded as per the three categories. Though theoretically the same farmer could participate in numerous outgrower schemes per crop, in these datasets, no farmer participates in more than one – therefore the unique identifier variables are "y2_hhid" and "zaocode."

SECTION 9

Within this module, a filter question is set up such that only households that processed any of the products harvested on the farm in the last 12 months are administered the subsequent questions. Specifically, information on crop name, by product name and quantity produced, quantity sold, associated prices and buyers and costs incurred due to labor/other inputs. Because the household can have more than one by-product associated with the same crop, it is necessary to use four variables to uniquely identify observations: "y2_hhid" to identify the household, "zaocode" to identify the crop, "ag09_02_3" to identify whether the product is deliberately processed or produced as a by-product of another process, and "ag09_03" to identify the product. Note that the same product can be both a processed product and a by-product.

SECTION 10A

This section focuses on livestock and questions are only administered to households that own any animals in the last twelve months.

For 15 animal categories, identified by the "lvstkcode", questions related to ownership, purchases, gifts received, diseases and animals lost because of them, thefts, sales and associated earnings, slaughtering and associated earning, labour activities of household and hired members associated with upkeep of animals, fodder costs for the animal and vaccinations provided to the animal.

SECTION 10B

This module focuses on livestock by products and asks questions on whether the household produced any product in the last 12 months, quantity produced, quantity sold, value of sold goods, buyers of by product sold. The variables that uniquely identify observations in this section are "y2_hhid" and "itemcode."

SECTION 10C

This module asks if traction/draught power or sire services were sold by a household in the past 12 months and the associated value of sales and buyers. The variables that uniquely identify observations in this section are "y2_hhid" and "itemcode."

SECTION 11

This module focuses on the number farm implements and machinery used or owned by the household in the past 12 months along with associated value if sold, whether the item was used, reasons for no usage, whether any of these items were rented or borrowed for use in the last twelve months and associated rents paid. The variables that uniquely identify observations in this section are "y2_hhid" and "itemcode."

SECTION 12

This module focuses on any extension services or advice that the household received for agricultural or livestock activities in the past 12 months through, government extension, NGO, Cooperative/Farmer's Association, Large Scale Farmer and Others. The section then asks what activity was advice sought for, subjective rating for advice received, and price paid for receiving advice. In both datasets in this section, the variables to uniquely identify observations are "y2_hhid" and "sourceid," but these two variables are not equivalent and it is not possible to merge the two datasets using these variables.

FILTERS

This section contains the screening questions from modules 2, 8, 9, and 10.

NETWORK

Throughout the various sections of the agricultural questionnaire, there are questions that refer to persons outside the household that are involved in the agricultural process. Examples include landlords, suppliers of inputs, harvest purchasers, outgrower partners, etc. To link these persons outside the household across sections, each is assigned a code when they first appear, and keep the same code throughout the questionnaire. For example, if the household both buys inputs from and sells crops to the same person, this person would have the same network ID in section 3 for the input purchases and in section 5 for the crops sales. The network roster file contains the location and category of each of these persons. The names have been removed for confidentiality reasons. The unique identifier variables for this section of "y2_hhid" and "nid," and these variables can be used to merge the network roster characteristics into the appropriate agricultural datasets.

## *FISHERIES QUESTIONNAIRE*

All households that answered Question 208 in Section V2 as yes should appear in this module. This means the household was involved in fishing, fish farming, or fish trading in the last 12 months. Note that the cover page is not included in the dataset because they do not contain any additional information beyond what is included in the household questionnaire cover page.

### SECTION B

This module focuses on the fishing calendar for the household on a monthly basis categorized as H for HIGH FISHING, L for LOW FISHING and N for NO FISHING. The seasonal definitions are independent from the months in which the fisherman worked. For example, if the month of March is typically considered as part of the high season in the region, but the respondent did not fish during this March because of illness, the respondent should still indicate "H" for high season in the blank below the month of March.

In some cases, the respondent was unable to distinguish between high and low season. In these circumstances, the interviewer recorded H (High) for all months in which any fishing takes place, and then marked "no" in the enumerator's perception question following the table.

### SECTION C

This module focuses on household labor in the months listed in Section B as high season. For members, that are listed as being involved in fishing in the last high season, questions on time spent fishing, time spent processing and time spent trading are recorded by number of weeks, days per week and hours per day.

The variables that uniquely identify observations in this dataset are "y2_hhid" and "indidy2." These two variables can be used to merge data from the fishery section with the characteristics of the individual fisherman from the household section.

### SECTION D

This section is divided into three data files. The first focuses on shared labor and expenses with anyone else through a partnership arrangement and asks questions on number of partners, percentage of contribution for input expenses and labor along with percentage of catch allotted to the household for sale or self consumption. The second and third sections cover hired labor. For hired labor, the number of fishing men, women and children and associated time investment is included alongside information on wages, share of boat catch and share of boat revenue. All questions focus on the months listed as high season in Section B. This section is at the household level and the variable to uniquely identify observations is "y2_hhid."

SECTION E

This module is divided into three data files.

The first part focuses on fishing input/gear used by any member of the household in the last high season and includes questions on the number of fishing gears operated, owned, purchased, rented and associated rental value for purchase. The unique identifiers for this dataset are "y2_hhid" and "gearid." The descriptions of the included types of gear are below.

| |
|---|
| Gear ID 1 &2: **Beach Seines**. These include Mosquito net, Chambo seine, Kambuzi seine, and Matemba seine. Similar in construction except for headline length and mesh size. |
| Gear ID3: **Long-lines:** Passive gear consists of a strong length of cord with mono-filament traces and hooks attached at intervals. The hooks are baited with pieces of fish. The long-line is then weighted to the bottom and is generally set overnight and lifted following morning. Long-line hooks are generally larger than those on handlines. **Handlines**: Consist of mono-filament nylon with hooks attached and at the bottom of the line a weight is attached. The hooks are baited with earthworms or Usipa depending on fish being targeted. |
| Gear ID 4: **Gill nets**: Rectangular gear usually surface set or bottom set and used normally as passive gear. Set in the morning and retrieved the following morning. But at times this net may be used as active gear in open-water operated like *chilimira* net; slowly dragged behind two boats; set in shallow water and fish chased into it by pounding the water-*chiombera.* |
| Gear ID 5: **Fish Traps *Mono***: Generally funnel-valve made of bamboo set in shallow river or lake areas to catch *chambo* and predators like *Mlamba* overnight. The fish trap may be used with a weir or fence, which serves to guide the fish into the trap. |
| Gear ID 6: **Cast Net**: Conical shape with footrope weighted with small stones. Generally used by two people, one paddles the other casts while standing in front of the canoe indicating to the paddler in which direction he wishes to be propelled. Immediately prior to the net being cast, the paddling ceases. As soon as the thrown net sinks to the bottom, the canoe is propelled forward so that the cast net is retrieved almost vertically. |
| Gear ID 7: **Large Fish Trap**: Wooden box with a lattice construction that allows water to pass through. The box has one entry point in the middle that allows fish to enter, but prevents them from leaving. A weight is attached to the box and food is kept inside to attract the fish. |
| Gear ID 8: **Night boat fishing**: One large boat is surrounded by smaller boats. The smaller boats put out lanterns to attract the fish toward the larger boat to be caught. |

The second part focuses on boats/engines used by any member of the household in the last high season and includes questions on the number of boats operated, owned, purchased, rented and associated rental value for purchase. The unique identifiers in this dataset are "y2_hhid" and "boatengine_id."

The third part focuses on other costs associated with purchases during the last high fishing season such as taxes, licenses, auction fees, wicks, rent for storage, transportation, buoys, thread for net sewing, beeswax/sealant, lubricant and chicken wire. The associated costs and units consumed are also included for each of these categories. The unique identifiers in this dataset are "y2_hhid" and "inputid."

SECTION F

This section focuses on the output of fishing activities. Questions related to type of fish caught, area fished in, quantity caught, fish processing mechanisms and sales and in house consumption are included in this section. The best effort has been made to categorize fish into groups, but due to the wide variation of local names for fish species across Tanzania, a large number "others" remain. Approximately 20% of households indicate having caught another type of fish than those on the list. Households are permitted to list up to five different types of fish, and the unique identifiers are "y2_hhid" and "fishid."

SECTION G

This section deals with fishing gear rented out in the last high season with a focus on number of units of each fishing gear rented and the rental amount received for it. The unique identifiers are "y2_hhid" and "gearid."

SECTION H

This section is only administered to those household members that were engaged in fish trading in the last high season. Questions on type of fish sold, average sales per week, quantity of fish purchased from other sources, costs related to fish trading are included. The data is contained in three data files. The first deals with purchases and sales of fish, and the unique identifying variables are "y2_hhid" and "fishcode." The second covers costs associated with the fish trading and the identifying variables are "y2_hhid" and "costcode." The final data file contains the screener questions at the end of the high season fishing activities.

SECTION I – N

Sections I through N repeat the same questions as above but for the low season.

## *COMMUNITY QUESTIONNAIRE*

For the purposes of this survey a "community" is defined as the village in rural areas and the mtaa in which the EA is located in urban areas. The community questionnaire was administered to a group of local leaders determined by the field supervisors. In general, in rural areas this group included the ward executive officer, village chairperson and the VEO, as well as other members from the village council. In urban areas the group included the ward executive officer, mtaa chairperson and possibly other local leaders. Note that not all EAs have a corresponding community questionnaire. Particularly in urban areas, EAs within the same ward share the same administration and therefore community level information. In the 2010/2011 dataset, there are 16 EAs that do not have individual level community information, but it should be considered to be the same as the EA within the same region, district, and ward in the dataset.

In addition, individuals that moved to new communities since the 2008/2009 would not have corresponding community information as this information was only collected for the originally selected EAs.

SECTION CA

This section covers basic information on community identification including region, district, ward, and enumeration area. It also includes the interviewers ID code, in most cases the supervisor, and the direct observation questions. All other variables in this section, including the name of the village and the GPS coordinates, are excluded for confidentiality purposes. Also included in this section, however, are the regional capital identifier and the location at which the prices were gathered in section CJ.

SECTION CB

This section looks at access to basic services in terms of distance and associated transportation costs for these services. Services included in this model are regional and district headquarters, presence of government and private primary and secondary schools, presence of government and private health clinics and hospitals, weekly and daily markets, access to financial services such as ATMs and mobile money agents, police station, court, community set water taps, and livestock services such as slaughter slabs, veterinary center, charcos dam and hide and skin bandas. The name of the institution is dropped for confidentiality purposes.

SECTION CC

Questions in this section look at recent construction projects and sources of funds and associated amounts. Categories included are road construction/maintenance, market construction/maintenance, water supply such as wells and pumps, school construction and maintenance at pre-primary, primary and secondary levels, health and veterinary services, irrigation schemes and grain storages.

SECTION CD

The first part of this module focuses on land use related issues with estimated percentages for how village land is used (cultivation, forest, pasture, wetland, residential, business etc.). The second part deals with reasons for re-allocation of land (if any), number of households affected and associated compensation. This section contains a fairly high percentage of missing values as sometimes key informants did not know all the information asked in the questionnaire.

SECTION CE

This module includes questions on the presence of community level agricultural cooperatives, number of farmers involved in such cooperatives and activities undertaken by the cooperative. Questions related to SACCOs and production of maize is also included in this section. The name of the nearest supplier of improved maize seeds (cm_e07) has been dropped for confidentiality purposes. There are 33 EAs in Dar es Salaam which did not receive this module because they were considered by the supervisor to be sufficiently urban that the module was not applicable.

SECTION CF

This section deals with demographic details and traditional norms within the community, including migration, languages spoken, number of divorce cases by type of divorce process, and wealth ownership. The names of the other languages spoken in the villages have been removed for confidentiality purposes, though this information may be available upon application from the Tanzania National Statistics Bureau.

SECTION CG

This section includes questions on village level governance such as number of assemblies, presence of ward tribunal, and related costs and procedure.

SECTION CH

This module focuses on issues related to water and sanitation, with questions on sources of drinking water, type of latrine facilities, and associated costs.

SECTION CI

This module consists of a roster of community leaders, with background questions such as gender, age, occupation, political affiliation, education, household assets, and food consumption patterns of the leader's household. The location of the leader's residence has been removed for confidentiality reasons.

SECTION CJ

This section contains information on market prices for the surveyed communities. The prices are listed for two locations, the first at the village level and the second at the district capital area. If district prices have already been recorded in another community questionnaire, the cluster ID of that questionnaire is entered at the top of the market prices table and district prices are skipped. The regional capital identifier and the location at which the prices were gathered are included in the section CA dataset. The GPS coordinates are removed for confidentiality purposes.

## Appendix A: Confidential Information, Geospatial Variables

To maintain the confidentiality of our respondents, certain parts of the TZNPS database have not been made publicly available. The confidential variables pertain to (i) names of the respondents to the household and community questionnaires, (ii) village and constituency names, (iii) descriptions of household dwelling and agricultural plot locations, (iv) phone numbers of household members and their reference contacts, (v) GPS-based household and agricultural plot locations, (vi) names of the children of the head/spouse living elsewhere, (vii) names of the deceased household members, (viii) names of individuals listed in the network roster, and (ix) names of field staff.

To increase the use of the TZNPS data, a set of geospatial variables has been provided by using the georeferenced plot and household locations in conjunction with various geospatial databases that were available to the survey team. The table in Appendix A provides the name, type, source, reference period, resolution, description, and source of each variable.

The geovariables are stored in two data files, one at the household-plot-level, and the other at the household-level. The plot-level file, named **Plot.Geovariables**, contains one geospatial variable measuring plot distance to household and the observations are uniquely identified by the combination of **y2_hhid plotnum**. The observations included in this file are rainy season plots that are owned and/or cultivated by the household and that have been visited for GPS-based land area measurement.

The rest of the geovariables are stored in **HH.Geovariables** and the observations are uniquely identified by **y2_hhid.** To partially satisfy the demand for georeferenced household and community locations while preserving the confidentiality of sample household and communities, we have computed the average of household GPS coordinates in each EA, applied a random offset within a specified range to the average EA value (following the MeasureDHS methodology) and provided the off-set EA latitudes and longitudes are part of **EA.Offsets.**

More specifically, the coordinate modification strategy relies on random offset of cluster center-point coordinates (or average of household GPS locations by EA in TZNPS2) within a specified range determined by an urban/rural classification. For urban areas a range of 0-2 km is used. In rural areas, where communities are more dispersed and risk of disclosure may be higher, a range of 0-5 km offset is used. An additional 0-10 km offset for 1% of rural clusters effectively increases the known range for all rural points to 10 km while introducing only a small amount of noise. Offset points are constrained at the district level, so that they still fall within the correct district for spatial joins, or point-in-polygon overlays. The result is a set of coordinates, representative at the EA level, that fall within known limits of accuracy. Users should take into account the offset range when considering different types of spatial analysis or queries with the data. Analysis of the spatial relationships between locations in close proximity would not be reliable. However, spatial queries using medium or low resolution datasets should be minimally affected by the offsets.

All geospatial variables have been produced by using the unmodified GPS data. These include extensive measures of distance, climatology, soil and terrain and other environmental factors.

Time-series on rainfall and vegetation have also been used to describe the survey agricultural season relative to normal conditions. These variables are intended to provide some understanding of how geophysical characteristics vary at the landscape level.

## Table 5: Household Geovariable Description

| Theme | Source | Dataset Title | Variable Name | Variable Type | Reference Period | Resolution | Description | Web |
|---|---|---|---|---|---|---|---|---|
| **Distance** | LSMS-ISA | Plot Distance to Household | **dist_hh** | Continuous | N/A | N/A | Plot distance to household | |
| | OpenStreetMaps, TANROADS | Household Distance to Main Road | **dist01** | Continuous | N/A | N/A | Household distance to nearest trunk road (as defined by TANROADS) | |
| | City population | Household Distance to Towns | **dist02** | Continuous | N/A | N/A | Household distance to nearest town of > 20,000 pop | http://www.citypopulation.de |
| | USAID FEWSNET | Household Distance to Key Market Centers | **dist03** | Continuous | N/A | N/A | Household distance to nearest major market (FEWSNET key market centers) | |
| | Tracks4Africa, roads, borders | Household Distance to Border Posts | **dist04** | Continuous | N/A | N/A | Household distance to nearest border post on main road | |
| | World Gazeteer Towns, Statoids | Household Distance to District Headquarters | **dist05** | Continuous | N/A | N/A | Household distance to to the headquarters of the district of residence | www.statoids.com |

| Theme | Source | Dataset Title | Variable Name | Variable Type | Reference Period | Resolution | Description | Web |
|---|---|---|---|---|---|---|---|---|
| **Climatology** | UC Berkeley | WorldClim Bioclimatic Variables | **clim01** | Continuous | 1960-1990 | 0.008333 dd | Average annual temperature calculated from monthly climatology, multiplied by 10 (°C) | http://www.worldclim.org/bioclim |
| | UC Berkeley | WorldClim Bioclimatic Variables | **clim02** | Continuous | 1960-1990 | 0.008333 dd | Average temperature of the wettest quarter, from monthly climatology, multiplied by 10. (°C) | http://www.worldclim.org/bioclim |
| | UC Berkeley | WorldClim Bioclimatic Variables | **clim03** | Continuous | 1960-1990 | 0.008333 dd | Total annual precipitation, from monthly climatology (mm) | http://www.worldclim.org/bioclim |
| | UC Berkeley | WorldClim Bioclimatic Variables | **clim04** | Continuous | 1960-1990 | 0.008333 dd | Precipitation of wettest month, from monthly climatology (mm) | http://www.worldclim.org/bioclim |
| | UC Berkeley | WorldClim Bioclimatic Variables | **clim05** | Continuous | 1960-1990 | 0.008333 dd | Precipitation of wettest quarter, from monthly climatology (mm) | http://www.worldclim.org/bioclim |

| Theme | Source | Dataset Title | Variable Name | Variable Type | Reference Period | Resolution | Description | Web |
|---|---|---|---|---|---|---|---|---|
| **Landscape Typology** | ESA and UC Louvain | GlobCover v 2.3 | **land01** | Categorical | 2009 | 0.002778 dd | Majority landcover class within approximately 1km buffer | http://ionia1.esrin.esa.int/ |
| | ESA and UC Louvain | GlobCover v 2.3 | **land02** | Continuous | 2009 | 0.002778 dd | Percent under agriculture within approx 1 km buffer | http://ionia1.esrin.esa.int/ |
| | IFPRI | IFPRI standardized AEZ based on elevation, climatology | **land03** | Categorical | | 0.008333 dd | Agro-ecological zones created using WorldClim climate data and 0.0833dd resolution LGP data from IIASA. | http://harvestchoice.org/production/biophysical/agroecology |

| Theme | Source | Dataset Title | Variable Name | Variable Type | Resolution | Description | Web |
|---|---|---|---|---|---|---|---|
| Soil & Terrain | NASA | SRTM 90m | **soil01** | Continuous | 0.000833 dd | Elevation (m) | ftp://xftp.jrc.it/pub/srtmV4/arcasci/ |
| | USGS | Slope (percent) | **soil02** | Continuous | 0.008333 dd | Derived from 90m SRTM | http://pubs.usgs.gov/of/2007/1188/, data provided USGS upon request |
| | AfSIS | Topographic Wetness Index | **soil03** | Continuous | 0.000833 dd | Derived from modified 90m SRTM. Local upslope contributing area and slope are combined to determine the potential wetness index: WI = ln (A s / tan(b) ) where A s is flow accumulation or effective drainage area and b is slope gradient. | http://www.ciesin.columbia.edu/afsis/bafsis_fullmap.htm# |
| | LSMS-ISA | Terrain Roughness | **soil04** | Categorical | 0.000833 dd | Derived from 90m SRTM using Meybeck relief classes and 5x5 pixel neighborhood | |
| | FAO | Harmonized World Soil Database | **soil05** | Categorical | 0.083333 dd | Nutrient availability | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |
| | FAO | Harmonized World Soil Database | **soil06** | Categorical | 0.083333 dd | Nutrient retention capacity | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |
| | FAO | Harmonized World Soil Database | **soil07** | Categorical | 0.083333 dd | Rooting conditions | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |
| | FAO | Harmonized World Soil Database | **soil08** | Categorical | 0.083333 dd | Oxygen availability to roots | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |
| | FAO | Harmonized World Soil Database | **soil09** | Categorical | 0.083333 dd | Excess salts | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |
| | FAO | Harmonized World Soil Database | **soil10** | Categorical | 0.083333 dd | Toxicity | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |
| | FAO | Harmonized World Soil Database | **soil11** | Categorical | 0.083333 dd | Workability (constraining field management) | http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ |

| Theme | Source | Dataset Title | Variable Name | Variable Type | Reference Period | Resolution | Description | Web |
|---|---|---|---|---|---|---|---|---|
| **Crop Season Parameters** | NOAA CPC | Rainfall Estimates (RFE) | **crops01** | Continuous | 2001-2011 | 0.1 dd | Average 12-month total rainfall (mm) for July-June | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crop02** | Continuous | 2001-2011 | 0.1 dd | Average total rainfall in wettest quarter (mm) within 12-month periods from July-June | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops03** | Continuous | 2001-2011 | 0.1 dd | Average start of wettest quarter in dekads 1-36, where first dekad of July =1 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops04** | Continuous | 2009-2010 | 0.1 dd | 12-month total rainfall (mm) in July-June, starting July 2009 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops05** | Continuous | 2009-2010 | 0.1 dd | Total rainfall in wettest quarter (mm) within 12-month periods starting July 2009 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops06** | Continuous | 2009-2010 | 0.1 dd | Start of wettest quarter in dekads 1-36, where first dekad of July 2009 =1 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops07** | Continuous | 2010-2011 | 0.1 dd | 12-month total rainfall (mm) in July-June, starting July 2010 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops08** | Continuous | 2010-2011 | 0.1 dd | Total rainfall in wettest quarter (mm) within 12-month periods starting July 2010 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |
| | NOAA CPC | Rainfall Estimates (RFE) | **crops09** | Continuous | 2010-2011 | 0.1 dd | Start of wettest quarter in dekads 1-36, where first dekad of July 2010 =1 | ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops10** | Continuous | 2001-2011 | 0.004176 dd | Average total change in greenness (integral of daily EVI values) within primary growing season, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops11** | Continuous | 2001-2011 | 0.004176 dd | Average timing of onset of greenness increase in day of year 1-356, where Jul 1 = 1, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops12** | Continuous | 2001-2011 | 0.004176 dd | Average timing of onset of greenness decrease in day of year 1-356, where Jul 1 = 1, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops13** | Continuous | 2009-2010 | 0.004176 dd | Total change in greenness (integral of daily EVI values) within primary growing season for July 2009 - Jun 2010, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops14** | Continuous | 2009-2010 | 0.004176 dd | Onset of greenness increase in day of year 1-356, starting July 1 2009, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops15** | Continuous | 2009-2010 | 0.004176 dd | Onset of greenness decrease in day of year 1-356, starting July 1 2009, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops16** | Continuous | 2010-2011 | 0.004176 dd | Total change in greenness (integral of daily EVI values) within primary growing season for July 2010 - Jun 2011, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops17** | Continuous | 2010-2011 | 0.004176 dd | Onset of greenness increase in day of year 1-356, starting July 1 2010, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |
| | BU | MOD12Q2 (DOY 185) Land Cover Dynamics from MODIS | **crops18** | Continuous | 2010-2011 | 0.004176 dd | Onset of greenness decrease in day of year 1-356, starting July 1 2010, averaged by district | ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005, DOY185 version provided upon request from MODIS Land Cover Group at Boston University |

## Appendix B: Calculation of Panel Weights

The methodology described in this paper builds upon published documentation from established panel surveys, such as the Panel Study of Income Dynamics [PSID], conducted since 1968 by the Institute for Social Research at the University of Michigan; and the British Household Panel Survey [BHPS], whose first 13 waves were conducted between 1991 and 2003 by Institute for Social and Economic Research at the University of Essex. Both the PSID and the BHPS are nationally-representative panel surveys in the USA and the UK respectively.

The weights are developed in eight steps:

1) Begin with the "base weights" or those calculated during the first round of the survey;
2) incorporate fair-share weights for composition changes;
3) derive attrition adjusted weights for all individuals, including split-off[10] households, then aggregate these weights to the household level;
4) post-stratify the pooled weights to known population totals.

Each of these steps in discussed in detail below.

*1) Base Weights from 2008/2009 Sample*

The panel weight calculations are based on the 2008/2009 household weights. These weights are based on the inverse probability of selection, EA level non-response correction, trimming of outlier weights, and a post-stratification correction[11]. These probability weights form the first component of the 2010/2011 calculations.

$$W_1 = W_{2008}$$

*2) Fair Share Correction*

Based on the tracking protocols, the tracking for split off rules for the TZNPS allow for the incorporation of people who now live with original sample members. For example a young adult living with his parents in 2008, may be 2010 have formed a new household, getting married and having a child. The wife and infant will be incorporated into the survey and thus require a probability of selection. Such corrections are routinely used to distribute weight to new sample members in panel surveys. See Rendtel and Harms (2009) for a discussion of several different methods of weight correction.

Because split-off individuals are tracked and interviewed in their new households, there are multiple ways that a household can become part of the survey.

- Either by being selected initially for the first round of the TZNPS.

[10] For the purposes of this note, 'parent' refers to the household found at the same location as the previous round of data collection, and 'split-off' refers to new households entering the sample through an individual originally resident in a parent household during a previous round. Since this distinction is arbitrary, however, there is no mathematical difference between the parent and split households.

[11] See BID for the first year of the panel for further details on the calculations of the base weights.

- By receiving a member that came from a household that was selected for the first round of the TZNPS.

In an ideal world, it would be possible to know the probability of selection that each new member brought into the household, and adjust the household weight accordingly. This is necessary since households receiving members have higher probabilities of selection (and therefore lower weights) because the household could have been selected in multiple ways. Since we cannot know the probabilities of every member, we must make simplifying assumptions. The first simplifying assumption is that the arriving members arrived together from one other household. This would be the case if a man and woman get married and set up a new household, or in the case of an older relative moving in with adult children. In certain cases, however, arriving members come from more than one household. Assuming only two source households underestimates slightly the probability of selection (and therefore over-estimates the weights). Incidence of these cases is believed to be relatively rare, and any resulting bias should be negligible. The second simplifying assumption we make is that the arriving members have the same probability of selection, on average, as those members that are already there. This would not be true on a case-by-case basis but would be true in the aggregate. With these simplifying assumptions, we add a factor of ½ for all households, 'split' or 'parent' that have new members arriving from other households. This takes into account the fact that they could have been selected in two ways, and assumes the probability of selection is equal.[12]

$$a_1 = \begin{cases} 1 & \text{otherwise} \\ \frac{1}{2} & \text{if new members} \end{cases}$$

Then the adjusted weights would be:

$$W_2 = W_1 * a_1$$

A limitation of the panel methodology is that the represented population is not identical to the 2010 Tanzanian household population, as it does not include immigrants in new households. Inclusion of these groups would necessitate refreshing the sample with new households. However, the represented population is close enough to the 2010 Tanzanian population to permit the desired cross-sectional estimates.

*3) Attrition Correction Factor*

All household panel surveys must tackle the problem of attrition, sample members selected for follow interview which cannot be located and/or interviewed. The methodology used to adjust weights for attrition in the UNPS follows Rosenbaum & Rubin (1984). We use predicted response probabilities from a logistic regression model based on the covariates to form the weighting classes or cells. This approach has also been adopted in the PSID; see for example, Gouskova (2008).

[12] New births and arriving children under age 4 do not count as 'new members' in this case because they could not have been selected in 2005.

The attrition correction in the case of the UNPS needs to take into account two distinct sources of attrition: entire households that are not found and split-off individuals that are selected for tracking but not found. The two potential options for the calculations are (1) to treat the split-off households as household heads and do the calculations at the level of the household, or (2) to treat the households that are not found as individuals and perform the calculations at the individual level. The first option is problematic as the characteristics of household heads are dissimilar to the characteristics of split-offs (see table 2). Therefore in the TZNPS, the second methodology was employed.

The total sample size for the second round of the NPS has a total sample size of 3924 households. This represents 3168 round one households, a re-interview rate of over 97 percent. In addition, of the 10,420 eligible adults (over age 15 in 2010), 9,338 were re-interviewed, a re-interview rate of approximately 90 percent.

To obtain the attrition adjustment factor the probability that a sample household was successfully re-interviewed in the second round of surveys is modeled with the linear logistic model at the level of the individual. A binary response variable is created by coding the response disposition for eligible households that do not respond in the second round as 0, and households that do respond as 1[13].

Then a logistic response propensity model is fitted, using 2005 UNHS household and individual characteristics measured in the first wave as covariates. Included covariates are:

- gender
- age
- marital status
- current school attendance
- years of education
- labor force participation
- physically or mentally handicapped
- household size
- rural / urban status
- household consumption
- square of household consumption
- residence in agricultural (crop) household
- residence in a livestock household
- residence in a fishing household
- residence in household owning enterprise
- residence in household receiving transfer income
- residence in dwelling with improved walls
- residence in dwelling with improved roof
- residence in dwelling with improved floor
- residence in a household with receiving wage income
- residence is a titled dwelling
- residence in a rental dwelling
- residence in household with at least one member owning mobile phone
- region of residence[14]

[13] Note that only household members who have died are excluded from the attrition calculations. In some rare cases, there are eligible household members who were selected for tracking but for whom the field teams did not actually search. Possible reasons could include international migration or lack of time on the part of the field teams.

[14] Zanzibar regions are aggregated into Unguja (region 51, 52, and 53) and Pemba (region 54 and 55) islands due to a lack of variable with the disaggregated categories.

In a few limited cases, values of unit level variables were missing from the 2008/2009 household dataset. These values were imputed using multivariate regression and logistic regression techniques. Imputations are done using the 'impute' command in Stata at the level of the UNPS strata (urban/rural and region). Overall, less than one percent of the variables required imputation to replace missing values.

The estimated logistic model is used to obtain a predicted probability of response for each household member in the 2010/2011 survey. These response probabilities were then aggregated to the household level (by calculating the mean), the using the household-level predicted response probabilities as the ranking variable, all households are ranked into 10 equal groups (deciles). An attrition adjustment factor was then defined as the reciprocal of the empirical response rate for the household-level propensity score decile.

Then the adjusted weights would be:

$$W_3 = W_2 * ac$$

*4) Post-stratification*

To reduce the overall standard errors, and weight the population totals up to the known population figures, a post-stratification correction is applied. Based on the projected number of households in the urban and rural segments of each region, adjustment factors are calculated. This correction also reduces overall standard errors (see Little et al, 1997).

$$W_{panel} = W_3 * ps$$

# Appendix C. Consumption Aggregate

This Appendix explains the steps involved in the construction of the consumption measure and describes the estimation of the nominal household consumption.

## 1.0 The construction of the consumption aggregate

Creating the consumption aggregate is guided by theoretical and practical considerations. First, it must be as comprehensive as possible given the available information. Omitting some components assumes that they do not contribute to people's welfare or that they do not affect the ranking of the population. Second, market and non-market transactions are to be included, which means that purchases are not the sole component of the indicator. Third, expenditure is not consumption. For perishable goods, mostly food, it is usual to assume that all purchases are consumed. However, for other goods and services, such as housing or durable goods, corrections have to be made. Fourth, a common reference period should be chosen. Typically each consumption module in a survey has a different reference period, for instance, education could refer to the last 12 months, food could refer to the last week, and health could refer to the last month. Following common practice in Tanzania, consumption will be reported per 28 days.

### 1.1 Food component

A few general principles are applied in the construction of this component. First, all possible sources of consumption are included. This means that the food component comprises not only consumption from purchases in the market or from meals eaten away from home but also food that was produced by the household or received as a gift. Second, only food that was actually consumed, as opposed to total food purchases or total home-produced food, enters into the consumption aggregate. Third, non-purchased consumed food needs to be valued and included in the welfare measure. The NPS gathers information on the amount spent on purchases and on the quantity purchased for all food items. A measure of prices, or rather a measure of unit values, can be obtained by dividing the expenditure by the quantity and can be used to value own-consumption or food received as a gift.

### 1.2 Non-food component

Data on an extensive range of non-food items are usually available: utilities such as water, kerosene, electricity, health, transportation, communications, recreation, education, furnishings, personal care, etc. Unlike food, the NPS only collects data on purchases of non-food items, that is, the survey assumes that the consumption of non-food goods and services coming from own-production, from gifts or from other sources is negligible and can be ignored. In addition, the

NPS does not gather information on quantities purchased because most non-food items are too heterogeneous to try to calculate prices.

Each non-food component is associated with a particular reference period, which reflects the frequency of that purchase or consumption. For instance, expenses on public transportation are collected for the last seven days, expenses on mobile phones and personal care are collected for the last month, and expenses on furnishings and small appliances for the last twelve months.

The information about some non-food goods and services needs to be excluded from the consumption aggregate because those items are not consumption. Payments of mortgages or debts are financial transactions and not consumption. Losses to theft are neither expenditure nor consumption. Remittances to other households are expenditures but not consumption. Expenditures on marriages, dowries, births and funerals are consumption but given their sporadic nature and the fact that the reported amounts are typically rather large, this consumption is left out to avoid overestimating the true level of welfare of the household.

### 1.3 Durable goods

Ownership of durable goods could be an important component of the welfare of the households. Given that these goods last for many years, the expenditure on purchases is not the proper indicator to consider. The right measure to estimate, for consumption purposes, is the stream of services that households derive from all durable goods in their possession over the relevant reference period. This flow of utility is unobservable but it can be assumed to be proportional to the value of the good. Information on the number of durable goods owned, their age, and their value (current or original) is required to estimate this component of consumption. Unfortunately, the NPS only provides data on the number of durable goods owned by the household. Calculating this consumption component would have involved making assumptions about their age, their current value and their lifespan. This might have resulted in an extremely imprecise estimation, thus it was decided to exclude this component from the consumption aggregate.

### 1.4 Housing

Housing conditions are considered to be an essential part of people's living standards. Nonetheless, in most developing countries limited or nonexistent housing rental markets pose a difficult challenge for the estimation and inclusion of this component in the consumption aggregate. As in the case of durable goods, the objective is to measure the flow of services received by the household from occupying its dwelling. When a household lives in a rented dwelling, and provided rental markets function well, that value would be the actual rent paid. If enough families rent dwellings, imputations can be made for those families that own their dwelling. It is common to include a question for homeowners asking them to provide the hypothetical rent they would pay for renting their dwelling. These self-reported rents can in principle be used to value the consumption the household gets from occupying its dwelling, but these amounts are not always credible or usable, particularly in rural areas where very few households rent. If imputed rents cannot be estimated, actual rents must be excluded from the

consumption aggregate for the sake of consistency. The NPS does not collect information on imputed rents and given that the number of households living in rented dwellings is fairly small, this component was excluded from the consumption aggregate.

## 2.0 Price adjustment

Nominal consumption of the household must be adjusted for cost-of-living differences. Temporal and spatial price adjustments are required to adjust consumption to real terms. Temporal differences are associated with the duration of the fieldwork (TSh 1,000 in October 2010 may not have the same value as in August 2011) as well as with the different recall periods (TSh 1,000 spent in the last month may not have the same value as in the last quarter or in the last year). Spatial differences are associated with the location of households interviewed in the survey (TSh 1,000 in Dar es Salaam may not have the same value as in Ruvuma).

The price index required to adjust nominal consumption could come partly or fully from the NPS. A price index is a combination of prices and budget shares in a base and a comparison period. The budget shares are the weights that each commodity has in the index and are equivalent to their share in the cost of the bundle being analysed. The NPS can provide information on budget shares for all items, but information on prices (unit values) only for food items. Two possible price indices could be constructed: a price index based only on food items (the assumption would be that non-food items show the same temporal and spatial differences than food items) or a price index that takes into account both food and non-food by combining information from the survey (food prices and weights for food and non-food items) and the official consumer price index (non-food prices).

Fisher price indices based only on food items were employed to adjust the nominal consumption aggregate for spatial and temporal price differences. Fisher price indices do a better job than Laspeyres or Paasche price indices at capturing differences in consumption patterns across domains as a consequence of differences in relative prices. They also avoid overstating or understating the true inflation (as would be the case with Laspeyres and Paasche respectively).[15] Price indices were estimated by stratum and quarter (a period of three consecutive months) and the base period comprises the entire period of each round of the NPS – that is, price indices were calculated separately for each round. A price index by stratum and month would have been ideal, but complications arose with the sample size because in some combinations of stratum and month few households were interviewed. Price indices by stratum and quarter might not be as precise as price indices by stratum and month but they provide more robust results. Fisher price indices by stratum and quarter were constructed using the following formula:

$$F_i = \sqrt{L_i P_i}$$

where *i* is a combination of stratum and quarter, L refers to a Laspeyres price index and P refers to a Paasche price index. The Laspeyres and Paasche price indices are defined as

[15] See Deaton and Tarozzi (2000).

$$L_i = \sum_{k=1}^{n} w_{0k}\left(\frac{p_{ik}}{p_{0k}}\right), P_i = \left[\sum_{k=1}^{n} w_{ik}\left(\frac{p_{ik}}{p_{0k}}\right)^{-1}\right]^{-1}$$

where $w_{0k}$ is the average household budget share of item $k$ in the country, $w_{ik}$ is the average household budget share of item $k$ in stratum and quarter $i$, $p_{0k}$ is the national median price of item $k$ and $p_{ik}$ is the median price of item $k$ in stratum and quarter $i$.

Food items that had been purchased by at least 10 households by stratum and quarter were included in the construction of the price indices. Residual or catch-all food categories were also excluded because their unit values effectively mix several items. The share of the bundle considered for the price indices with respect to total food consumption is similar in both rounds of the NPS: it stands at around 67% at the national level and goes from 63% in rural mainland to more than 80% in Dar es Salaam and Zanzibar. Median unit values were estimated for the price indices because the median is less sensitive to outliers than the mean.

Table 2.1 shows the Fisher food price indices for each round of the NPS. Spatial price differences across strata remain fairly constant over time. The most expensive stratum is Dar es Salaam whereas the cheapest is rural areas in mainland. The cost of living in other urban areas in mainland and Zanzibar is relatively similar. Temporal price differences across quarters are noticeably larger during the NPS2, thus reflecting a higher inflation in the second round compared to the first round.

**Table 2.1: Fisher food price indices by stratum and quarter, NPS1 and NPS2**

| NPS1 | Oct-Dec 2008 | Jan-Mar 2009 | Apr-Jun 2009 | Jul-Sep 2009 |
|---|---|---|---|---|
| Dar es Salaam | 1.08 | 1.18 | 1.20 | 1.15 |
| Other urban | 1.00 | 1.04 | 1.04 | 1.04 |
| Rural | 0.92 | 0.86 | 0.92 | 0.96 |
| Zanzibar | 1.03 | 1.06 | 1.07 | 1.07 |

| NPS2 | Oct-Dec 2010 | Jan-Mar 2011 | Apr-Jun 2011 | Jul-Sep 2011 |
|---|---|---|---|---|
| Dar es Salaam | 1.05 | 1.11 | 1.17 | 1.18 |
| Other urban | 0.90 | 0.97 | 1.06 | 1.08 |
| Rural | 0.87 | 0.86 | 0.98 | 1.02 |
| Zanzibar | 0.89 | 0.98 | 1.06 | 1.07 |

**Updating monetary figures across rounds of the NPS**

Price indices will also be required to update monetary figures across both rounds of the NPS. The price indices from Table 1 are used to adjust nominal consumption for cost of living differences within each round of the NPS. Yet it would not be correct to compare real

consumption at NPS1 prices with real consumption at NPS2 prices. Either NPS1 figures should be adjusted to NPS2 prices or NPS2 figures should be adjusted to NPS1 prices.

Fisher price indices based only on food items were employed to adjust consumption for spatial and temporal price differences across rounds of the NPS. It was assumed that non-food goods and services show the same temporal and spatial price differences across rounds than food items. Price indices were estimated for the entire country and for the full extent of each round: the base period was the 12 months of the NPS1 and the comparison period was the 12 months of the NPS2.

Food items that had been purchased by at least 50 households in the country were included in the construction of the price indices. As with the previous price indices, residual food categories were also excluded and median rather than mean unit values were used. The share of the bundle considered for the price indices with respect to total food consumption is similar in both rounds of the NPS: it stands at around 98%. The Fisher food price index across the NPS1 and the NPS2 was estimated at 1.21, that is, the cost of an average food bundle consumed in the country increased by 21% between rounds of the NPS. This inflation will be employed to adjust the consumption aggregate and the poverty lines across the NPS1 and the NPS2.

## 3.0 Household composition adjustment

The final step in constructing the welfare indicator involves going from a measure of standard of living defined at the household level to another at the individual level. Ultimately, the concern is to make comparisons across individuals and not across households. Two types of adjustments have to be made to correct for differences in composition and size. The first relates to demographic composition. Household members have different needs based mainly on their age and gender, although other characteristics can also be considered. Equivalence scales are the factors that reflect those differences and are used to convert all household members into "equivalent adults". For instance, children are thought to need a fraction of what adults require, thus if a comparison is made between two households with the same total consumption and equal number of members, but one of them has children while the other comprises only adults, it could be expected that the former will have a higher individual welfare than the latter. Unfortunately there is no agreement on a consistent methodology to calculate these scales. Some are based on nutritional grounds, but while a child may need only 50% of the food requirements of an adult, it is not clear why the same scale should be carried over non-food items. It may very well be the case that the same child requires a larger proportion than the adult in education or clothing.[16]

The second adjustment focuses on the economies of scale in consumption within the household. The motivation for this is the fact that some of the goods and services consumed by the household have characteristics of "public goods". A good is said to be public when its consumption by a member of the household does not necessarily prevent another member from consuming it as well. Examples of these goods could be housing and durable goods. For example, one member watching television does not preclude another from watching too. Larger households may need to spend less to be as well-off as smaller ones. Hence, the bigger the share

[16] See Deaton and Muellbauer (1986) or Deaton (1997).

of public goods in total consumption, the larger the scope for economies of scale. On the other hand, private goods cannot be shared among members – once one household member has consumed them, no other member can. Food is the classic example of a private good and, for instance, in poor economies, where food represents a sizeable share of the household budget, little room exists for economies of scale.

Poverty analysis in Tanzania employs an adult-equivalent scale to implement these two adjustments (see Table 3.1). In general, children are thought to consume less than adults and women less than men. An alternative and common practice would have been to use a per capita adjustment for household composition. This is a special case of both adjustments and implies that children consume as much as adults and there is no room for economies of scale. In other words, all members within the household consume equal shares of the total consumption and costs increase in proportion to the number of people in the household. In general, per capita measures will underestimate the welfare of households with children with respect to families with no children, and the welfare of large households with respect to families with a small number of members.

**Table 3.1: Adult-equivalent scale by gender and age**

| Age (years) | Male | Female |
|---|---|---|
| 0-2 | 0.40 | 0.40 |
| 3-4 | 0.48 | 0.48 |
| 5-6 | 0.56 | 0.56 |
| 7-8 | 0.64 | 0.64 |
| 9-10 | 0.76 | 0.76 |
| 11-12 | 0.80 | 0.88 |
| 13-14 | 1.00 | 1.00 |
| 15-18 | 1.20 | 1.00 |
| 19-59 | 1.00 | 0.88 |
| 60 and more | 0.80 | 0.72 |

## Appendix D: List of Data Checks

Household Questionnaire

- Values fell beyond five standard-deviations from the sample mean.
- Age was not consistent for an individual's educational level.
- Total consumption was less than the sum of purchased and home-produced consumption.
- Sum of money spent on education does not equal the total amount recorded.
- The unit of measure was inconsistent across the questions for a single consumption item in Section K (i.e., purchased amounts were recorded in kilograms but total amounts were

recorded in grams).

- Non-zero monetary expenditure on a given item was accompanied by missing values for total consumption.
- The range of consumed goods went outside the established boundaries.
- More than 1 head of household was listed or there was no head of household listed.
- Individual modules of the questionnaire was not recorded.
- Individual id's were not consistent with the household roster.
- An individual was engaged in multiple primary activities or no primary activity.
- No non-farm businesses were listed if the household received non-farm loans or remittances.
- Duplicate household id's, unknown household id's, and item code id's were recorded.
- The skip patterns within the questionnaire were not properly followed.

Agriculture Questionnaire

- The purpose of the plot not listed.
- Crop, byproduct, and plot names/id's are not listed.
- Crop and byproduct quantity produced, harvested, and sold not listed.
- Total costs, expenses, and losses incurred while farming not listed.
- The year and number of plants and trees on a plot was not recorded.
- The last harvest date was not listed.
- The distance of the plot from the road and market not listed.
- The skip patterns within the questionnaire were not properly followed.

Community Questionnaire

- There are duplicate EA and community leaders codes listed.
- EA codes are not listed in each questionnaire module.
- There are missing or incorrect entries.
- There are an incorrect number of row items listed.

Fishery Questionnaire

- There is no fishery data but the respondent indicated that they engage in fishery activities.
- The number of weeks, days per week, and hours per day that the individual engaged in fishery activities is not recorded.
- The species of fish caught, bought, sold, or traded not recorded.
- The quantity and unit of fish caught, bought, sold, or traded not recorded.
- The location of where the individual engages in fishery activities not recorded.
- The type of fishing gear used or rented by individuals not recorded.
- The number of household members or hired labour engaged in fishery activities not recorded.
- The costs and expenses incurred from engaging in fishery activities not recorded.

## Appendix E: List of Definitions

GULIO (local market)
Gulio is a local market mostly at the Village level which can occur at any frequency – once a week, twice etc. One can get any sort of goods at a Gulio and it mostly operates in rural areas. Farmers get together at a certain place once a week and sell their produce. Note that these sellers travel across the country – it could be that every Monday they operate in Village A, every Tuesday in Village B, every Wednesday in Village C and so on.

SOKO KUU (main market)
Soko kuu is the main market that people go to that usually operates daily and they operate at the village or ward level. Soko Kuu is the primary market for all goods. The difference between Soko Kuu and Gulio is that Soko Kuu operates daily while Gulio operates on a fixed schedule basis, which is why they are referred to as the Main Market.

SOKO (market)
Soko is a small market located at street corner that sells few items like vegetables, and some other household goods. Soko's are frequently found in Dar for example. Usually these markets do not carry high value goods and have fewer items than a Soko Kuu.

MNADA (auction)
Mnanda is an English auction for a very specific good – like a cow for example. A minimum price is set, beyond which the bidders can go up to any amount. The highest bidder wins. It is an open auction in the sense that all the bidders find out how much each is bidding and can competitively bid more.

M/BIASHARA BINAFSI (private business person)
This is usually a vendor or a hawker who sells goods by walking door to door – like cigarette/water sellers or vegetable vendors in Dar.

DUKANI/MCHUUZI (grocery local merchant)
This is a small shop owner around street corners that sell very specific items. This could be a shop for stationery, shop for buying everyday essentials like bread, toothpaste, etc… or even a medical shop. Dukani/Mchuuzi sell very specific goods and are fixed shops owned by merchants unlike the Biasharas who are travelers.

## Appendix F: Tanzanian Educational System

Tanzania has 13 years of formal schooling – D1 to D7 and F1 to F6.

D1 - Grade 1 (1st year)
D2 - Grade 2 (2nd year)
D3 - Grade 3 (3rd year)
D4 - Grade 4 (4th year)
D5 - Grade 5 (5th year)
D6 - Grade 6 (6th year)
D7 - Grade 7 (7th year)

F1 - Form I (8th year)
F2 - Form II (9th year)
F3 - Form III (10th year)
F4 - Form IV (11th year)
F5 - Form V (12th year)
F6 - Form VI (13th year)

Prior to independence, there used to be a D8 – Grade 8. Additionally, all classes used to be taught in English but following independence, the Tanzanian educational system switched to being based in Swahili.

If an individual does not proceed to Form I (F1), they can take the MS+ Course. This is a vocational course – for jobs such as carpentry - that ranges from three months to a year. For a student to proceed to Form V (F5), they must the Form IV (F4) national level exam, which is also known as O+. It is important to note that the O+ does not constitute an extra year school. It is simply a required final exam an individual must take to advance educationally. So an individual could have finished Form IV (F4) but have failed the O+, thus not proceeding to Form V (F5). Students must also take a national level exam, A+, after completing For VI (F6). If they pass the A+ exam with a certain grade, they go directly to University (U1 through U5).

If one does not pass the A+ exam, one can do the Diploma course or drop out of school. If the individual chooses to participate in the Diploma course, afterward completion, they can enroll into University.

U1 – 14th year
U2 – 15th year
U3 – 16th year
U4 – 17th year
U5 – 18th year
U5+ -- 18th plus year

The Diploma course can range from one to three years. Acquiring a Diploma degree in Tanzania can qualify an individual to be a primary school teacher. However, participating in the Diploma course does not technically add additional years of education to an individual's record. This is

because universities treat A+ certification and Diploma's equally for admission. Therefore, to calculate the number of educational years an individual, who attended the Diploma course, has is to add one year to their current university level (U1 through U5+). For example, a student with a Diploma who is in U2 would have 16 years of schooling.

## Appendix G: Description of TASCO

In Section E of the Household Questionnaire (HH_SEC_E.dta), the TASCO codes are used for questions 16, 31, and 45. Respondents were asked to describe what kind of job/work they did. Based off the respondent's description, the TASCO codes were assigned. Depending on the specificity of the job/work description affects if there is a two or three digit TASCO code. Respondents were asked to be specific as possible but in some cases, their responses did not allow for a three digit TASCO code to be assigned. The following list is all of the potential TASCO codes and those used within the survey.

**MAJOR GROUP 1: LEGISLATORS, ADMINISTRATORS AND MANAGERS**

11 Legislators and Administrators
- 111 Legislators
- 112 Senior Government Executive
- 113 Village Leaders
- 114 Senior Administrators of Special-Interest Organizations

12 Company Directors and Corporate Managers
- 121 Company Directors and Non-Government Chief Executives
- 122 Specialised Managers and Senior Administrators
- 123 Production and Operations Managers and Senior Administrators

13 Small Business Managers and Managing Supervisors
- 131 Small Business Managers and Managing Supervisors

**MAJOR GROUP 2: PROFESSIONALS**

21 Physical, Mathematical, and Engineering Science Professionals
- 211 Physical Scientists and Related Professionals
- 212 Mathematicians, Statisticians, and Related Professionals
- 213 Computing Professionals
- 214 Architects, Engineers, and Related Professionals

22 Life Science and Health Professionals
- 221 Life Science Professionals
- 222 Medical and Health Professionals (Except Nurses)
- 223 Nursing Professionals

23 Teaching Professionals
- 231 College, University, and Higher Education Teaching Professionals
- 232 Secondary Education Teaching Professionals
- 239 Other Teaching and Related Professionals

24 Other Professionals
- 241 Business and Administrative Professionals
- 242 Legal Professionals
- 243 Archivists, Librarians, and Related Information Professionals
- 244 Social and Related Science Professionals
- 245 Artistic Professionals
- 246 Religious Professionals

**MAJOR GROUP 3: TECHNICIANS AND RELATED PROFFESIONALS**

31 Physical, Mathematical, and Engineering Science Associate Professionals
- 311 Physical, Science, and Engineering Technicians
- 312 Computer Assistants and Equipment Controllers
- 313 Optical and Electronic Equipment Controllers
- 314 Ship and Aircraft Controllers and Technicians
- 315 Building, Safety, Health, and Quality Inspectors

32 Life Science and Health Associate Professionals
- 321 Life Science Technicians and Related Workers
- 322 Modern Medicine and Health Associate Professionals (Except Nurses)
- 323 Nursing and Midwifery Associate Professionals
- 324 Traditional Medicine Practitioners and Faith Healers

33 Teaching Associate Professionals
- 331 Secondary Education Teachers, Associate Professionals
- 332 Technical/Vocational Education Teachers
- 333 Primary Education Teachers
- 334 Pre-Primary Education Teachers
- 335 Special Education Teachers, Associate Professionals
- 339 Other Teaching Associate Professionals

34 Other Associate Professionals
- 341 Finance and Sales Associate Professionals
- 342 Trade Brokers and Business Services Agents
- 343 Administrative Associate Professionals
- 344 Government Associate Professionals
- 345 Social Work Associate Professionals
- 346 Creative and Performing Art, and Artistic Entertainment, and Sports Associate Professionals
- 347 Religious Associate Professionals
- 348 Other Associate Professionals

**MAJOR GROUP 4: CLERKS**

41 Office Clerks
- 411 Secretaries, Keyboard Operators, and Registry Assistants
- 412 Numerical Clerks
- 413 Material Recording and Transport Clerks
- 414 Library, Mail, and Related Clerks
- 415 Other Office Clerks

42 Customer Service Clerks
- 421 Cashiers, Tellers, and Related Clerks
- 422 Client Information Clerks and Telephone Operators

**MAJOR GROUP 5: SERVICE WORKERS AND SHOP SALES WORKERS**

51 Personal Service Workers
- 511 Travel Attendants and Guides
- 512 Housekeeping and Restaurant Services Workers, Institutional
- 513 Housekeeping and Restaurant Services Workers, Domestic
- 514 Personal Care Workers

515 Astrologers, Fortune-Tellers, and Related Workers
516 Other Personal Service Workers
52 Protective Service Workers
520 Protective Service Workers
53 Salespersons, Demonstrators, and Models
531 Salespersons and Demonstrators
532 Stall and Market Salespersons
533 Fashion and Other Models

**MAJOR GROUP 6: SKILLED AGRICULTURAL AND FISHERY WORKERS**

61 Skilled Agricultural and Fishery Workers
611 Farmers and Crop Skilled Workers
612 Animal Producers and Skilled Workers
613 Forestry and Related Skilled Workers
614 Fishery Workers, Hunters, and Trappers
62 Subsistence, Agricultural, Forestry, Fishery, and Related Workers
621 Subsistence Agricultural, Forestry, Fishery, and Related Workers

**MAJOR GROUP 7: CRAFT AND RELATED WORKERS**

71 Extraction and Building Trades Workers
711 Miners and Blasters Stone Cutters and Carvers
712 Building Frame and Related Trades Workers
713 Building Finishers and Related Trades Workers
714 Painters, Structural Cleaners, and Related Workers
72 Metal and Machinery Trades Workers
721 Metal Moulders, Welders, Sheet-Metal Workers, Structural Metal Preparers, and Related Workers
722 Blacksmiths, Toolmakers, and Related Workers
723 Machinery Mechanics and Fitters
724 Electrical and Electronic Equipment Fitters, Installers, and Repairers
73 Precision, Handicraft, Printing, and Related Trades Workers
731 Precision Workers in Metal, Diamonds, Plastics, Rubber, Paper, and Other Related Materials
732 Potters, Glass Formers, and Related Workers
734 Handicraft Workers in Wood, Textile, Leather, and Related Materials
735 Printing and Related Trades Workers
74 Other Crafts and Related Trades Workers
741 Food and Related Products Procession Trades Workers
742 Cabinet Makers, Wood Treaters, and Related Trades Workers
743 Textile and Garment Trades Workers
744Pelt, Leather, and Shoemaking Trades Workers
749 Other Craft and Related Trades Workers, NEC

**MAJOR GROUP 8: PLANT AND MACHINE OPERATORS AND ASSEMBLERS**

81 Industrial Plant Operators
811 Mining and Mineral-Processing Plant Operators

812 Metal-Processing Plant Operators
813 Glass and Ceramics Kiln and Related Plant Operators
814 Wood-Processing and Papermaking Plant Operators
815 Chemical-Processing Plant Operators
816 Power-Generating and Related Plant Operators
82 Stationary Machine Operators and Assemblers
821 Metal and Mineral Products Processing Machine Operators
822 Chemical Products Machine Operators
823 Rubber, Plastics, and Leather Products Machine Operators
824 Wood Products Machine Operators
825 Printing, Binding, and Paper Products Machine Operators
826 Textile Products Machine Operators
827 Food and Related Products Processing Machine Operators
828 Assemblers
829 Other Stationary Machine Operators and Assemblers
83 Drivers and Mobile Machinery Operators
831 Railway Engine Drivers and Related Workers
832 Motor Vehicle Drivers and Riders
833 Agricultural, Earthmoving, Lifting, and Other Mobile Material-Handling Equipment Operators
834 Ships Deck Crew and Related Workers

**MAJOR GROUP 9: ELEMENTARY OCCUPATIONS**

91 Sales and Services Elementary Occupations
911 Street Vendors and Related Workers
912 Shoe Cleaning and Other Street Services Elementary Occupations
913 Domestic Helpers and Cleaners and Related Workers
914 Building Caretakers and Window Cleaners
915 Messengers, Watchers, and Related Workers
916 Garbage Collectors and Related Labourers
919 Other Sales and Services Elementary Occupations
92 Agricultural, Forestry, Fishery, and Related Labourers
921 Agricultural, Forestry, and Fishery Labourers
93 Labours in Mining, Construction, Manufacturing, and Transport
931 Mining and Construction Labourer
932 Manufacturing Labourers
933 Transport Labourers
934 Hand Packers, Weighers, and Related Elementary Workers

**MAJOR GROUP 0: DEFENCE FORCES**

01 Defense Forces
011 Tanzania People's Defense Forces
012 National Service (JKT)

**MAJOR GROUP X: WORKERS NOT CLASSIFIED BY OCCUPATIONS**

XI New Workers Seeking Employment

X11 Fresh Graduates and Under-Graduates Seeking Employment
X12 Fresh Secondary School Leavers and Dropouts Seeking Employment
X13 Fresh Primary School Leavers and Dropouts Seeking Employment
X14 Fresh Literates Seeking Employment

X2 Workers Reporting Occupations Unidentifiable or Inadequately Described
X21 Workers Reporting Occupations Unidentifiable or Inadequately Described

X3 Workers Not Reporting Any Occupation
X31 Workers Not Reporting Any Occupation

## Appendix H: List of ISIC codes

In Section E of the Household Questionnaire (HH_SEC_E.dta), the ISIC codes are used for questions 17, 32, 46, and 53. Respondents were asked to describe their association to different trades and businesses. Based off the respondent's description, the ISIC codes were assigned. Depending on the specificity of trade/business description affects if there is a two, three, or four digit ISIC code. Respondents were asked to be specific as possible but in some cases, their responses did not allow for a three or four digit ISIC code to be assigned. The following list is all of the potential ISIC codes and those used within the survey.

A - *Agriculture, forestry and fishing*
- 01 - Crop and animal production, hunting and related service activities
- 02 - Forestry and logging
- 03 - Fishing and aquaculture

B - *Mining and quarrying*
- 05 - Mining of coal and lignite
- 06 - Extraction of crude petroleum and natural gas
- 07 - Mining of metal ores
- 08 - Other mining and quarrying
- 09 - Mining support service activities

C - *Manufacturing*
- 10 - Manufacture of food products
  - 101 - Processing and preserving of meat
  - 102 - Processing and preserving of fish, crustaceans and mollusks
  - 103 - Processing and preserving of fruit and vegetables
  - 104 - Manufacture of vegetable and animal oils and fats
  - 105 - Manufacture of dairy products
  - 106 - Manufacture of grain mill products, starches and starch products
  - 107 - Manufacture of other food products
  - 108 - Manufacture of prepared animal feeds
- 11 - Manufacture of beverages
- 12 - Manufacture of tobacco products
- 13 - Manufacture of textiles
- 14 - Manufacture of wearing apparel
- 15 - Manufacture of leather and related products
- 16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
- 17 - Manufacture of paper and paper products
- 18 - Printing and reproduction of recorded media
- 19 - Manufacture of coke and refined petroleum products
- 20 - Manufacture of chemicals and chemical products
- 21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations
- 22 - Manufacture of rubber and plastics products
- 23 - Manufacture of other non-metallic mineral products
- 24 - Manufacture of basic metals

25 - Manufacture of fabricated metal products, except machinery and equipment
26 - Manufacture of computer, electronic and optical products
27 - Manufacture of electrical equipment
28 - Manufacture of machinery and equipment n.e.c.
29 - Manufacture of motor vehicles, trailers and semi-trailers
30 - Manufacture of other transport equipment
31 - Manufacture of furniture
32 - Other manufacturing
33 - Repair and installation of machinery and equipment

D - *Electricity, gas, steam and air conditioning supply*

35 - Electricity, gas, steam and air conditioning supply

E - *Water supply; sewerage, waste management and remediation activities*

36 - Water collection, treatment and supply
37 – Sewerage
38 - Waste collection, treatment and disposal activities; materials recovery
39 - Remediation activities and other waste management services

F - *Construction*

41 - Construction of buildings
42 - Civil engineering
43 - Specialized construction activities

G - *Wholesale and retail trade; repair of motor vehicles and motorcycles*

45 - Wholesale and retail trade and repair of motor vehicles and motorcycles
46 - Wholesale trade, except of motor vehicles and motorcycles
47 - Retail trade, except of motor vehicles and motorcycles
471 - Retail sale in non-specialized stores
472 - Retail sale of food, beverages and tobacco in specialized stores
473 - Retail sale of automotive fuel in specialized stores
474 - Retail sale of information and communications equipment in specialized stores
475 - Retail sale of other household equipment in specialized stores
476 - Retail sale of cultural and recreation goods in specialized stores
477 - Retail sale of other goods in specialized stores
478 - Retail sale via stalls and markets
479 - Retail trade not in stores, stalls or markets

H - *Transportation and storage*

49 - Land transport and transport via pipelines
491 - Transport via railways
492 - Other land transport
4921 - Urban and suburban passenger land transport
4922 - Other passenger land transport
4923 - Freight transport by road
493 - Transport via pipeline
50 - Water transport
51 - Air transport
52 - Warehousing and support activities for transportation
53 - Postal and courier activities

I - *Accommodation and food service activities*

55 – Accommodation
56 - Food and beverage service activities
561 - Restaurants and mobile food service activities
562 - Event catering and other food service activities
563 - Beverage serving activities

J - *Information and communication*
58 - Publishing activities
59 - Motion picture, video and television programme production, sound recording and music publishing activities
60 - Programming and broadcasting activities
61 – Telecommunications
62 - Computer programming, consultancy and related activities
63 - Information service activities

K - *Financial and insurance activities*
64 - Financial service activities, except insurance and pension funding
65 - Insurance, reinsurance and pension funding, except compulsory social security
66 - Activities auxiliary to financial service and insurance activities

L - *Real estate activities*
68 - Real estate activities

M - *Professional, scientific and technical activities*
69 - Legal and accounting activities
70 - Activities of head offices; management consultancy activities
71 - Architectural and engineering activities; technical testing and analysis
72 - Scientific research and development
73 - Advertising and market research
74 - Other professional, scientific and technical activities
75 - Veterinary activities

N - *Administrative and support service activities*
77 - Rental and leasing activities
78 - Employment activities
79 - Travel agency, tour operator, reservation service and related activities
80 - Security and investigation activities
81 - Services to buildings and landscape activities
82 - Office administrative, office support and other business support activities

O - *Public administration and defense; compulsory social security*
84 - Public administration and defense; compulsory social security

P – *Education*
85 - Education

Q - *Human health and social work activities*
86 - Human health activities
87 - Residential care activities
88 - Social work activities without accommodation

R - *Arts, entertainment and recreation*
90 - Creative, arts and entertainment activities
91 - Libraries, archives, museums and other cultural activities
92 - Gambling and betting activities

93 - Sports activities and amusement and recreation activities

S - *Other service activities*

94 - Activities of membership organizations

95 - Repair of computers and personal and household goods

96 - Other personal service activities

T - *Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use*

97 - Activities of households as employers of domestic personnel

98 - Undifferentiated goods- and services-producing activities of private households for own use

U - *Activities of extraterritorial organizations and bodies*

99 - Activities of extraterritorial organizations and bodies