Pre-Analysis Plan

1. Introduction

Micro-insurance is a division of micro-finance, which means providing insurance products to low income households or individuals who have little savings, in order to reduce their vulnerability and save them from disastrous liabilities. Micro-insurance products are tailored to meet the needs of vulnerable households or individuals.

This project will involve marketing micro-insurance product, meaning rainfall index insurance product to rural farmers (in villages where there is no formal insurance presently available) and to urban migrants in Ouagadougou who originate from villages (close to Ouagadougou) and have relatives engaged in farming. The intervention is based on an existing rainfall index insurance product developed by the organisation PlaNet Guarantee for small-scale rural farmers in Burkina Faso. Subscribers can insure themselves against rainfall shortages in a specified location within the organization's coverage area. Rainfall is measured using high resolution satellite data (at the level of 10 square kilometers). Payouts are triggered if rainfall falls below a threshold specified at three stages of plant growth (germination, second-stage growth, fruit and flowering).

For individuals selected, the marketing process will be as follows: (i) a group presentation explaining the concept of insurance, and details about how PlaNet Guarantee's product works, the pricing scheme, etc. with the relevant information circulated on flyers; (ii) follow-up telephone calls with participants to gauge their interest in the product, and answer any questions that may arise; (iii) for participants who express interest in a subscription, scheduled house visits to carry out the subscription process.

The intervention will proceed in a similar manner in urban areas and in rural villages with the following differences: (a) participants from the same village will all be exposed to the same treatment, while urban migrants who originate from the same village may be subject to different treatments; (b) rural farmers receive the offer to purchase insurance for their own household, while urban migrants will potentially have the opportunity to name a beneficiary (potentially a rural relative) other than himself/herself. All rural farmers in the study villages and urban migrants from these villages (i.e. who were household members in any of the households located in the study villages) will be deemed eligible for the intervention.
As stated before, the study will be conducted in rural villages located in regions close to the capital city Ouagadougou and with migrants based in Ouagadougou. This combination of locations will be chosen to ensure that the study includes sufficient number of farmer-migrant pairs. Burkina Faso was chosen for the study because of the dominance of rainfed agriculture among smallholder farmers, the erratic nature of rainfall during the growing season, the near absence of formal insurance that can help farmers cope with weather-related shocks and the strong prevalence of rural-urban migration by individuals who retain strong ties with their rural households.

The intervention is based on an existing rainfall index insurance product developed by the organisation PlaNet Guarantee for small-scale rural farmers in Burkina Faso. Subscribers can insure themselves against rainfall shortages in a specified location within the organisation's coverage area. Urban migrants will be given the opportunity to purchase insurance for agricultural plots farmed by their rural relatives, with the contract specifying indemnity payments to be paid either to the subscriber or directly to their rural relative. The research questions are as follows:

(1) How does the uptake rate of rainfall index insurance among urban migrants compare with that of rural farmers? How do the respective uptake rates evolve over time?

(2) How does the uptake of rainfall index insurance by urban migrants affect agricultural investments, production choices and consumption smoothing of their rural relatives? How does this compare with the effects of insurance uptake by rural farmers?

(3) How does the uptake of rainfall index insurance by urban migrants affect their economic decisions, including transfers between urban migrants and rural relatives?

Four main assumptions are behind our theory of change. The first key assumption behind is that urban migrants with rural family members engaged in farming will find rainfall index insurance attractive at market prices. This assumption was tested during Phase 1 of the study. The data on the uptake of rainfall index insurance by urban migrants in Phase 1 of the project, demonstrated that this is more cost-effective (cost per subscription) than marketing the product directly to rural farmers. We believe that evidence of this nature, when presented to the key industry players, will attract private insurers to the market and thus increase the supply of formal index-based insurance without subsidization by governments or international development agencies.

The second assumption is that for urban migrants who provide assistance to their rural families in times of need, access to rainfall index insurance will reduce their risk burden, and thus allow them to pursue more profitable economic opportunities. Rural farmers who obtain access to formal insurance via their urban relatives will be able to smooth consumption more effectively. In addition, they will be encouraged to make farming and investment decisions that involve more risk but entail higher expected profits.

The third assumption is that when urban migrants subscribe to the proposed insurance product, this would also provide rural farmers indirect experience with formal insurance. We hypothesize that this experience can help rural farmers improve their understanding of formal insurance, trust in formal insurance providers and, in the long run, improve direct uptake by rural farmers. (The enclosed diagram provides a schematic view of our theory of change).

The last assumption is that the Phase 2 evaluation will provide more accurate data on the cost-effectiveness of marketing the rainfall index insurance product to urban migrants. In addition, the evaluation will allow us to investigate how uptake of insurance affects urban-rural transfers, and its effect on investments, production choices and consumption of rural households. As we are not planning to conduct an endline survey with urban migrants, the evaluation will not tell us how the product affects their economic decisions.

Based on our findings from the Phase 2 evaluation, we plan to engage in communication and engagement activities with policymakers and key players in the insurance industry which will allow us to explore the final step of our theory of change.

2. Outcomes

a. Intermediate outcomes

The intermediate outcomes of the intervention are as follows:

- ✓ Take-up of the rainfall index insurance which is measured as 1 if a paired rural household-urban migrant link purchases insurance and 0 otherwise.
- ✓ The value of the capital or acreage insured, measured in CFA insured for the capital, and in hectares for the acreage.

b. Final outcomes

The expected final outcomes of the study are organized in three broad categories as follows:

i. *Risk coping and management mechanisms*
   a. Land area allocated to food crops and cash crops; and to crops with different levels of sensitivity to variations in rainfall;
   b. Seasonal migration from rural households and participation in non-farm economic activities;
   c. Off-farm activities;
   d. Education: measured as enrollment of children between 6 and 15 years old;
   e. Livestock (herd size and sales);

ii. *Investments in agriculture*
   a. Cultivated area: measured as hectares of land cultivated;
   b. Fertilizer use, measured in kilograms of fertilizers per hectare of cultivated land;
   c. Labour input, measured in labour per unit time per hectare of cultivated land, at different stages of the farming cycle;
   d. Improved seeds, measured as acreage planted with improved varieties;
   e. Agricultural productivity, measured as the value of harvest (net of marketed input costs) per hectare.

iii. *Consumption smoothing and income growth*
   a. Consumption smoothing by rural farmers, measured as changes in consumption. Insurance should shelter consumption from income shocks;
   b. Income growth: measured as the change in household total income between the baseline and the endline.

3. Methodology

a. Sampling

For rural households, we will employ a multi-site cluster randomization technique using blocking before randomizing the villages to the different treatments. In cluster randomized block design, communes will be the block, while villages will be the clusters. For urban migrants, we will randomize at the migrant level. Hence, the study sample will consist of rural farm households and their migrant relatives who reside in Ouagadougou. Building on the approach tested in the pilot study, the sampling procedure will include 3 steps that can be described as follows:

1. We start by conducting a household census in sampled villages. During the census, households are asked whether they have relatives who live in Ouagadougou with whom they are in regular contact. The census thus produces two groups of households, one group without urban migrants and another with urban migrants.

2. At the time of administering the full survey, we collect full contact details including telephone numbers of the urban migrants from the rural households. Based on the experience of the pilot, the enumerators will be instructed to make the first call to the migrant at the time they collect the information from the rural household. Thus incorrect phone numbers can be corrected immediately.

3. We will use the information provided by the rural households to track their migrant relatives in Ouagadougou.

The end result of this process is a random sample of rural households and a random sample of urban migrants who are related to the rural households.

b. Randomization design

We randomize at the village level in the rural area, and at the individual level in the city. In villages randomized to the treatment group, the insurance provider will offer its standard product, i.e. market rainfall index insurance to rural farmers. Migrants from both treatment and control villages who live in Ouagadougou will be randomly assigned to two groups. The first migrant group will be offered to purchase the insurance product to cover their relatives in the villages, while the insurance product will not be marketed to the second migrant group (the control). To simplify the exposition, we denote by V0 and V1 the control and the treated villages, and by U0 and U1, the control and treated urban migrants. Combining rural households and urban migrants produces four experimental groups that can be described as follows.

1. V0U0: The pure control group: the insurance product is not marketed to rural households or to their migrant relatives.
2. V0U1: The insurance is marketed to the urban migrants, but not to their rural relatives.
3. V1U0: The insurance is marketed to rural households, but not to their urban relatives.
4. V1U1: The insurance is marketed both to rural households and to their urban relatives.

c. Sample size and power calculation

First, we compute the sample size required to detect an increase in take-up of 21.7% (as obtained in the pilot study), starting from a baseline of zero take-up. We assume an 11 ppt difference between treatment groups, i.e. half the take-up rate observed during the pilot. Pairs of farmers and migrants (who share a family link) will be randomized across the treatment and control groups described above. Assuming $\varkappa$ =0.80 (statistical power) and α=0.05 (statistical significance), we will need 254 pairs of farmer-migrant pairs to detect a 21.7% change in take-up and 11 ppt difference between groups (see Table 01 attached). Thus, we require 1016 urban migrants for the four experimental groups. Since randomization is at the level of the farmer-migrant pair, we do not adjust the evaluation sample for intra-cluster correlation.

We then compute the number of rural households required to attain a sample of 1016 urban migrants. In Phase 1, we successfully tracked 69% of migrants (170/247), and successfully interviewed 72% of those tracked (124/170). The Phase 1 data shows that households with migrants in Ouagadougou were linked to 1.45 migrants on average. We plan to oversample such households so that they constitute 75% of the rural sample (we deem this to be feasible based on the Phase 1 data). Based on these figures, our calculations indicate that we need a sample of 1600 rural households.

For the rural survey, we will start with 80 villages, and randomly sample 20 households in each village. We will block villages at the commune level with 8 villages per commune. We then proceed to estimate the minimum detectable effect (MDE) for the economic outcomes of interest: consumption, fertilizers, cultivated area, and education, using data from the Burkina Faso 2010 DHS (for education) and from the Ministry of Agriculture of Burkina Faso for the remaining outcomes (the construction of the latter variables are described in Kazianga and Wahhaj (2017). The data are summarized in Table 02 attached. We calculate the MDE under two scenarios. In the first case, we treat any rural household in a treated link as treated. We also account for the proportion of variance explained by blocking. We assume 20% take-up rate. Under these assumptions and with the figures shown, we can detect a minimum change equivalent to roughly 40% of the standard deviation. This MDE corresponds to CFA 198,802 increase in consumption, 17.6 KG increase in fertilizers, 1.7 ha increase in cultivated area, and 19% increase in current enrollment for education.

In the second case, we consider comparing V0U1 with V0U0, and V1U1 with V1U0. In this case, we can again ignore the intra-cluster correlation (as when comparing take-up) and block at the village level. Assuming again a take-up rate of 20%, the MDE is, in this case, about 0.30 for the outcomes we consider.

These MDEs, while large, do not account for the additional power that we will gain from having a baseline and for controlling for additional covariates. They are, on balance, very conservative.

d. Data sources (census, surveys and administrative data)

At the start of the intervention, IPA will produce a list of targeted participants together with contact details, based on information collected during the (baseline) rural and urban household surveys. PlaNet Guarantee will then make contact with the targeted participants according the description of the intervention provided above. Throughout the intervention, the organisation will keep record of

their activities with each individual; i.e. whether attempts at contact were successful, whether the participant attended the organisation's presentation, whether and when a follow-up call was made, whether a house visit was conducted and the outcome of the house visit. The organisation will also make available data on subscription, details of the contract (e.g. who was the named beneficiary), indemnity payments, etc.

Quantitative surveys will be administered to rural households and to urban migrants (IPA will coordinate the data collection). We will work with the insurance provider to collect take-up data and basic information on household and migrants between the baseline and the end-line surveys. The survey questionnaire will be programmed on SurveyCTO and tested on the field during a pilot. In addition, the survey questionnaire will be translated into the local language by a professional. For the urban survey, and building on the experience of the pilot study, we will organize telephone call sessions to get in touch with all of the migrants indicated by their rural relatives as living in Ouagadougou. Then, our call center officer will attempt to call back all migrants for whom contact information would have been collected during the rural survey. On site verification of survey will consist of backchecking about 10% of households interviewed the previous day, selected randomly. Administrative data on subscription will be collected by Planet Guarantee.

4. Estimations

a. Balance checks

In the first step of the analysis, we will compare household characteristics and all primary and intermediate outcomes variables in the baseline data across the four treatment/control arms to check whether balance was achieved in assigning household-migrant pairs to different arms. Balance tests across treatment arms will be done within a regression framework so that we can include randomization strata. The balance specification includes dummies for each of the three treatment arms, and dummies for randomization strata, clustered at the village level.

First, the balance checks will report the means and standard deviation for the treated and control groups and a test of significance for each treatment pair (including all pairwise combinations possible between treatment and control groups).

Second, we test the hypothesis that the means are equal across the treatment arms and the control groups by regressing each variable of interest on the treatment dummy variables and province fixed effects. For each variable, we will report two series of tests: (i) t-tests on the difference between each specific treatment arm and the control, and (ii) an F-test on the equality of the means of all groups. We will also report an F-test from a regression of treatment on all the primary and secondary outcomes.

These tests provide evidence of the balance between treatment and control group. Successful randomization should ensure that the experimental arms do not differ along observable and (presumably) unobservable characteristics. Because we will be testing a larger number of variables, imbalances may occur by chance. Such imbalances should not be considered a source of concern if they do not occur for more than 5% of the variables, when the significance level for hypothesis testing is set at the 5% level.

b. Main effects

We plan to estimate the impact of the intervention in two ways. First, we propose to conduct a comparison of means for the intermediate outcomes (take-up, value of the insured land parcel), and final outcome variables (land under cultivation, crop composition, farming inputs, seasonal migration and participation in nonfarm economic activities, consumption variability, livestock, school enrolment of children) across the four treatment/control arms.

Second, we plan to estimate equations of the form

$$\Delta Y_{iv} = \beta_0 + \beta_1 Treated_{iv} + \beta_2 Rainfall_{iv} + \beta_3 Treated_{iv} \times Rainfall_{iv} + \varepsilon_{ivt} \quad \textbf{[1]}$$

where ΔY_{iv} is the change in outcome variable Y for household i in village v between the baseline and the endline surveys, $Rainfall_{iv}$ is a measure of rainfall experienced by the household following the intervention. And $Treated_{iv}$ is a binary variable indicating the treatment status for the urban migrant linked to household i.

We will conduct the estimation for two subsamples:

(i) the subsample of rural-urban household-migrant pairs in which the rural households were not exposed to the intervention.;
(ii) the subsample of rural-urban household-migrant pairs in which the rural households were exposed to the intervention.

Thus, the estimates based on the first subsample would provide an estimate of the impact of marking insurance to urban migrants when no index insurance is being marketed in rural areas; while the estimates based on the second subsample provides an estimate of the impact of marketing insurance to urban migrants when index insurance is being marketed in rural areas.

We will use OLS to estimate regression (1). We will also run separate regressions including:

- Baseline covariates: The vector of covariates will include household size and composition by age and gender, land size, and wealth measured at baseline.
- Province fixed effects and province-specific trends among regressors in order to control for underlying difference across provinces and district-wide change over the programme period.

For binary outcomes, we will conduct robustness checks using probit or logit models. All parameters of interest will be assessed using two-sided tests, at usual significance levels (10, 5 and 1%), and on the full sample. The unit of analysis is (i) the rural household, (ii) the urban migrant, and (iii) the pair of rural household-urban migrant. Standard errors will be clustered at the level of the unit of randomization, i.e. at the village level. Since we consider several outcomes, following Anderson (2008), we plan to implement multiple hypothesis testing within each broad category of outcomes. We will employ Hocheberg step down procedure and p-values adjusted for multiple outcomes inference (Family-wise Error Rate Control and False Discovery Rate Control), as well as the procedures suggested by List et al. (2016).

c. Heterogeneous effects

We also plan to adapt equation [1] to explore heterogeneity in the impact of intervention. Equation [1] already includes one dimension of heterogeneity, namely the rainfall experience of the rural household. Other dimensions of heterogeneity include

(a) the social network of the rural household for sharing risk or information, measured at baseline;
(b) household assets that may be used for self-insurance, measured at baseline;
(c) the 'quality' of the rural-urban migrant link – in terms of the closeness of the relationship, or the frequency of transfers – in the baseline;
(d) occupation of the migrants, in terms of formal sector (i.e. steady and certain income flows) and informal sector (uncertain and variable income flows).

These other dimensions of heterogeneity will be introduced into equation [1] in the same manner as the rainfall measure; in the form of an interaction term involving the treatment status variable.

5. Addressing incomplete data

We anticipate two potential sources of incomplete data: i) attrition across the survey rounds, and (ii) failure to track some rural migrants.

a. Attrition

A regression of the probability of attrition on treatment status and individual observable characteristics at the baseline will be run in order to check differential or systematic attrition along observable dimensions. Should we get differential attrition with respect to the treatment status, we would run Lee Bounds (Lee, 2008) as a robustness check for the impact exercise. We would also assess the nature of the attrition issue and try to argue on the nature of any attrition-induced bias, specifically whether said bias is upward or downward.

Yet, some level of attrition is unavoidable in longitudinal surveys. We anticipate a default attrition rate of about 5% per year based on some previous surveys from Burkina Faso. To minimize attrition, we plan on tracking farmers who left the sample villages. In particular, we will implement the following strategies:

- We will actively track households and migrants who left the sample villages or the city and who reside within 45 kilometers from their original residence.
- For those who reside more than 45 kilometers away or who cannot be successfully tracked, we will administer a limited version of the survey to their neighbors who did not move.

b. Missing data from migrants

We will build on lessons learned during the pilot study (Kazianga and Wahhaj, 2018) to maximize the rate of migrants who are successfully tracked. At the time of administering the full survey, we collect full contact details including telephone numbers of the urban migrants from the rural households. Based on the experience of the pilot, the enumerators will be instructed to make the

first call to the migrant at the time they collect the information from the rural household. Thus, incorrect phone numbers can be corrected immediately.

c. Dealing with outliers.

To deal with outliers, all continuous variables are winsorized at the 98th and 2th percentiles within the household survey, at the most disaggregated level possible. Component variables are not winsorized. For example, the value of per capita consumption of millet over the last 7 days is winsorized, not the household total consumption value. Specifications will be checked using both the original and windsorized variables. The stability of coefficients will be tested through standard hypothesis testing (Wald test).

6. Calendar and budget

We expect the project to last between 2018 and 2021.

References

Anderson, M. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, Perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Kazianga, H. and Wahhaj, Z. (2017). Intra-household Resource Allocation and Familial Ties. *Journal of Development Economics*, 127: 109-132.

Kazianga, H. and Wahhaj, Z. (2018). Will Urban Migrants Formally Insure their Rural Relatives? Family Networks and Rainfall Index Insurance in Burkina Faso. *University of Kent Economics Discussion Paper Series* 18/03.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071-1102.

John List & Azeem Shaikh & Yang Xu, 2016. "Multiple Hypothesis Testing in Experimental Economics," *NBER Working Paper* No. 21875