

Notes on the Togo Informal Firms Data and Replication Package

Context:

This data was collected in the context of an impact evaluation with a sample of 1,500 informal firms in Lomé, Togo. 500 entrepreneurs in the sample were invited to a Business Edge training, an internationally accredited managerial training program developed by the International Finance Corporation (IFC). An additional 500 entrepreneurs in the sample were invited to an innovative training on personal initiative developed by Prof. Frese from the Leuphana University of Lüneburg and his team. The remaining 500 entrepreneurs served as a non-treatment control group for the impact evaluation.

Sampling:

To be eligible for the program, businesses had to be informal (they could not be registered at the Chamber of Commerce or the Center of Enterprise Formalities at the time of their application), have less than 50 employees, exist for at least 12 months, and be in any sector other than agriculture (as there was a different project targeting the training of agricultural firms). All of the targeted businesses were in the greater Lomé area. To apply, companies needed to fill out a form and to provide proof that they had existed for at least 12 months, such as receipts, memberships to professional associations or trade unions, business bank or micro finance accounts etc. Applications were accepted either online or in paper format at the project's partner institutions, which included micro finance institutions, artisan and women's associations, the Regional Chamber of Artisans, and a government institution working with the informal sector.

In order to mobilize eligible companies to apply to the program, the project launched a four-month communication campaign using traditional means such as radio and television but also working with local partners and going door to door.

The selection process was conducted in two steps. In the first step, the list of entrepreneurs to survey in the baseline survey was randomly selected from the database of eligible applicants. Approximately 1,800 entrepreneurs were selected with the goal of surveying 1,500. As such, entrepreneurs who were no longer interested or who could not be located could be replaced with others from our sampling frame. After conducting the baseline survey, we used the more precise data to divide the sample of 1500 interviewed companies into three groups: 500 in the managerial training, 500 in the entrepreneurial training, and 500 in the control group.

During the first step of the selection process, we used the application data to generate strata based on average monthly sales and sector of activity. The sales variable in the application form included the following brackets of average monthly sales (in FCFA):

- 0 – 50,000
- 50,001 – 125,000
- 125,001 – 250,000
- 250,001 – 375,000
- 375,001 – 500,000
- 500,001 – 625,000
- 625,001 – 750,000
- 750,001 – 875,000
- 875,001 – 1,000,000
- 1,000,001 – 1,125,000
- 1,125,001 – 2,250,000
- Greater than 2,250,000

The sector of activity variable was grouped into 1) Production 2) Construction 3) Commerce 4) Other services. There were 47 distinct strata. We then randomly selected the same number of companies in each strata. In strata where there were fewer companies than the equal share of companies to select per strata, the remaining spots were equally distributed among the other strata. The idea was to ensure a diverse population of companies and to maintain as much representativeness of types of companies as possible. Furthermore, the project team was interested in supporting a higher number of high potential entrepreneurs in production and with higher average sales. This method, while giving equal chance to entrepreneurs with the same characteristics, weighted the sample in favor of those with higher sales and those in smaller sectors. As aforementioned, this selection process provided a list of 1,794 eligible companies to provide to the survey firm for the baseline survey.

To assign the treatment groups, we separated the sample into strata based on gender, sector and whether they had a business practice score above or below the median score. The sector variable consisted of three sectors: 1) Production 2) Commerce 3) Other services. The business practice score variable, however, unfortunately had a coding error. Instead of randomly assigning those with the median score to be either above or below, it divided the sample into two parts using the random variable. Finally, within these strata we formed triplets based on profits and randomly allocated treatment within the triplet.

Questionnaires:

The questionnaires were drafted in French and translated into Ewe and Kabiyé using an approach of translation, backtranslation and validation of the translation. English versions of the questionnaires are also available for the convenience of researchers; however, they were translated after the survey and did not follow the same rigorous validation procedure. In case of any translation discrepancy, the original wording in French should prevail.

The codes for the sector of activity are in the document entitled Supporting Materials.

Raw data:

Questions that are sensitive for reasons of confidentiality have been removed from the dataset. For example, all of the open-ended questions have been removed. The coded answers that were derived from these questions are included, however. The guide to coded questions explains how the open-ended answers were converted into the coded variables included in the dataset.

All financial information is in Francs CFA (XOF) and has not been deflated. The CPI used to deflate the variables for our analysis has been included in the dataset. As specified in our registered pre-analysis plan, to deflate our variables we used the consumer price index published by the National Institute for Statistics and Economic and Demographic Studies in Togo, which is a public establishment attached to the Ministry in charge of statistics. The name of the institution in French is Institut National de la Statistique et des Etudes Economiques et Démographiques (INSEED-TOGO).

All datasets were created in STATA 14 but have been saved in STATA 13 compatibility format for your convenience.

Do files:

To replicate the analysis, you must first do the following:

- Create a folder for replication that has the following subfolders. Please be sure to name the subfolders in exactly this way, or the code will not properly replicate.
 - Data
 - Do files
 - Results files
 - Log files
- Copy the raw datasets into the Data folder. This includes:
 - "BL_raw_an_old.dta"
 - "FU1_raw_an_old.dta"
 - "FU2_raw_an_old.dta"
 - "FU3_raw_an_old.dta"
 - "FU4_raw_an_old.dta"
- Copy all do-files into the Do files folder. This includes:
 - "Togo IE Master Replication_final.do"
 - "Togo IE Data Preparation Replication_final.do"
 - "Togo IE Data Analysis Replication_final.do"
- Open the do-file "Togo IE Master Replication_final.do". Add your file path for the replication folder in line 47 of this do-file (replacing the current text "your filepath to the replication folder here").
- Run the Master do-file. The results will be exported into the Results files subfolder you created.

All analysis was done in STATA, with the exception of the Monte Carlo simulations for the mediation analysis, which was done in R.

****Note:** The command `estadd` cannot be install using `ssc install`. It is used in the replication do-files. In order for everything to run properly, you must first manually install the command `estadd` by searching for it using `"help estadd"`.