TECHNICAL NOTE

Addis Ababa and Dire Dawa, Ethiopia

Disaster Poverty Household Survey

WORLD BANK GROUP

GFDRR
Global Facility for Disaster Reduction and Recovery

1 Overview

Content of this document: This document provides information about the Disaster Poverty Household Survey (DPHS). It describes the DPHS series, the survey and sampling design and the questionnaire used, and it discusses some data considerations, including outlier treatment and anonymization process.

Objective of the survey: The DPHS is designed to collect information that can be used to assess the relationship between disasters (exposure, vulnerability, and capacity to recover) and poverty in the urban environment. The data can be used to explore policy-relevant research topics related to climate change adaptation, urbanization, urban poverty, and more.

Content of the database: DPHS data contains information on household characteristics, household expenditure, living conditions and household experience with disasters. Household characteristics include household size and member level information on religion, education and labor. Household expenditure is collected using the Survey of Well-being via Instant and Frequent Tracking (SWIFT) methodology, which estimates household income (or consumption expenditure) based on non-monetary variables that are highly correlated with poverty. Information on living conditions covers housing quality, asset ownership, access to services and jobs, rent and housing costs and tenure arrangements. Information on experiences with disasters includes direct and indirect impacts of historic disasters on household assets, education, health and labor access, as well as impacts on public services. There is also information on coping behaviors and perception of risk of future exposure. The DPHS can be customized to collect information on different disasters. So far, it has mainly focused on the impacts of urban flooding.

Addis Ababa and Dire Dawa application: The DPHS in Addis Ababa and Dire Dawa was conducted in May and June 2017, with the objective to assess the role of poverty in disaster risk, focusing primarily on urban flooding but also other hazards.

This project was a collaborative effort between Global Facility for Disaster Reduction and Recovery (GFDRR), the Poverty Global Practice and Urban, Disaster Risk Management, Resilience and Land Global Practice (GPURL). Data collection was carried out by UDA Consulting under the supervision of the World Bank.

Data files and other resources

- DPHS_AddisAbabaDireDawaEthiopia_Data_2017: DPHS data in STATA format (.dta)

- DPHS_AddisAbabaDireDawaEthiopia_SWIFT: SWIFT (household expenditure) data in STATA format
- DPHS_AddisAbabaDireDawaEthiopia_Questionnaire: DPHS Questionnaire in excel

Citation requirements:

- The World Bank. Disaster Poverty Household Survey (DPHS), Addis Ababa and Dire Dawa, Ethiopia 2017. Ref: *DPHS_AddisAbabaDireDawaEthiopia_Data_2017*. Dataset downloaded from microdata.worldbank.org on [date].

2 The survey

2.1 Description

Name of the study: Disaster Poverty Household Survey, Addis Adaba and Dire Dawa, Ethiopia

Geographical coverage: Addis Adaba and Dire Dawa

Sample size: 1197 (810 in Addis Adaba and 387 in Dire Dawa)

Date of the survey: May and June 2017

Primary Investigators: Alvina Erman (World Bank), Silvia Malgioglio (World Bank), Nobuo Yoshida (World Bank), Stephane Hallegatte (World Bank)

Collaborators: GFDRR, Poverty Global Practice, GPURL and UDA consulting

Funding: GFDRR

2.2 Sampling design

Satellite images of Addis Adaba and Dire Dawa were used to divide both cities into 100mx100m grids and among those, 173 and 81 grids in Addis Ababa and Dire Dawa respectively were randomly selected. In each selected grid, a 10 x 10 meters secondary dot grids were created. Then, in each secondary grid, 5 households were randomly assessed for inclusion. If the house corresponded to the characteristics of a residential and “low-income/slum” dwelling, it was included in the sample. While the sampling was carried out in a manner to assure representativeness at the city level, caution should be taken before generalizing results generating from this data for the entire city population. This is because the sample intended to sample slum dwellers and low-income households (based on factors

that are detectable in high-resolution satellite imagery and visible from above, such as quality of roofing and dwelling size, size of plot, etc.).

3 Questionnaire modules

- K: Household information: pre-filled before the interview
- B: Household member roster
  - Educational attainment
  - Labor participation
  - Health
- C: Household enterprises
- M: Asset ownership
- H: Housing and services
  - Tenure arrangements
  - Land rights
  - Credit for land
  - Housing costs and rent
  - Tenure security
  - Housing quality
  - Access to services
  - Agricultural practices
  - Social protection (UPSNP)
  - Source of income
  - Remittances
- I: Investments in housing
- E: Shock roster (flood, landslide, fire)
- F: Experience with shocks
- P: Idiosyncratic shocks
- W: Preventive measures flood
- U: Perception of future risk
- L: Food Insecurity (Reduced Coping Strategy Index CSI-R)
- D: Depression

4 Data considerations

4.1 Anonymization of the dataset

Protecting the privacy of survey respondents is of the outmost importance to the World Bank. To make sure the data cannot be used to identify individual households in the dataset, a technique of statistical disclosure control (SDC), as described in Benschop et al. (2021), was applied. It helped identify variables that included unique information about households. After identifying the high-risk variables, necessary adjustments were made to make sure the SDC analysis provided satisfactory results, i.e., low risk of re-identification. Results can be shared upon request. The following data editing was done for anonymization purpose:

- Precise location data, such as GPS coordinates, and 10 x 10 meters grids were dropped
- Personal information, such as names and phone numbers were dropped
- The number of religions reported was reduced from 6 to 3 categories, the number of ethnicities from 14 to 4 categories, marital status from 6 to 4 categories
- Household size exceeding seven household members was categorized as “above 7 members”
- Household member information for 7th member and above was dropped to avoid reconstruction of the household size variable.

4.2 Outlier treatment

Continuous variables may present some measurement errors. A technique of outliers' treatment is recommended. Some of these variables are:

- *b11d_commuting_time*: How long does it take to get to the workplace? (minutes)
- *h34*: Over *the* last 3 months, how much did you send cash or gifts to family (In *Birr*)
- h6_c: How much do you pay in rent each time? (In *Birr*)
- *h8*: If somebody else wants to rent a dwelling just like this today, how much money would he/she has to pay monthly? (In *Birr*)

An established method to identify outliers is to tag the observations that deviate from the mean by a set number of standard deviations. Three standard deviations are commonly used.

Figure 1 includes STATA code that can be used to tag outliers[1].

Figure 1: Codes for the identification of outliers for the variable *h6_c (rent)*

```
* Create dummy for outliers
foreach var of varlist h6_c {
    quietly summarize `var'
    g Z_`var'=(`var' > 3*r(sd)) if `var'<.
    list `var' Z_`var' if Z_`var'==1

}
/*
      +-----------------+
      |   h6_c   Z_h6_c |
      |-----------------|
 681. |  99999        1 |
1072. | 100000        1 |
      +-----------------+
*/
```

These outliers can then be replaced in function of the variable studied. For instance, for the variable h6_c (amount of rent), a hedonic regression can be applied to identify drivers of rent, which can be used to produce predicted rent values that can then replace the outliers identified.

5 SWIFT Methodology

Household consumption data are costly to collect and significantly increases the duration of interviews. Beyond budgetary and data processing issues, it also reduces the quality of the data by reducing the space available for the other questions in the survey and by increasing the risk of survey fatigue of respondents. To avoid these issues, the survey adopted the SWIFT approach to estimate household expenditures and poverty rates. The SWIFT methodology collects household data using a short list of questions that can be integrated into the questionnaire and computes an estimated household income (or consumption expenditure) based on non-monetary variables that are highly correlated with poverty. SWIFT uses survey-to-survey imputation based on official household data and produces estimates comparable to official data. More details on SWIFT are provided in Yoshida et al. (2021).

The data on expenditure are in *DPHS_ AddisAbabaDireDawaEthiopia_SWIFT* dataset, which

[1] Additional checks may be conducted to analyze the presence of outliers. The technique in Figure 1 assumes that the distribution of the variable is normal. This may not be the case, even after using a logarithmic transformation. Other transformations and for which kinds of variables to use them are explained in Ravallion (2017). Outliers may influence the mean and the median of the distribution. More robust methods of outlier treatment may be necessary, for instance, the median absolute deviations (MAD) method (Belotti et al., 2021, Rousseeuwand Croux, 1993).

can be merged with *DPHS_AddisAbabaDireDawaEthiopia_Data_2017* dataset using the key variable *hhid*.

References

BELOTTI, F., G., VECCHI, G., AND MANCINI (2021): "OUTDETECT: Stata module to perform outlier detection and diagnostics for welfare analysts," *Statistical Software Components* S458932, Boston College Department of Economics.

BENSCHOP, T. AND WELCH, M. (n.d.): "Statistical Disclosure Control for Microdata: A Practice Guide", Retrieved 30 March, 2022, from https://sdcpractice.readthedocs.io/en/latest/

RAVALLION, M. (2017): "A concave log-like transformation allowing non-positive values," *Economics Letters*, 161, 130-132.

ROUSSEEUW, P. J., CROUX, C. (1993): "Alternatives to the median absolute deviation," *Journal of the American Statistical association*, 88(424), 1273-1283.

YOSHIDA, N., X. CHEN, S. TAKAMATSU, K. YOSHIMURA, S. MALGIOGLIO, S. SHIVAKUMARAN, K. ZHANG, D. ARON. (2021): "The Concept and Empirical Evidence of SWIFT Methodology," The World Bank