Saving, by default

Pre-analysis plan

Outline

1. Introduction
2. Description of the sample.
3. Data sources.
4. Timeline
5. Hypotheses
6. Treatment effect equations
7. Multiple outcomes and multiple hypothesis testing
8. Attrition
9. Outcomes with limited variation
10. ITT – ATE
11. Other exploratory outcomes

1. Introduction

The first objective of most financial inclusion plans consists in having as many people as possible opening a bank account. A consecutive important step, for those who opened an account, is to receive payments on that account.

Due to a *statu quo* bias, the *'default option'* may strongly influence human behavior. For those who receive money on their bank accounts, that money is saved, by default. Unless they take the active step of withdrawing it from their account, their money will stay on the account. On the other hand, we expect people paid in cash to spend their money. Unless they take the active step to deposit it on their account or to freeze it under another form.

Based on that hypothesis, the objective of this study is to test whether moving beyond the first step of opening bank accounts and reaching the second step where people are paid on their accounts indeed substantially increases savings levels. And, if it does, whether the increase can be attributed to the '*default option'* behavior that we mentioned.

The study is carried out in collaboration with the Indian NGO Basix Sub-k. It builds on the recent expansion of the *business correspondent model* (BCM) in India. In collaboration with formal banks, Basix Sub-k is opening bank accounts in different States over the country. In rural areas, the NGO selects one villager to become the local banker (the *business correspondent sub agent* or BCSA). He receives training, a receipt machine, a finger print recognition device and a mobile phone connected through the mobile network to the partner bank. The BCSA can then use those devices to perform standard transactions on their accounts: deposits, withdrawals and transfers.

In villages where Basix Sub-k has recently opened BCSA accounts, we do weekly interviews during 7 to 13 weeks, to gather detailed information about the evolution of their household composition and the various earnings and expenditures of their household members. Because those surveys are extremely demanding, each participant is offered Rs150 after each interview.

Phase 1

To test the above hypotheses we designed the study as follows. We operate in an area with recent BCSAs. We select a random sample of people who opened an account (group A) and a random sample of people who did not. That second sample (of people without account) is subject to the first treatment: we give an account to two third of them. After this first treatment, we are left with three groups: A – already had an account, B – received a new account, and C – did not have an account and is not offered one.

The second treatment in the experiment is a randomization of the way the weekly compensation is paid. Half of the respondents with a BCSA account receive Rs150 directly on their account (treated), while the others receive it in cash (control).

The design is schematized in Figure 1 below.

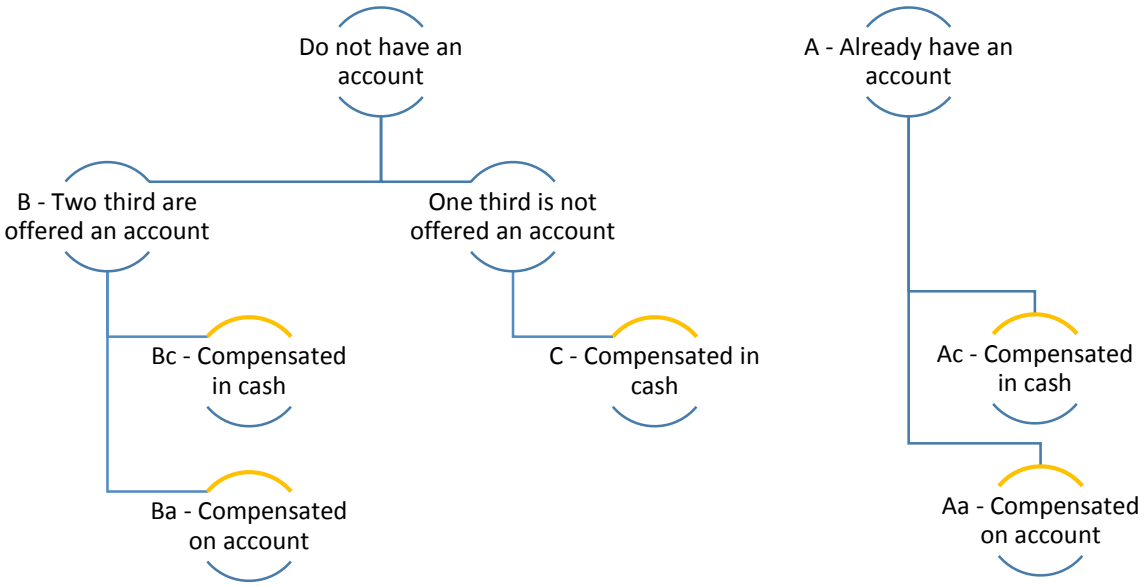

Figure 1: Experimental design in Phase 1.

We now have 5 different groups: Ac had an account and is paid cash, Aa had an account and is paid on it, Bc just received an account and is paid cash, Ba just received an account and is paid on it, C does not have an account and is paid cash. The comparison between these five groups will provide clear information on the impacts of the different steps of the financial inclusion process.

The first randomization, providing bank accounts to *people who initially did not self-select into the new banking system*, has two main goals:

- Measuring the first step of the process: having a bank account versus not having one
- Evaluating the importance of the initial self-selection. Existing studies suffer from the caveat that they only observe impacts on the pool of self-selected individuals (corresponding to our group A). Those impacts may not be representative of what a large scale financial inclusion plan (with accounts opened for *everyone*) would achieve.

The second randomization (cash versus account) aims at measuring the second step of the financial inclusion process: going from having a bank account but receiving cash payments, to having a bank account and being paid on that account. Importantly, given that (i) all respondents have an account (in groups A and B); (ii) there is no direct cost in depositing or withdrawing; (iii) all respondents in a village receive their compensation at the same location, close to the BCSA; we can attribute the treatment effect to the '*default option*' described above.

The main comparison groups are summarized in Figure 2 below. This is done for illustrative purposes and we intend to do further comparisons as will be clear from the specification of our main regressions in Section 6.

	C	Bc	Ba	Ac
Bc	The effect of obtaining a new bank account, while still receiving cash payments			
Ba	The combined effect of obtaining a new bank account and being paid on that account	Conditional on having obtained a new bank account, the differential effects of cash and account payments		
Ac	The effect of having chosen to open a bank account (self-selection), while still receiving cash payments	Conditional on being paid cash, the differential effects of having obtained a new bank account and having chosen to open one earlier (self-selection)		
Aa	The combined effect of having chosen to open a bank account (self-selection), and being paid on that account		Conditional on being paid on the account, the differential effects of having obtained a new bank account and having chosen to open one earlier (self-selection)	Conditional on having chosen to open an account (self-selection), the differential effects of cash and account payments

Figure 2: Summary of comparison groups.

Phase 2

One month after the last week of interviews of phase 1, we will re-start the weekly surveys for an additional period of four weeks. During these weeks, we will compensate everyone in cash as to test whether the initial treatments have impacts that last in the medium term.

2. Description of the sample

2.1 Sample size and design

The unit of observation is the head, or spouse of the head of a household. To ensure the geographical balance of the sample, we cluster the sample by villages (there is one BCSA per village). To guarantee that we could carry out the desired heterogeneity analysis in terms of gender, we stratify the sample by gender.

In each village we further distinguish three groups of villagers. We first have a random sample of people who already had opened an account with the BCSA before the intervention takes place. We also have a random sample of people who did not have an account yet. Among them, two third

(randomly selected) are asked to open an account. Basix Sub-k takes care of all costs and paperwork in order to open those accounts.

Our power calculations indicated that we need 18 villages and 32 respondents per village and thus a total of 576 respondents. Table 1 summarizes the different groups and treatments in one village.

<table>
<tr><td colspan="2">Group A – 14 respondents</td><td colspan="2">Group B – 12 respondents</td><td>Group C – 6 respondents</td></tr>
<tr><td colspan="2">Already had an account.</td><td colspan="2">Did not have an account, but opened one.</td><td>Did not have an account and were not asked to open one.</td></tr>
<tr><td>Aa
7 paid on account</td><td>Ac
7 paid cash</td><td>Ba
6 paid on account</td><td>Bc
6 paid cash</td><td>C
6 paid cash</td></tr>
</table>

Table 1: Sampling.

Each of the five arms (Aa, Ac, Ba, Bc, C) are stratified by gender in the following manner:

- Aa and Ac: In 9 village 3 men and 4 women are selected, in the other 9 villages 4 men and 3 women are selected
- Ba, Bc and C: 3 men and 3 women are selected in each village

2.1 Selection of villages

The villages are chosen in collaboration with the partner NGO Basix Sub-k according to the following criteria:

- Geography: rural, so the villages do not have their own commercial bank branch, and with clusters of villages sufficiently close to one another, so that the survey team can travel between villages in a reasonable amount of time. Furthermore, we excluded South India as our main local collaborator does not have expertise in this region.
- Sampling requirement: each of the 18 villages must include a high enough number of people with and without accounts.

Among the different areas where Basix Sub-k worked when we planned the survey, only two satisfied the first criterium: Moradabad district in Uttar Pradesh and an area of three bordering districts in Chhattisgarh. We choose to work in Chhattisgarh as the poverty rates are lower in this area, and as the villages are further away from bigger towns. The villages selected are located in the Magarload block in the district Dhamtari (5), in the Rajim block in the district Gariyabandh (7) and finally in the Abhanpur block in the district Raipur (6). Given that all BCSAs had sufficient villagers with and without an account, there was no need to make a selection based on the second criterium.

2.2 Selection of respondents in each village

The respondents in group A are randomly selected from the BCSA's customer lists. The respondents in the groups B and C, who are not BCSA customers, are randomly selected from the voter lists.

The households that already have a bank account with another institution (not the BCSA), such as a post office, cooperative bank, rural bank, or public sector or private commercial bank are excluded from the sample. There are two exceptions to this rule:

- We allow for those accounts if they were opened to receive welfare scheme or MGNREGA payments from the government and are used for that purpose only.

- We also accept cooperative bank accounts if they were not opened for any other purpose than the payment of paddy or other grains.

2.3 Randomization

1st randomization:

Two thirds of the sampled people without an account are offered one. The randomization is blocked by village and gender.

2d randomization:

Half the respondents in each block (defined by the groups – A and B, gender and village) are attributed to treatment.

The randomizations are done with the software Stata.

2.4 Variables included in tests of randomization and attrition balance

We will check the randomization and attrition balance on the variables defined in Section 5 below.

3. Data sources

We will use two data sources. First, Basix Sub-K will provide information on the use of the BCSA accounts. Second, first-hand data is collected by the research team through weekly household surveys.

3.1 Bank account details

Basix Sub-k will provide the data recorded by the bank. The data contains information on all the deposits, withdrawals, and transfers made or received by the respondent or any other household member during the period of the experiment.

3.2 Survey data

Phase 1

The respondents will be surveyed every week during approximately 7 to 13 weeks. The exact number of weeks in each village will depend on the time at which we can start surveying the village.

The survey consists mainly in questions about:

1) The evolution of the household composition in the past 7 days.
2) All cash/kind/labor/financial flows (earnings and expenditures) of the household members in the past seven days.

Phase 2

A few weeks later, the same survey is carried out again.

4. Timeline

Summer 2014: selection of villages

Fall 2014: baseline survey

January 2014: opening of new accounts

February - May 2014: weekly interviews and treatment (phase 1)

June - July 2014: Weekly interviews (phase2)

5. Hypotheses

5.1 Outcome variables

All our outcome variables are based on the bank account's data and on financial information from our surveys. We have this information on a weekly basis for the duration of the experiment. We will use both weekly values and averages over the whole course of the experiment.[1]

The first hypothesis that we want to test is whether the treatment increases the savings on the person's account. The outcome variable in this case is the **account balance (Y1)**, as recorded in Basix Sub-k's data. In addition to using (i) the final balance, we also plan to use (ii) the average balance on the account; (iii) the number of days the respondent had a positive balance; (iv) the maximum balance that was recorded. Villagers paid on the account are advantaged in these measures, since Rs150 are deposited on their accounts every week. To undo this advantage, we only include amounts that were at least 24 hours on the account for (ii), (iii) and (iv). This means that the money that is immediately withdrawn after we deposited it is not included in the calculation of the balances.

In the second part of the analysis, we use the survey data to test the treatment effect on the other household's savings and financial flows. We first consider the treatment effect on the household's **total net monetary flows (Y2)**, which is the sum of all the money that enters the household:

- Income from farming and livestock
- Income from forest products
- Income from wage employment and self-employment
- Income from renting out assets (land, machinery, animals, etc.)
- All transfers received (remittances, public transfers, gifts, and other private transfers)
- Loans and credits repaid

from which we deduct the total amount of money that leaves the household:

- Purchase of goods and services for consumption
- Purchase of inputs (farming, livestock, business)
- Purchase of durables assets

[1] For example, for the BCSA account, we will use the average balance per week, and the average over all the weeks.

- Rents paid for hiring assets (land, machinery, animals, etc.)
- All transfers made by household members
- Loans and credits paid
- Insurance payments

As a complement to this broad measure of monetary savings, we test the treatment effect on:

- **the households savings in informal groups (such as SHGs) (Y3)**
- **other financial assets, such as jewelry, money guarded by others, etc. (Y4)**

5.2 Covariates

Individual measures:

- C1 Caste
- C2 Age of the respondent
- C3 Married or not
- C4 Education : dummy equal to 1 if the respondent can read and write
- C5 Occupation dummies (employed*agriculture; self-employed*agriculture, not working)
- C6 Accounts held by the respondent (other than BCSA)
- C7 Group membership

Household characteristics

- C8 Land owned in acres
- C9 House quality: dummy by dwelling type
- C10 Number of adult members and children in the household (for Y2, Y3 and Y4 only)
- C11 Accounts held by the household (for Y2, Y3 and Y4 only)
- C12 Groups membership of the household members (for Y2, Y3 and Y4 only)

Distance to the BCSA

- C13 Geographic distance between the home and the BCSA.

Others:

- C14 Village fixed effects.
- C15 Time fixed effects (in the panel regressions only).

5.3 Heterogeneous effects

We plan to check for heterogeneous treatment effects along five dimensions:

- H1 Gender of the respondent.
- H2 Group category (defined as A, B and C in section 1).
- H3 Whether or not the respondent is in charge of the household savings.
- H4 Trust in the BCSA and the banks: each respondent is asked at baseline whether they trust the BCSA and the banks. We build a trust index equal to one if the answer to both questions is “quite a bit of trust” or “a lot of trust”. Otherwise, the index is equal to zero.

- H5 Time preference: a binary variable equal to one if the respondent makes the most impatient choice in our two baseline questions about (monetary) time preferences.

6. Treatment effect equations

6.1 Phase 1

First treatment: opening a bank account

Here we cannot use the dependent variable Y1, because the bank account data is only available for those who have an account, and not for group C. We first regress the outcome of interest on treatment status; i.e. we include ***O***, a binary variable equal to one if the individual was offered an account. The only covariates here are the village fixed effects $\boldsymbol{V_j}$:

(1) $$Y_{ij} = \alpha + \beta_1 O_{ij} + V_j + \varepsilon_{ij}$$

where Y_{ij} is the outcome for individual *i* in village *j*. We also study heterogeneity in the treatment effects by interacting the treatment with the variables defined in the "heterogeneous effects" section above. We run both separate regressions for each of the variables listed, and a joint regression including interaction terms for all the variables listed at once.

Equation (1) becomes:

(2) $$Y_{ij} = \alpha + \beta_1 O_{ij} + \beta_2 H_{ij} + \beta_3 O_{ij} * H_{ij} + V_j + \varepsilon_{ij}$$

where ***H*** stands for the vector of variables defining the heterogeneous effects of interest.

We also estimate both equations including a set of covariates **X** (defined in Section 5).

The equations are estimated using OLS. Finally, when the dependent variables are measured weekly, we follow a panel OLS procedure with time and village fixed effects.

We cluster the standard errors at the village level. Since we do not expect any negative treatment effect, we will use one-sided tests of the treatment coefficients in our main specification.

Second treatment: testing the *default*

This treatment only involves groups A and B, since group C does not have an account. We run exactly the same regressions as above, but the treatment variable is now ***A***, a binary variable equal to one if the individual was compensated on his account. The equations become

(3) $$Y_{ij} = \alpha + \beta_1 A_{ij} + V_j + \varepsilon_{ij}$$

And

(4) $$Y_{ij} = \alpha + \beta_1 A_{ij} + \beta_2 H_{ij} + \beta_3 A_{ij} * H_{ij} + V_j + \varepsilon_{ij}$$

As for the first treatment, we will also estimate both equations including a set of covariates **X** (defined in Section 5).

The equations are estimated using OLS. Finally, when the dependent variables are measured weekly, we follow a panel OLS procedure with time and village fixed effects.

We cluster the standard errors at the village level. Since we do not expect any negative treatment effect, we will use one-sided tests of the treatment coefficients in our main specification.

6.2 Phase 2

In phase 2, we follow exactly the same procedure as defined in 6.1, but this time using the dependent variables measured in the Phase 2 survey.

7. Attrition and non-response

We will check whether survey attrition is correlated with the treatments. If that is the case, we will estimate Lee-bounds (Lee 2009).

We will also follow Kling, Liebman and Katz (2007). We obtain lower bounds of the treatment effects by replacing missing observations in the treatment (control) arms by the corresponding arm's mean value minus (plus) 0.05, 0.10 and 0.20 standard deviations of the control group. Upper bounds of the treatment effects are constructed in a symmetrical way.

No imputation for missing data from item non-response will be performed. We will check whether item non-response is correlated with treatment status following the same procedures as for survey attrition, and if it is, construct bounds for our treatment estimates that are robust to this.

8. Outcomes with limited variation

We follow David McKenzie's approach: *"In order to limit noise caused by variables with minimal variation, questions for which 95 percent of observations have the same value within the relevant sample will be omitted from the analysis and will not be included in any indicators or hypothesis tests. In the event that omission decisions result in the exclusion of all constituent variables for an indicator, the indicator will be not be calculated"* (Development Impact blog, The World Bank, October 28th 2012).

9. ATT –ATE

We focus on intention-to-treat effects in our key results. Indeed, we cannot force all respondents to participate in each and every weekly interview and hence we cannot guarantee that they will all receive the exact same treatment.

References

Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz, ''Experimental Analysis of Neighborhood Effects,'' Econometrica, 75, 83–119, 2007.

Lee, David, "Training, wages, and sample selection: Estimating sharp bounds on treatment," NBER Working Paper No. 11721, 2009.