Data Analytics to Generate Deeper Insights on Smallholder Households

Background Reading

Expert Workshop for Feedback

Thursday 9 February 2017, 2:00-5:30
Ledger Hotel
Dar es Salaam, Tanzania

RESEARCH FRAMEWORK

In this document, we summarise the three-layered research framework that we have developed for this analysis. The framework is based on a thorough literature review, conversations with a range of industry experts and discussions with CGAP staff.

We begin by taking a broad view of the key factors that can indicate a smallholder household's ability to and likelihood of accessing and using different financial services. We then zoom in on the financial behaviours of smallholder households, to understand the different mechanisms that are used to meet economic and social needs. Finally, we look at what the data can show about the business case for serving smallholder households, and what kind of approaches we might take to segmenting the market.

In each of the three sections below we have included a table with research questions. The questions are numbered to be consistent with chapter 7 of our inception report, and refer to a data source and a type of analysis. Numbers in the 'type of analysis' column refer to the annex at the end of this document.

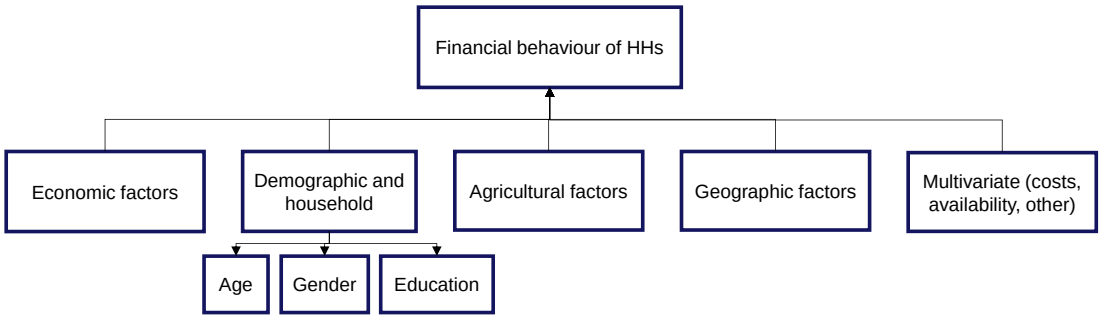
1. HOW AND WHY DO SMALLHOLDER FAMILIES ACCESS AND USE FINANCIAL SERVICES?

As a starting point for our analysis, we plan to explore some high level questions around the access to and usage of financial services by smallholder families.

Taking the six survey datasets as the basis for our analysis, we will look to isolate the key determinants of financial decision making in smallholder households, with an emphasis on those in Mozambique, Uganda, Bangladesh, Tanzania, Cote d'Ivoire and Nigeria. Where possible, we will bring in relevant data from Finscopes, FSP maps and FII here.

The analysis will use a range of econometric techniques based around regression analysis. Essentially we are hypothesising that access to or usage of a particular financial service, or type of financial services (e.g. digital), is a function of a wide range of variables. These include data points around personal, social, economic, household, environmental, agricultural and other characteristics.

Figure 1: Access and use of financial services - research tree



Depending on the data (in particular the characteristics of the dependent variable), we will use standard multivariate linear regression, probit/logit, and multinomial logistic regression. Please see below the research questions that we aim to address with our analysis. Please refer to the inception report annexes for more detail on the type of analysis we envisage for each of the research questions below.

Table 1: Research questions (Section 7.1)

Thematic Area	RQ#	Research Questions	Data	Analysis [1]
Economic factors	7.1.1	Which factors determine income and expenditures of small-holder farmers?	Surveys	6
	7.1.2	How do more formally banked HHs compare to informally banked HHs with respect to commercialisation, total income, physical asset portfolio, expenditure?	Surveys and Diaries	3, 4
	7.1.3	How is access to and use of financial solutions associated with the most positive livelihood outcomes (i.e. low poverty, high income, low income volatility) for smallholder households?	Surveys and Diaries	3, 4, 6
Demographic and household factors	7.1.4	How does the financial activity of female household members differ from male, and how do the gender roles tend to vary between countries?	Diaries	3, 4
	7.1.5	How does the behaviour of older family members compare to younger family members (e.g. in terms of financial activities, diversity of incomes, risk appetite, overall livelihood strategies)	Surveys and Diaries	3, 4
	7.1.6	How do educated HH members compare to less educated HH members in terms of income generating activities, making HH expenditures, asset purchases/sales, and savings/borrowing behaviour?	Surveys and Diaries	3, 4
	7.1.7	What are the important factors in determining propensity to use digital financial services? How important is age in this?	Surveys	4,6
Geographic factors	7.1.8	By linking the geotagged Surveys data to data from FSP maps where available, what can we say about the role of geography and relative location in determining usage of financial services?	Surveys, FSP Maps & Diaries	4
Agricultural factors	7.1.9	How do the access to and use of financial services relate to linkage strength with value chains? What other features of agricultural activity are related to usage of formal and informal tools? Can we explore the importance of connections to markets as a determinant of income, poverty, assets, livelihood strategies, digitization, etc.	Surveys and Diaries	1,4,5
Other variables	7.1.10	How are the outcomes of access to and use of finance associated with the development stage of the financial industry in the country?	Surveys and Diaries	2
	7.1.11	What are the key drivers behind the usage of informal financial tools? What can the data tell us about the relative role of cost, flexibility, availability and other factors in making these decisions?	Surveys and Diaries	4,6

[1] Numbers refer to types of analysis in Table 4: Summary of types of analysis (see Annex).

2. HOW DO SMALLHOLDERS MANAGE THEIR INCOMES AND EXPENDITURES FOR RESILIENCE, INVESTMENT AND OTHER NEEDS AND WANTS?

This section will map research questions using the diaries data on to initial findings from our analysis of national surveys (and other macro sources), to conduct a much deeper review of financial activities of the target populations. The primary data source will be the smallholder diaries data from Tanzania, Mozambique and Pakistan, which will be complemented with additional data where possible. Though the statistical significance of these data are limited, they provide a unique opportunity to examine in more detail some of the key issues that have emerged from our initial analysis. The data will help to tell stories and add narrative to some of the trends identified in our earlier analysis. Some of the questions we ask of smallholder diaries will be deeper explorations of questions already explored, however we will use a range of additional techniques to analyse these very different data sets. These include the following.

Cash flow modelling and graphical analysis

The unique time series allows us to track dynamic trends over the course of the agricultural and annual cycles. By modelling how income and borrowings map on to expenditures and savings we can track how liquidity is managed and investigate how decisions are made. By comparing within and between households this provides a simple framework for understanding how key variables such as gender, age and income affect financial behaviour.

Consumption smoothing analysis

An initial scan of the data shows us that there is variation in how different households manage their income to finance expenditure and investment over time. Where some households manage to smooth consumption to a certain degree and save up for larger, potentially production-related, investments, others go through 'hunger periods' and struggle to make larger investments. Patterns of how different types of households make small and regular transactions versus larger and lumpy ones can tell us a lot about variations in behaviour.

Key moment analysis

The data suggest that there are certain important transactions that occur in the financial cycle of a household. Examples might be the purchase of an asset, investing in a savings product, or taking out a loan. Here we will build on Rutherford's concept of *useful lump sums.* These large lump sum financial transactions can form pivotal points in the lives of smallholder households and we will identify and examine these in detail. Another example could be households that transition towards commercialisation or from using informal services to formal. Using the diaries data, we will classify and rank households according to *degree of commercialisation* or *financial product usage* on a monthly basis to identify households that transition in relation to other households over time.[2] In particular we will look at the behaviour around these key moments, to see what lead and lagged indicators are significant in understanding these decisions, and the effects that they have.

We will analyse these data to provide insights around how smallholder households manage money, manage risk and accumulate assets. Some of the research questions that we intend to study in this analysis include the following.

[2] See also Box 1 below for more detail on smallholder commercialisation indices.

Table 2: Research questions (Section 7.2)

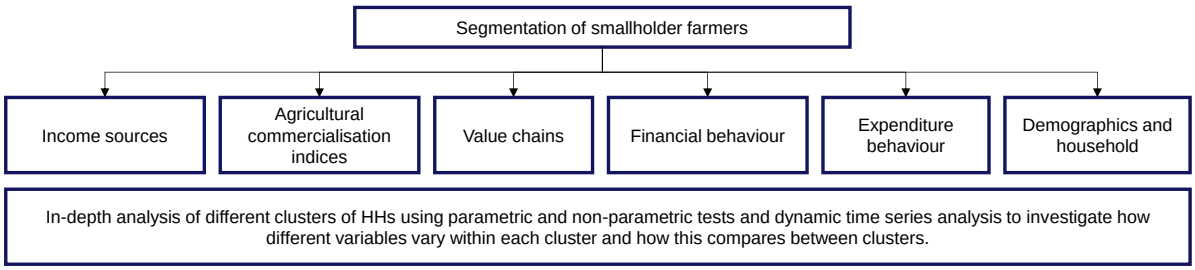
Thematic Area	RQ#	Research Questions	Data	Analysis
Money management	7.2.1	How does the composition of household income and expenditures vary across time? How are different income streams (e.g. agricultural sales, labour, remittances) and expenditure streams (e.g. education, health, non-agricultural investments, agricultural investments, general consumption) related and what are the covariance patterns between them?	Diaries	1,3,5
	7.2.2	For households switching from informal to more formal financial services over the year, what are the key drivers behind these decisions?	Diaries	2,3
	7.2.3	What is the role of women in managing the flow of funds and how does this vary between country datasets?	Diaries	3,4
	7.2.4	How does income relate to behaviour around expenditure, savings and borrowing? And what are the effects of income volatility on livelihood outcomes?	Diaries	3,5
	7.2.5	For smallholders using credit, what types of income contribute to loan repayment? What can the data tell us about the role and relative importance of agricultural cash flows and non-agricultural cash flows in loan repayment?	Diaries	3,5
Risk management	7.2.6	What can we learn about the nature and severity of shocks to cash flows and what is the role of risk mitigation tools?	Diaries and Surveys	3,4,6
	7.2.7	How does income volatility relate to other variables, such as poverty level, crop choices and diversification, and other aspects of livelihood strategies?	Diaries	3,5
Asset accumulation	7.2.8	For households who are seen to significantly increase or decrease their incomes, switch from non-commercial to commercial activities or significantly change livelihood strategies within the annual cycle, what are the key variables that allow these transitions to take place?	Diaries	2,4
	7.2.9	How are households investing in physical assets? Which assets complement income generation, which assets are 'locking in liquidity' and are therefore a form of savings and which assets are for household usage?	Diaries	2

3. SEGMENTATION AND BUILDING A BUSINESS CASE

The final section will bring together the macro and the micro analysis and answer some questions that are more specific to the needs of financial institutions and other key stakeholders. This will be a more forward-looking section, to understand the practical implications of the analysis and how it can be used to build a business case for more effectively serving smallholder households.

In working to build models for segmentation, we will look to more behavioural approaches, seeking not just to look for market segments by socioeconomic characteristics and incomes but also how they use their money (expenditures) and how they interact with financial services. In this way we hope to build a framework that is more accessible to financial institutions.

Figure 2: Segmentation of smallholder farmers



We will aspire to add depth to the understanding of smallholder farmer's lives by making use of the granularity of the diaries to segment the sample by demographics, expenditure patterns, financial usage patterns and income patterns. As we segment the sample in different ways we will aim to identify clusters of households with commonalities across different variables. We will then conduct a deeper analysis of these commonalities. Here we will compare households both in terms of the annual averages of variables of interest and behaviour of both households and individual household members over time. We will make use of techniques mentioned under 7.2 (consumption smoothing analysis, key moment analysis and cash flow modelling and graphical analysis) to look at dynamic trends. In preparation of this analysis we have produced a master data set that will allow us to conduct this type of analysis.

In this section we will particularly focus on investigating alternative ways of segmenting and analysing smallholder families and their behaviour. Firstly we will aim to expand on the conventional definitions of smallholder commercialization by making use of the granularity of the diaries data. Second we will aim to segment households in new and alternative ways and see how this might be interesting for FSPs in terms of customer profiling. One example would be to consider that not all households might choose commercialisation as their desired way out of poverty. While some do, others might want to focus more on waged employment of self-employment and are using agriculture only as another complementary income source. Households focusing on progressing their agri-business then might need a different financial product portfolio than households who are aiming for waged employment. As such we will not only profile HHs according to income but will also explore alternative routes to pick-up on profiles by looking at expenditure, type of income mix, borrowing behaviour, saving behaviour, to create profiles of HHs that we can compare.

Table 3: Research questions (Section 7.3)

Thematic Area	RQ#	Research Questions	Data	Analysis
Segmentation	7.3.1	Who is a smallholder farmer? What are the key criteria in defining smallholder families as a market?	Diaries and Surveys	2
	7.3.2	What different segments of smallholder households can be defined according to their household characteristics and source of income mix? (segmentation by demographics and income)	Diaries	1
	7.3.3	Are there differences in how various subgroups of smallholder households – by gender, different income levels, or education – spend their income (aggregately or proportionally)? (segmentation by expenditure)	Diaries and Surveys	1,3,4
	7.3.4	Can we develop an index of commercialisation, in order to segment smallholder farmers by the level of commercialisation of their agricultural activities? (segmentation by commercialisation)	Diaries	1,4
	7.3.5	What profile of smallholders is most likely to demand and actively use certain financial products (e.g., savings, layaway, mobile money, general short- and long-term credit, input credit) and formal and informal tools? How does this differ between meeting general household needs and priorities, and those linked to agriculture? (segmentation by financial usage)	Surveys	6
	7.3.6	How can we use what we've learned to build on previous segmentation models (e.g. no/loose/tight value chains)?	Diaries and surveys	1,4
Building a business case	7.3.7	What are the gaps between current smallholder demand and the financial solutions available (e.g. current terms do not meet seasonal constraints)?	Diaries and surveys	1,2,3,6
	7.3.8	To what extent can we view smallholder households as businesses? How are their cash flows and working capital needs similar/different to a small business?	Diaries	2
	7.3.9	What incentives, motivations or aspirations might drive uptake and usage of DFS among smallholder farmers? Do they differ according to farmer segments?	Surveys	4,6
	7.3.10	What kind of transactions are indicative of a good customer for a financial institution? Conversely, what are the 'red flag' behaviours that indicate likelihood of a poor customer?	Diaries	6
	7.3.11	Can the data be useful in building credit reference models for smallholders? Are there elements in the data that can give smallholders a credit history or information trail FSPs can use to make lending decisions?	Diaries	4,6
	7.3.12	What can we learn from transaction behaviour about the velocity of money in rural communities? And what are the implications for design of DFS?	Diaries	2,3

Annex: Description of types of analysis

Throughout our analysis we will make use of a common set of analytical techniques. Please see Table 4 below for a short summary of each technique as well as an applied example.

Table 4: Summary of types of analysis

Ref #	Type of analysis	Summary	Example
1	Segmentation / Cluster Analysis	Segmentation approaches identify clusters of customer attributes or behaviour. From these we can define (and estimate the size of) discrete sub-groups of the population that help to understand the differing needs within the market.	We will identify different types of smallholder farmers by developing a commercialisation index where we will use information on the share of agricultural income over total income as a proxy for commercialisation. See also Box 3.
2	Summary Statistics and Frequency / Cross-frequency Analysis	Summary statistics are used to make broad comparisons. It provides statistical measures such as, [1] Mean, median and mode, [2] Maximum, and minimum, [3] Variance and standard deviation, etc.	We have provided an example of mean comparisons in Box 1 above. We also do provide cross-frequencies for variables in the diaries data in Annex 5.
3	Graphical Analysis	Graphical analysis can show the trend and changes of data over time. It is also used to show the differences between different data groups. Type of graphical charts are, [1] Column and bar charts, [2] Line graph, [3] Pie chart, [4] Area chart. etc.	We will use graphical analysis for most research questions and specifically in the segmentation part of 7.3 to visualise how different clusters of smallholders compare in different terms of financial behaviour or between gender and age.
4	Parametric and non-parametric approaches	Parametric or non-parametric approaches can be used to test the null hypothesis in a statistical analysis. [1] Parametric tests includes paired t-tests, unpaired t-tests, one way ANOVA [2] Non-parametric tests includes Wilcoxon Rank Sum test, Mann-Whitney U test, Median Comparison, Kruskal-Wallis test.	This type of analysis can be used to test whether the difference between two groups is statistically significant. We will use these tests in Section 7.1 to, for example, look at how more formally banked HHs compare to informally banked HHs with respect to commercialisation, total income, physical asset portfolio, expenditure.
5	Correlation Analysis	Correlation analysis is used to assess how two variables move in relation to each other.	This analysis does not explain relationships between variables but simply show whether variables move positively or negatively in relation to each other and to which degree. Amongst others we will use this technique in Section 7.2 to analyse how levels of income correlate with income volatility.
6	(Multi) Regression analysis	Regression analysis is used to identify any relationship between a dependant variable and one or more independent variables. The following regression techniques will be adopted in the analysis, [1] Multiple regression analysis, [2] Logit/Probit regression, [3] Multinomial logistic regression	Regression analysis will be used throughout our analysis – both for key moment analysis, other time series analysis as well as when comparing average values in a cross-section under 7.1. Example of questions where we will use regression analysis are: [1] What profile of smallholders is most likely to demand and actively use certain financial products and formal and informal tools? [2] How does this differ between meeting general household needs and priorities, and those linked to agriculture?