USER GUIDE
Inside the Production Function - The Effect of Financial Contracts on Growing Firms' Technology Use. Evidence from a field experiment in Uganda [*]

Francesco Loiacono, Stockholm University[†]
Selim Gulesci, Bocconi University
Andreas Madestam, Stockholm University
Miri Stryjan, Ben-Gurion University of the Negev[‡]

2017
December

[†]For questions about the data sets, contact francesco.loiacono@ne.su.se.
[‡]For questions about the user manual, contact miristry@bgu.ac.il.

Contents

1 Background

The data files contain data from interviews with Ugandan small business owners that are part of a randomized experiment evaluating the effect of loan contract structure on business performance. The aim of the project is to examine how alterations of the standard microfinance contract structure affect borrowing business owners' use of the loan and their business performance, measured in a variety of ways.

The experiment was carried out in collaboration with the NGO BRAC Uganda. BRAC is a large non-profit organization founded in Bangladesh in 1974, currently active in 12 developing countries in Asia, Sub-Saharan Africa and the Caribbean. BRAC was launched in Uganda in 2006 and is currently one of the largest development organizations and microfinance institutions in the country. Its core activity is microfinance which encompasses both microcredit groups and the Small Enterprise Lending (SEP) program that targets existing business owners.

Beginning in 2014, we collaborated with the BRAC SEP program in 76 local branch offices in Central, Western, and Eastern Uganda. Every firm in the sample is a BRAC borrower in the catchment area of one of these branch offices, belonging to one of the business sectors we had pre-selected, that has been approved by BRAC credit officers and is eligible for a loan. Upon being eligible, firms are asked to participate in a Baseline survey. Once the survey is conducted with firms that gave their consent to participate, they are randomized into one of the treatment (and control) arms. Data was collected from clients in several waves: before assignment to a treatment or control group (Baseline), during the loan cycle (Diaries), and after completion of the loan (end-line). Each of these types of data are described in more detail in the section on Survey design.

Data was collected electronically using tablets and stored on the survey platform SurveyCTO. Surveys were programmed using the application SurveyCTO.

2 Experiment Design

2.1 Overview

Every firm in the sample is a BRAC borrower that has been approved by BRAC credit officers and is eligible for a loan. The loans offered through BRAC Uganda's SEP program are individual liability loans offered to business owners. The minimum loan size is 2.5 Million UGX, or

approximately 900 USD, and maximum loan sizes follow a ladder where first time borrowers maximum loan size is 8 Million UGX (2,900 USD) while repeat borrowers are allowed larger loans. The median loan size at the start of our experiment was 3 Million UGX.* Prospective borrowers are evaluated by BRAC credit officers at the local office closest to the borrower's business location. A borrower needs to be the sole business owner, registered with the tax authorities, a permanent resident of the branch office area (s)he is applying in, and is not allowed to have outstanding loans with BRAC or other lenders. In addition, a borrower needs to provide collateral (land) amounting to the value of the loan and two guarantors who will be responsible for the repayments in case the borrower fails to repay the loan. A loan cycle lasts for 1 year during which the loan is paid off in 12 equal installments at an annual interest rate of 25 percent.

Upon being eligible, and after completing the Baseline survey, firms were then randomized into one of the treatment (and control) arms. Randomization was implemented at the individual client level across 76 branches throughout Uganda. The local research officer was then informed of the treatment and - in case the client was not assigned to the control group - met again with the client to explain the altered contract terms.

2.2 Treatments

The borrowers were randomly allocated into one of the following 6 groups:

T1. "Early repayment voucher": Firms in treatment group 1 were allowed to skip the first 2 repayments (months 1-2 in the 1-year loan cycle);

T2. "Late repayment voucher" firms in treatment group 2 were allowed to skip the last 2 repayments (months 11-12 in the 1-year loan cycle);

T3. "Flexible repayment voucher" firms in treatment group 3 were allowed to skip any 2 repayments of their own choosing (any 2 months in the 1-year loan cycle);

T4. "Flat repayment voucher" firms in treatment group 4 received rebates on all repayments such that the total loan repayment over the 1-year loan cycle is equivalent to the total repayment in each of treatments T1-T3;

---

*Using the nominal exchange rate at the launch of the experiment (November 1, 2014) the minimum SEP loan size corresponded to 912 USD, the median to 1,100 USD and the maximum for first time borrowers corresponded to 2,900 USD. Corresponding real values, using the World Bank PPP adjusted exchange rate for 2014, are 2450 USD, 2,940 USD and 7,840 USD, respectively.

T5. "Subsidy voucher" firms in treatment group 5 received a cash grant equivalent in value to 2 repayments (one sixth of the principal plus the interest payment). This grant was paid to the firms on the same day as their loan disbursement (i.e. at the beginning of their loan cycle);

C. "Control" firms in the control group received the standard BRAC loan contract as described above.

The corresponding variables in the data belong to the `treatment_data.dta` (described in detail below). They include: `arm, flex, early, late, flat, control, sub`.

2.3 Stratification and treatment group size

The randomization was stratified (Bruhn and McKenzie, 2009) by region[†], sector (manufacturing vs. retail), and previous experience with BRAC SEP loans (new vs. repeat borrower). Furthermore, firms that entered into the sample consecutively within a stratum were assigned blocks and the randomization was conducted *within* each block. In this way, we effectively stratify the sample by the order in which firms enter into our sample within an area, sector, and previous loan experience. The sample size (390 firms per treatment group and 390 for the control group) is based on calculations to detect a 0.20 S.D. shift in key outcomes with 80 percent power under a 95 percent confidence level.

2.4 Practical treatment implementation

In practice, the treatments were implemented by giving the clients vouchers that they then used instead of repayment for certain installments. The treatments were designed such that the value of the repayment voucher provided each firm with a subsidy, or rebate equivalent to two monthly installments. In other words, without discounting, the value of the subsidy provided in T1-T5 is one sixth of the total amount (principal plus interest) that the firm owed BRAC.

[†]Each local BRAC office ("branch") is served and supervised by a BRAC Area Office and some of the lending activities of the local offices in a given Area offices are overseen by the same Area office staff. The 76 branches in our study are grouped into 15 such Areas.

3   Data collection procedure

The survey data (presented in more detail below) was collected through interviews with business owners that had taken loans with BRAC SEP program, and were conducted at their business premises. In total 23,764 interviews were made with 2,410 clients.

3.1   Training and supervision of interviewers (enumerators)

The field team consisted of 42 interviewers, hired on three occasions: October 2014, February 2015 and April 2015. From September 2016 to June 2017, as the number of monthly interviews gradually became smaller, the team was gradually reduced accordingly.

The interviewers, also referred to as "enumerators" in the local context, are local university graduates hired especially for this project through the Human Resources Unit of the collaborating organization, BRAC Uganda. Pre-screening questions were designed by the research team with the help of the Human Resources department in order to assess their quantitative and their personal relationship skills, as well as their experience with data collection using tablets or phones. The interviewers were trained to conduct the survey in several training sessions, planned and conducted by one of the coauthors of this project, with the assistance of a Ugandan field supervisor. The initial training session for the Baseline survey and business Diaries lasted 9 days and covered extensively Baseline as well as business Diaries. At the end of this training session, the research team chose the most suitable candidates to undertake the data collection, based on their scores on homework assigned at the end of each training day, their active participation in the classes and their field exam, which consisted in a series of mock interviews with business owners, residing close to the training venue and who agreed in answering to the mock questionnaires. Language skills and knowledge were determinants when deciding the geographical placement of each candidate once hired.

Refresher training sessions were thereafter conducted every month to address the most common mistakes committed by the interviewers as well as the most relevant issues encountered when dealing with the respondents in remote areas. Finally, the training for the End-line survey started in October 2015, in line with the timeline of the project.

The trainings were conducted mainly in English and Luganda (the common local language spoken in Uganda). During the training period, questionnaires were translated from English into the main languages spoken in the areas where the study took place, namely Luganda,

Lusoga, Lutoro, Lunyankole/Lukiga and Lugisu. The questionnaires covered general concepts concerning loans (interest rates, repayment timing, etc.), business practices (definition of profits, employees' contracts and agreements, etc.), and access to credit (sources of funds for micro and small businesses).

3.2 Teams of interviewers in each geographic location

The interviewers were organized in teams of minimum 1 to maximum 3 members. Each team was associated with a local BRAC Area office, following the administrative area division used by the BRAC-SEP program. The areas were the following: Kampala (East, South, North and West); Kasangati; Luwero; Mukono; Jinja; Iganga; Mbale; Masaka; Kabwohe; Mbarara; Fort Portal and Mubende. Fort Portal was later split in 2, forming the additional administrative area of Kasese. Similarly, Iganga was divided in 2 and the additional area of Bugiri was created. Kampala North and Masaka were reshaped and this created the additional administrative area Mpigi. The data collection teams were re-allocated accordingly in order to cover all the areas under the SEP program.

4 Data

There are three types of data:

1. Survey data on the clients, collected through interviews conducted by research assistants hired by the research team, at the business premises of the client's business.

2. Experiment data: data on the treatment status of each client.

3. Administrative data: identifiers, basis for stratification and division into blocks, as well as client loan size.

4.1 Survey structure and survey data

The survey data includes variables collected in multiple waves for each borrowing business ("client"). Interviews were conducted with the person responsible for the managerial decisions at the business, in the large majority of cases this was the business owner. We refer to

the respondents as "clients" since they are borrowing clients of our collaborating organization, BRAC.

- Baseline data: collected before the clients were assigned to a treatment. (`Interview_nb=0`)

- Diaries: collected bi-monthly, or with higher frequency for some clients (`Interview_nb=1-11`)

- Endline data: collected 12 months after the Baseline, after completion of the loan cycle (`Interview_nb=12`)

Clients entered the BRAC lending program on a rolling basis and surveys were conducted accordingly. Baselines were thus conducted between October 17, 2014 and April 14, 2016, Diaries were conducted between December 1, 2014 and March 21, 2017 and Endlines were conducted between November 13, 2015 and June 26, 2017.

Each of the three types of survey data are described in more detail below.

4.1.1 Baseline and Endline surveys

Before assigning them to a treatment, we collected Baseline data on each client. The Baseline survey instrument includes detailed information about the firm's history and present firm operations, as well as about individual and household characteristics of the business owner/client. The business operation data includes information about sales and profits volumes, expenditures, the level and type of business assets held at Baseline, and the number and type of workers employed in the business. Individual characteristics of the business owner include gender and education as well as measures of financial literacy, time and risk preferences. We also collect comprehensive data on the business owner's household, including household size, income, data on other household members and their occupations, and other business activities run by the household.

Upon completion of the client's loan cycle, 12 months after the Baseline survey, an Endline survey was conducted, collecting the same type of information that was included in the Baseline.

4.1.2 Business Diaries

In addition to the Baseline and Endline surveys, bi-monthly business Diaries were conducted with all the clients in our study. A subset of clients were visited more than bi-monthly. Apart

from providing us with regular data on sales and profits, the Diaries focused on changes made in the business, as well as in the owner's household, between each visit of the interviewer.

4.1.3 Raw and created variables

There are two datasets that contain survey data. One is called `raw_data.dta` and one is called `created_data.dta`. See Section 5 for more details. All the survey data variables in both datasets can be merged to the other dataset using the unique identifier `idclient`. Further description about each variable, in what survey-rounds it was collected, and if it is raw or created, is provided in the codebooks.

4.2 Experiment data

This data includes: information on what treatment the client was allocated to.

The experiment data variables are collected in the data file called `treatment_data.dta`. It can be merged to the other dataset using the unique identifier `idclient`. Further description about each variable is provided in the codebook `Variables_treatment_CODEBOOK`.

4.3 Administrative data

This includes data on the unique id, area and local branch that the client belongs to, as well as the loan size of the loan taken as part of the project, and the timing of this loan and of various surveys collected on the client (the ID and location data are anonymized and translated into numbers). Variables related to the randomization structure (block and strata indicators) are also included here.

The administrative data variables are collected in the data file called `admin_data.dta`. It can be merged to the other dataset using the unique identifier `idclient`. Further description about each variable is provided in the codebook `Variables_admin_CODEBOOK`.

5 Using the data

The data files are kept in a repository at Stockholm University with DOI: `10.17045/sthlmuni.5734887`. There are 4 different datasets provided in this package. All the datasets can be merged to the other datasets using the unique identifier `idclient`.

5.1 Survey data

5.1.1 Data files

The survey data is divided in two data files: `raw_data.dta` and `created_data.dta`. The raw data contain all the variables from the questionnaire, with minimal modification. The only modification done since they were collected is renaming of variables so they have more intuitive names, and basic cleaning: dummies have been converted to values of 0 for "no" and 1 for "yes" rather than the automatically generated values of 1 and 2). The created data are variables that have been created based on the raw variables. For example, whereas the "year of birth" of the respondent is a raw variable, "age of the respondent" is a variable created by computing the difference between the year that the survey was collected and the reported year of birth.

5.1.2 Codebooks and questionnaires

To facilitate the understanding of the variables, we provide two types of documentation: Questionnaires (survey instruments) and codebooks. Each section of the questionnaire has a headline or label (appearing in column c: `label`) which is provided also in the corresponding codebook, to enable users to easily find the underlying survey question for each variable. The codebooks are arranged according to the division into raw and created variables, and administrative and treatment variables have separate codebooks.

1. Raw data variables codebook: `Variables_raw_CODEBOOK` lists the name and label of each variable as well as the type, the format of each variable and its position in the raw dataset. There are several observations for each client.

2. Created data variables codebook: `Variables_created_CODEBOOK` for each variable, in addition to the variable name this codebook contains a label or short description, and additional description about how the variable was created, and information regarding in what questionnaire rounds (Baseline, Diaries, End-line) the variable appears. The variables in this codebook are grouped according to their topic and according to if they have outcomes expressed in monetary values (costs, profits, sales) or non monetary (e.g. dummies, categories). There are several observations for each client.

3. `Variables_admin_CODEBOOK` and `Variables_treatment_CODEBOOK` contain one observation for each client.

10

5.1.3 Variables in different survey datasets

When using the survey data for analysis, it is important to notice that not all variables were included in all survey rounds. In the provided codebook for the survey data, information is provided regarding for what survey rounds each variable is available. The variables can be divided into the following 5 groups depending on which rounds they appear in. In the Stata data files (`.dta`), each group is associated with a "character", indicated after the description of each group below, divided into 5 spreadsheets. In the raw variables codebook, what survey rounds a given variable is available for is indicated through the name of the spreadsheet (more details on each spreadsheet are provided below) while in the created variables codebook, three columns indicate by YES/NO if a variable is available in the Baseline, Diaries, and Endline, respectively.

1. All interview variables: these are included in Baseline, Diaries and Endline. They include most questions that ask about business and household activities in the past 30 days (1 month). Spreadsheet: `ALL`

2. Baseline and Endline variables: Include all variables asking for sales/profits/costs/household income over the past 12 months; as well as details about other household owned businesses and details about other loans, shocks, transfers. Spreadsheet: `BL_EL`.

3. Diaries and Endline variables: Variables indicating change over time: investments, repairs, sold, stolen or damaged assets, household expenditures by categories, New employees and their type of payment, experience before joining the firm and if the owner knew them before hiring them. Spreadsheet: `Diary_EL`.

4. Baseline only: These include variables related to personal background of the respondent, their digit span results, time preferences and his/her parents that cannot change over time, and questions about loans taken, transfers and shocks in the past 2 years. Spreadsheet: `BL_only`.

5. Endline only: These include questions about other loans taken (apart from the loan that the experiment relates to), transfers and shocks in the past 1 year, i.e. the year of the experiment. It also includes a series of questions about the experience with the BRAC loan taken as part of the experiment. Spreadsheet: `EL_only`.

5.2 Missing variables conventions

In the raw data, whenever we do not have an informative answer for a question, the variable entry can be either missing (".")， 88, or 99. //

- Missing values ("." in Stata) appears for questions that were not asked to the respondent. There are two types of such cases: (a) the respondent was not interviewed for a particular survey round (diary), or (b) the data comes from a follow up question and was not relevant for the respondent in question. Example: Are there any goods that you produce in this business? If the answer is "yes", follow up questions will be asked about the names of the items. If the answer is no, the value of the follow up questions will be missing (".").

- If the respondent was unable to answer a question, the number 88 is used as a generic code for "Don't know".

- If the respondent refused to answer a question, the number 99 is used as a generic code for "Refusing to answer".

  In the `created_data.dta`, entries with the values of 88 and 99 have been replaced by missing values ("." in Stata). When creating new variables based on the raw data, this needs to be taken into account.

5.3 Conversion of monetary variables

In the provided data files, all monetary outcome values are expressed in their value in Ugandan Shillings. To make Baseline surveys carried out in different months and years more comparable, the values can be deflated to their October 2014 value using the CPI values from Bank of Uganda[‡]. We recommend to use October 2014 as the base-month for such deflation, since this was the month when the data collection was started. For conversion to USD we recommend to use the 2014 average rate of the World Bank PPP adjusted exchange rate for Uganda.

The deflation code used by the authors is provided in the file `Deflation.rtf`.

[‡]See `https://www.bou.or.ug/bou/rates_statistics/statistics.html`

5.4 Stratification, blocks and clustering

The stratification and block randomization needs to be taken into account when analyzing the data. The strata are defined by

- Geographical area that the client's local branch office initially belonged to. There were 15 such areas at the beginning of the project. As explained above, additional areas were added over the course of the project, but the initial area division is the relevant one for the stratification. The variable is called `area_loan`

- Previous loan experience: (i) new or (ii) repeat borrower. The variable is called `new`

- Sector (coarsely defined): (i) manufacture/service or (ii) retail. The variable is called `retail`

Firms that entered into the sample consecutively within a stratum were assigned blocks and the randomization was conducted *within* each block.The variable indicating the block is called `block`.

Whenever running regressions using several observation for the same client, we recommend adding the following Fixed effects (indicators): `block`, `Strata` and allowing for clustering of the standard errors at the client level (using the unique identifier variable `idclient`). Controlling for the identity of the interviewer can also be relevant, using the variable `RPOid`.

References

- **Bruhn, M. and McKenzie, D.** 2009. In pursuit of balance: Randomization in practice in development field experiments. American Economic Journal: Applied Economics, 1(4), pp.200-232.