

Monitoring COVID-19 impacts on firms in Ethiopia

SAMPLING DESIGN

SAMPLING DOCUMENTATION

15 / May / 2020



Authors: Girum Abebe, Tom Bundervoet, Christina Wieser

OVERVIEW



This document describes the sampling design of the High Frequency Phone Survey of Firms (HFPS-F), implemented as a response to COVID-19. The survey aims at assessing the dynamics of the impacts of COVID-19 on establishments in Ethiopia. The COVID-19 pandemic and its effects create an urgent need for timely data and evidence to help monitor and mitigate the impact of the crisis. Due to limits on face-to-face surveys the HFPM-F is undertaken via the phone.

The HFPM-F will monitor the economic activities and responses to the COVID-19 crisis, particularly its effects on firm operations, revenues and jobs, by calling a sample of establishments every three weeks between mid-April and mid-September 2020 for a total of eight survey rounds. The final dataset will consist of a panel of 800 establishments (500 in Addis Ababa and 300 in four other cities; in the other cities, there will be seven instead of eight rounds). To account for non-response and attrition, the team aims at successfully interviewing 1,050 establishments (650 in Addis Ababa and 400 in other cities) in round 1.

SAMPLING METHODOLOGY



The sampling procedure was undertaken in three steps. First, the team cleaned the list of registered establishments in Ethiopia, received from the Ministry of Trade and Industry (MoTI), by removing establishments with missing or invalid phone numbers. Second, all phone numbers of the cleaned list of establishments were shared with Ethio Telecom and only active phone numbers were kept constituting the sampling frame. Third, with this sampling frame, the team drew a random sample of establishments, stratified by establishment size (proxied by capital) and sector.

STEP 1: CLEANING SAMPLING FRAME



To get to a sampling frame, the team relied on a list of registered establishments received from MoTI. The sampling frame included 403,039 establishments in Addis Ababa and 142,170 establishments in Mekelle, Adama, Bahir Dar, and Hawassa. Given that the survey is administered over the phone, we had to weed out establishments without a phone number or with an invalid one. Therefore, the list of establishments was cleaned by removing establishments with a missing or invalid phone numbers. Invalid phone numbers were deemed as those with (i) less than or more than 10 digits; (ii) an invalid telephone code; and (iii) invalid characters in the number string. The cleaned list of contained 389,927 establishments in Addis Ababa and 138,679 establishments in the other cities.

STEP 2: CHECKING FOR ACTIVE PHONE NUMBERS



The cleaned list of establishments was shared with Ethio Telecom, the only telecom provider in Ethiopia to validate the phone numbers. Ethio Telecom checked on the status of all numbers and only numbers deemed as "active" by Ethio Telecom were retained and constitute the sampling frame. The verification with Ethio Telecom was undertaken to avoid a scenario in which establishments were in the sample but could not be reached due to an invalid phone number. The sampling frame consists of 288,660 establishments in Addis Ababa and 118,523 establishments in the other cities.

STEP 3: DRAWING THE SAMPLE



The sampling frame consists of all establishments registered with MoTI with a valid phone number in Addis Ababa, Mekelle, Adama, Bahir Dar, and Hawassa. We selected two survey domains (Addis Ababa and other cities) which were considered as explicit sampling strata. The sample size was set by a tight budget envelope and the team decided on a panel of 500 establishments in Addis Ababa and 300 establishments in other city after round 8 of the survey operation.

The sample is stratified by establishment size (proxied by its capital) and industry classification, both variables were available in the sampling frame. We considered two industry classifications (industry and services) and three firm size groupings: micro (below the 25th percentile in terms of capital), small and medium (25th to 75th percentile of capital), and large establishments (above the 75th percentile of capital) to be included in the stratification process. The use of capital to proxy firm size is highly imperfect, as capital is only weakly related to employment size in Ethiopia. The sample consists of six strata:

1. micro establishments in industry
2. micro establishments in services
3. small and medium establishments in industry
4. small and medium establishments in services
5. large establishments in industry
6. large establishments in services

In Addis Ababa, we oversampled industry as the percentage of firms operating in the industry sector was too small and there was concern about representativeness if attrition is high. In the other cities outside of Addis Ababa, we only focused on the manufacturing and tourism sector as these are the sectors of most interest to the program of the World Bank in these cities. There is, however, only one sector in services (tourism) and a proportional sample would have led to few establishments in services (tourism). To avoid this, we oversampled for services to ensure that 40 firms out of the 200 (20 percent) operate in the service sector in each city.

The sample was drawn using a simple random sample without replacement. Expecting a high non-response rate, we drew a sample of 1,450 establishments for Addis Ababa and 800 (200 in each city) establishments in other cities.

SAMPLING WEIGHTS



To obtain unbiased estimates from the sample, the information reported by the establishment needs to be affected by a sampling weight (or raising factor) w_h . Given that very little information is provided in the list of registered establishments—we have 3 characteristics (capital, sector, location) available—we decided to undertake a weighting class adjustment. We thus constructed cells (region by sector by size) to get a correct count of establishments in each class and divided the number of establishments in the survey by the number of establishments in the sampling frame for each class. We then applied the weights to the sample to ensure the distribution in the survey matches the distribution in the sampling frame.

REPLACEMENT OF ESTABLISH- MENTS



In each domain (Addis Ababa and other cities), additional establishments are sampled to serve as replacement establishments in case of non-response and attrition. As part of the data collection phase, each establishment is called at least three times over three consecutive days (for a total of nine attempts) before replacing an establishment from the list of replacements.