



**HAL**  
open science

## **RUME (2010) Survey Manual**

Isabelle Guérin, Govindan Venkatasubramanian, Sébastien Michiels, Marc Roesch,  
Arnaud Natal

► **To cite this version:**

Isabelle Guérin, Govindan Venkatasubramanian, Sébastien Michiels, Marc Roesch, Arnaud Natal. RUME (2010) Survey Manual. French Institute of Pondicherry; IRD - Institut de recherche pour le developpement. 2010. ⟨hal-04153130⟩

**HAL Id: hal-04153130**

**<https://hal.science/hal-04153130v1>**

Submitted on 6 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



# RUME

RURAL MICROFINANCE AND EMPLOYMENT SURVEY

<https://odriis.hypotheses.org/rume>

## Survey Manual

2010





**RUME research team:**

Dr Isabelle Guérin (IRD, CESSMA and IFP), Dr Govindan Venkatasubramanian (IFP),  
Dr Sébastien Michiels (IFP), Dr Marc Roesch (CIRAD).

**RUME data management team:**

Dr Sébastien Michiels (IFP), Arnaud Natal (Univ. Bordeaux, IFP).

**Survey coordinator:** Dr Isabelle Guérin ([isabelle.guerin@ird.fr](mailto:isabelle.guerin@ird.fr))





This data collection was carried out within the  
*Observatory of Rural Dynamics and Inequalities in South India.*



<https://odriis.hypotheses.org/>



# Contents

- 1 Introduction of the RUrAl Microfinance and Employment survey** **7**
- 2 Survey methodology** **9**
  - 2.1 Coverage and sampling . . . . . 9
  - 2.2 Quality and reliability of data . . . . . 10
  - 2.3 Ethics . . . . . 11
  - 2.4 Follow up and longitudinal data . . . . . 12
- 3 Files description** **13**
  - 3.1 Content . . . . . 13
  - 3.2 Merging files . . . . . 13
  - 3.3 Data format . . . . . 13
- 4 File content and variables** **15**
  - 4.1 Household database . . . . . 15
  - 4.2 Occupations database . . . . . 23
  - 4.3 Loans database . . . . . 24
  - 4.4 Lenders database . . . . . 25



# 1 | Introduction of the RUrAl Microfinance and Employment survey

THE *RUrAl Microfinance and Employment survey* (RUME), conducted in 2010, is the first wave of a multi-topic household survey covering 405 households and 10 villages in Cuddalore and Viluppuram districts in rural Tamil Nadu (South India). These households provide information for 1928 members. The RUME survey took place within the French Institute of Pondicherry (IFP) and was coordinated by Isabelle Gu erin (IRD, CESSMA), Marc Roesch (IRD-CIRAD) and G. Venkatasubramanian (IFP). The RUME survey was part of a broader programme including quantitative and qualitative data collection tools at IFP (Labour, Finances, and Social Dynamics). It was initiated within a comparative research project funded by the French National Agency for Research (programme: *Les Suds, Aujourd'hui*). The project started in 2007, at a time where microfinance was emerging as a key development policy for alleviating rural poverty. Drawing on a comparative perspective between South-India, Mexico and Malagasy, and intense data collection, both qualitative and quantitative, the RUME project's main objective was to explore the linkages between rural finance and rural employment –including diversification and migration– with a view to contributing to the discussions and interventions in rural development and poverty and vulnerability reduction.

Since 2016, the programme Labour, Finances, and Social Dynamics has merged with another IFP programme –Labour, Skills, Social networks and Mobility in India (LAKSMI)-- to form the Observatory of Rural Dynamics and Inequalities in South India (ODRIIS). In addition to one-off surveys, the ODRIIS's objective is to conduct surveys over time to allow for a longitudinal analysis of the transformations of Tamil rural society. Two other waves of surveys (NEEMSI-1 and NEEMSI-2) have since been conducted.



## 2 | Survey methodology

### 2.1 Coverage and sampling

#### 2.1.1 A focus on a small region to understand and contextualise the complexity of change

The RUME survey is part of the Indian tradition of “village studies”, while being on a regional scale, broader than the village. The RUME survey is not intended to be representative but to complement national data and surveys with an in-depth analysis in a given context. National surveys are obviously crucial for measuring national and regional trends, but the lack of detailed information on institutional contexts and change processes limits their contribution to both understanding change and policy design.

#### 2.1.2 Village selection: common trends and diversity

RUME data collection took place in 10 rural villages in Tamil Nadu, located at the border between Viluppuram and Cuddalore districts, in the area formerly known as “South-Arcot”. Villages include Manappakam, Semakottai, Manamthavizhthaputhur, Natham, Korattore, Karumbur, Oraiyur, Govulapuram, Elamthampattu, Kuvagam.

We chose this region and the 10 villages after a long process of mapping the different Tamil regions. Although they are located in a small area, the 10 villages reflect several dynamics characteristic of the Tamil rural economy and its diversity: a mix of irrigated and dry farming, two nearby industrial towns (Neyveli and Cuddalore) and a regional business centre (Panruti) and varying degrees of remoteness. Villages were selected depending on ecotype systems (half irrigated villages, half dry villages) and accessibility and distance to main roads and small towns (Panruti, Viluppuram, Cuddalore).

Our sample does not claim to be representative but rather to illustrate the diversity of economic and rural dynamics.

### **2.1.3 Household selection: caste as a key factor**

In rural Tamil Nadu, the influence of caste remains crucial, both spatially, economically (with a strong fragmentation of labour markets according to caste), socially (endogamy, making of identities and hierarchies) and politically.

The region shows a high numerical importance of Dalits (former untouchables), who in 2010 represented about half of the village population in this region. In order to be able to compare the processes of change between castes (or jatis, subcastes) and the role of caste in these changes, “middle” and “upper” castes have been overweighted

To choose households, we use the random route sample method: enumerators, by a team of two, interviewed a household every five houses. The survey takes place in each household place of living, and the data collection process takes around two hours.

### **2.1.4 Units of analysis**

The household is our main unit of analysis but some questions are asked at the level of individuals such as employment questions.

To ensure comparability with national surveys, we use the definition of household and head of the household of the Government of India (2011) used in the Census of India.

A household is then a group of persons who usually live together and take their meals from a common kitchen unless the exigencies of work prevent any of them from doing so. Persons in a household may be related or unrelated or a mix of both. The important link in finding out whether it is a household or not is a common kitchen.

While the head of the household is a person who is recognised as such by the household, she or he is generally the person who bears the chief responsibility for managing the household affairs and decides on behalf of the household. In general, it was he who answered the questionnaire.

## **2.2 Quality and reliability of data**

Collecting reliable and quality data is a major concern of the RUME survey. In rural areas with a low level of education, populations have their own visions and understandings of labour, finance, relationships or the State, as anthropology has long shown. In addition, there are the usual biases, well known to statisticians (memory bias, social desirability bias, gender bias, interviewer bias, etc.).

The RUME survey was constructed to strike a balance between categories that make sense to local people and more general categories that are useful for comparison (with other regions of India and abroad). Prior ethnographic work over several years has

provided excellent knowledge of local contexts, terminologies used, units of measurement employed, and the functioning of labour, credit and land markets. A major characteristic of the RUME survey is that the same researchers are involved in the questionnaire and ethnographic surveys. For example, with regard to income, it is a known fact that in contexts of high informality where people often change jobs and combine several jobs, measuring income is a challenge. Here, good knowledge of the different labour markets, prevailing wages and seasonal variations was essential to assess the quality of the answers and to guide people in their questions. Similarly, the reliability of the data on debt comes from a good knowledge of the different sources of debt and the terminologies used (less shameful than the term “debt”) and the prices usually charged.<sup>1</sup>

The questionnaire was designed in English and then translated into Tamil by local researchers specialising in economics, sociology, and development studies. The questionnaire, in paper format, was administered by five local male enumerators (and two fieldwork supervisors) to the household head of each selected household, who answered for all members. In the NEEMSIS waves (2016-17 and 2020-21), this was modified: use of tablets, female interviewers and separate interviews with women on certain sensitive issues.

## 2.3 Ethics

Conducting surveys is time-consuming for the populations interviewed, while the benefits for them are never guaranteed. Beyond the usual ethical rules (explaining the reasons for the survey, asking for consent, allowing people to stop a questionnaire in progress, anonymising data), our team is constantly questioning the ways in which we can conduct our research in an ethical manner. This is done in different ways:

- financial compensation for the families surveyed (at the end of the survey);
- mobilising funds in the event of a serious crisis for the poorest people in the village;
- restitution of the results to the village communities, carefully selecting certain outcomes so as not to harm or create local conflicts;
- as in the case of ethnographic surveys, the creation of long-term reciprocal relationships, based on the exchange of information, advice and friendships;
- using our findings to advance understanding of inequalities and social policy design.

---

<sup>1</sup>This is explained in detail in the book *The Indebted Woman. Kinship, Sexuality and Capitalism* by Isabelle Guérin, Santosh Kumar and Govindan Venkatasubramanian, Stanford University Press, 2023.

## **2.4 Follow up and longitudinal data**

RUME collects data from 405 households and 1928 individuals. Among them, 388 have been surveyed again in 2016-17 in the framework of the NEEMSI-1 survey, and 382 have also been surveyed again in 2020-21 in the framework of the NEEMSI-2 survey. Thus, among the 405 households of RUME, 388 have two points in time (2010 / 2016-17), and 382 have three points in time (2010 / 2016-17 / 2020-21).

## 3 | Files description

### 3.1 Content

RUME household data consists of 4 files:

1. `RUME-HH.dta`
2. `RUME-occupations.dta`
3. `RUME-loans_mainloans.dta`
4. `RUME-lenders.dta`

The first file, `RUME-HH.dta`, is the main file. It includes all household questionnaire variables except occupations, lenders, and loans modules, which are sub-modules and kept in separate `.dta` files since they contain more than one observation by individual. `HHID2010` (household identifier) and `INDID2010` (individual identifier) combined are unique identifiers in this database.

In `RUME-occupations.dta`, `RUME-lenders.dta`, and `RUME-loans_mainloans.dta`, `HHID2010` and `INDID2010` combined are not unique identifiers but `occupationid`, `credittid` and `loanid` are respectively unique identifiers.

### 3.2 Merging files

All files, can be merged together using `HHID2010` and `INDID2010`.

### 3.3 Data format

The third column on the pdf questionnaire shows the lists of choices for categorical and dummy variables. In Stata, dummy variables and single choice categorical variables are numerical with labels. Multiple choice categorical variables are strings. Other variables are either integer or decimal (numerical) or text (string).

. *Missing values.* Values are missing in early questionnaire versions because some questions were addressed later in the survey.

66 *Not applicable.* The question is irrelevant in view of the previous question(s).

77 *Other.* For categorical variables with “Other” option.

88 *Don't know.* The respondent doesn't have the answer to the question.

99 *No response.* The respondent doesn't want to reply.

## 4 | File content and variables

### 4.1 Household database

---

File:	RUME-HH.dta
ID:	HHID2010, INDID2010
Unit:	one observation is one household member

---

Household database is the main database of RUME since it includes most variables of the household questionnaire. One observation corresponds to one household member, that is identified with the combination of *HHID2010* and *INDID2010*. Indeed, *HHID2010* represents the unique identifier of a household in RUME, and *INDID2010* represents, within a household, a unique individual identifier.

Other household questionnaire modules (*RUME-occupations*, *RUME-lenders* and *RUME-loans\_mainloans*) are part of separate databases where *HHID2010* and *INDID2010* are no more the unique identifiers since individuals have several occupations, lenders, and loans. In those sub modules, *occupationid*, *creditid* or *loanid*, combined with *HHID2010* and *INDID2010* are unique identifiers.

Each household belongs to one village *village* and one village area *villagearea* (“Ur” or “Colony”).

#### 4.1.1 General information

*address*. Key variable in order to draw the household migration trajectory at least in three points in time (2010/2016-17/2020-21).

*caste*, *religion*, and *comefrom* identify the caste, the religion and the origin of the household, i.e. since how long the household is in the village.

*numfamily*. Number of household members.

*jatis* is equivalent to *caste* in the questionnaire. We complete this variable with *caste*, which constitutes a simpler variable to analyse castes: 1. *Lowest*, 2. *Middle*, 3. *Upper*.

## Household member module

The respondent is not necessarily the household head in the **Household member module** (*relationshiptohead*).

Each individual has an ID (*INDID2010*) and a name (*name*).

*livinghome* tells us about each members's status in the household. Individuals in the first category *Yes* are considered as living home because they are staying with the household head. Individuals in the second category *No* are considered as not living home because they are not staying with the household head.

*relationshiptohead*. Household questionnaire respondent is not necessarily the household head. The status as "head" doesn't depend on age but is left to the respondent's appreciation: who is considered as the head of this household.

*education*. Level of education of each household member.

*studentpresent*. Identifies whether the individual considered is still student.

*typeeducation*. Complete *education* to understand the type of education.

### 4.1.2 Employment

*dummyworkedpastyear*. Variable repeated for each household member and who is more than 10 years old. It includes unpaid workers, excluding houseworkers. Even though he/she might not be working on the interview date, the answer is *Yes* if she/he worked over the previous year. When the answer is *Yes*, occupation module has data about this individual's occupation(s). See **RUME-occupations** database.

#### Self-employment

These questions exclude farm businesses since there is a module in household questionnaire dealing specifically with agricultural occupations. See **Agriculture** module in **RUME-HH.dta**.

*kindselfemployment*, *yearetablissement*, *businesscastebased*, *businessskill*, and *businessamountinvest* provide basic details on the activity, respectively the type, the year of establishment, the caste-based, or not, the learning and the investment.

**Business source investment** This section provides details about the investment (source and loss) and details about the source: nature of the lender, caste.

**Income** In order to analyse the volatility of income, *goodincomeperiod*, *averageincomeperiod*, and *lowincomeperiod* define the periods when income is respectively good, average and low, with the approximate amount.

**Business labourers** If the business involves several people's labour (*dummybusinesslabourers=1*), module **Business labourers** opens. Labourers were listed as much as possible during the interview and some businesses involve lots of labourers, hence a high number of observations for some occupations (suffix *1* to *7*).

*relationshipbusinesslabourer1-7* variable informs about the nature of the relationship. There are several variables about labourers' socioeconomic characteristics and jobs (caste, type of job and salary, pension, bonus, insurance).

**Sell the product or service on credit basis** If the business involves sell on credit, module **Sell the product or service on credit basis** opens. It aims to analyse the sale of self-employed on credit: the usual period, coping strategies, etc.

**Business in the past** If the individual has done any business in the past (*dummypastbusiness=1*), module **Business in the past** opens to obtain information on the type of business and the reason for the cessation of this business.

**Doing own business** For all individuals above 10 years old, this module provides information about the interest in doing own business. This include, for instance, the type of business and the needs in terms of investment, education, etc.

### **Salaried job**

Wage jobs include a wide range of occupations from daily employment to permanent jobs, piece rate to monthly wages, but they have the common characteristics that the worker is under the supervision of an employer. Questions in this module are also asked to unpaid workers who said they worked over the past year (excepting housewives). This section includes farm and non-farm jobs and the term "salaried" job here doesn't mean we talk about formal employment. Wage job questions are related to the workplace, the employer's characteristics and the terms of the employment.

As workers can have several jobs variables are repeated for each salaried job (prefix from *1* to *4* are used).

### **Problems in your work since 2008 (due to economic crisis)**

As data collection takes place after the economic crisis of 2008, this module aims to analyse the consequences in terms of employment.

### **4.1.3 Migrations**

*dummymigration*. Yes if the household member considered has migrates for work. The job has to be outside the village and the household member stays outside more than

one day. This module provides a range of information such as the type of job, place, distance, travel cost, advance.

#### 4.1.4 Remittances

##### Remittances received

*dummyremreceived*. Yes if at least one household member received money as a remittance over the previous year (household level question).

Remittances questions (source name, relation, amount, etc.) are repeated for each remittance source, which can be identified looking at the variables' prefix *1* to *3* (the data shows maximum three sources).

##### Remittances sent

*dummyremsent*. Yes if the individual considered sent money as a remittance over the previous year (household level question).

Remittances questions (name of recipient, location, frequency, amount) are repeated for household members who have sent remittances and for each remittance recipient, which can be identified looking at the variables' prefix *1* to *3* (the data shows maximum three recipients).

#### 4.1.5 Financial practices

Questions include a wide range of semi-formal and informal financial transactions, beyond formal transactions.

##### Loans and main loans

*dummyloans*. Yes if at least one household member has taken a loan during the year 2009 (household level question).

For information on loans and main loans, see the database [RUME-loans\\_mainloans](#).

##### Lenders

See the database [RUME-lenders](#).

##### Lending money

*dummylendingmoney*. Yes if the individual considered lent money over the previous year (individual level question).

Then variables are duplicated with suffixes *1* to *4* as individual can lend to several individuals.

### **Recommendation given**

The module starts with *dummyrecommendgiven* which identifies whether a household member has recommended someone for a loan in the past year.

As a household may have recommended several individuals the variables are repeated using the suffixes *1* to *2*.

### **Recommendation received**

The module starts with *dummyrecommendreceived* which identifies whether a household member received recommendation for a loan in the past year.

### **Gold**

This module identifies the quantity of gold held and the quantity pledged.

### **Chit funds, savings, insurance**

Each of these three modules starts with a dummy variable to know if at least one household member belongs to a chit fund (*dummychitfund*), has a bank account (*dummysavingaccount*), or has a formal insurance (*dummyinsurance*).

The second variable identifies whether the individual under consideration is concerned (*chitfundbelongerdummy*, *savingsownerdummy*, *insuranceownerdummy*).

Members can belong to several chit funds, have several bank accounts or several insurances. In that case, variables are duplicated with suffixes *1* to *2* for chit funds, *1* to *3* for bank account, and *1* to *2* for insurances, in order to have the same information for each chit fund, bank account and insurance.

## **4.1.6 Agriculture**

Agriculture module is household level questions (in **RUME-HH**, same values will appear for all household members).

### **Land**

A sets of questions, each starting with a dummy, are asked:

- Land at present: *ownland*, *leaseland*;
- Land purchase and selling last 5 years: *dummylandpurchased*;
- Own land leasing over the last 3 years: *dummyleasedland*;
- Leasing someone else's land over the last 3 years: *dummyleasingland*.

## Cropping

This module opens if the household own or lease land from someone else at present and if the household cultivates at least one crop (*productlist* answer).

*productlist*. List of crops. Variables are repeated using suffixes 1 to 5 as households may have several crops.

Besides incomes from cropping, this module also deals with cultivation costs (production expenses and labour cost).

**Stopped or interested crops** This sub-module identifies whether the household is interested in growing crops or has stopped. Questions are repeated using suffixes 1 to 4 as households may be interested in growing several crops, or has stopped several crops.

**Other products from agriculture** In addition to the previous module **Cropping**, this sub-module identifies other products from agriculture. Variables are repeated using suffixes 1 to 2 as households may have several other crops.

**Labourers** If the agriculture involves several people's labour (*dummylabourers=1*), sub-module **Labourers** opens. The sub-module is at household level and identifies the number of labourers *productworkers*, the total wage *productlabourwage* for instance.

## Livestock

Questions asked no matter if the household owns and leases land. Livestocks can represent an important share of the household income and is significant of the household's wealth.

*dummylivestock*. Have you got livestock?

Livestock variables recoded by kind of livestock using a suffix: *\_cow*, *\_goat*.

To account for the flow of livestock over the last 3 years, RUME has details cattle loss and details cattle sold questions.

## Farm equipment

It contains owned farm equipment.

*equipmentlist*. List of farm equipment. Variables are repeated using suffixes 1 to 2 as households may have several equipments.

#### 4.1.7 Consumption and assets

RUME accounts for consumption practices, consumer durables and goods. All variables are at the household level, meaning that in **RUME-HH**, same values will appear for all household members.

##### Expenses

*foodexpenses*. Food expenses are measured on a weekly basis on average.

*healthexpenses*. Health expenses are recorded on an annual basis.

*ceremonieexpenses*. Festivals expenses include festivals for household members but also festivals outside the household, and are recorded on an annual basis.

*educationexpenses*. Education expenses are measured on a yearly basis.

*deathexpenses*. Death expenses are recorded on an annual basis.

##### Good details

*listgoods*. List of potential consumer durables owned by the household.

Variables are repeated using suffixes *1* to *14* as households may have several consumer durables.

Good property is significant of the household's wealth and public schemes distributing free consumer durables, hence questions about the year of purchase and how the good was bought (credit, free, cash, etc.).

#### 4.1.8 Housing and facilities

This module accounts for the household's housing status (owner, tenant) at the date of the interview, and the house facilities (water, electricity).

#### 4.1.9 Public service work

This module gather informations on public service such as president, ward member, temple committee, school committee, political party designation, etc. These may bring some money as honorarium or no income include SHG leadership, NGO committee.

*dummyserv* identifies whether the individual considered is involved in a public service work.

If yes, details are requested. Variables are repeated using suffixes *1* to *2* as an individual may have several public service works.

#### **4.1.10 Memberships**

This module gather informations on the participation to political public events last year (such as political meeting, trade union activity), and to SHG, association or cooperative.

## 4.2 Occupations database

---

File:	<code>RUME-occupations.dta</code>
ID:	<code>HHID2010, INDID2010, occupationid</code>
Unit:	one observation is one occupation

---

This module gathers information about all occupations during the past year, i.e. the twelve months before the interview date. For each occupation, the interviewees declared the type of occupation (*kindofwork*), the name in details (*occupationname*), an estimated annual income (*annualincome*), and whether or not the individual has stopped this activity (*stopworking*).

The observation unit is the occupation identifier (*occupationid*).

## 4.3 Loans database

---

File:	RUME-loans_mainloans.dta
ID:	HHID2010, INDID2010, loanid
Unit:	one observation is one loan

---

Two sets of variables are available to document the borrowing practices of the households (**Loans** and **Main loans** modules). Both sets are combined in **RUME-loans\_mainloans** database.

Several questions were then asked for each loan: *loanamount*, *loanlender*, *loanreasongiven*, etc. These variables are in **RUME-loans\_mainloans**.

Each line corresponds to one loan, identified by the variable *loanid*.

In a second stage, three loans per household were selected among the pool of all loans. Denominated in the questionnaire as “main loans”, there are not necessarily the largest loans. Households had to select the loans whose repayments were regarded as the most crucial, with the constraint that one of these main loans had to be a woman’s one if such loan existed in the household. For these three loans per household, additional questions (*borrowerservices*, *interestloan* or *guarantee* for instance) were asked and are included in **RUME-loans\_mainloans**. The corresponding variables are recorded on the lines of the loans chosen as main loans, and are consequently missing for the other loans.

## 4.4 Lenders database

---

File:	<code>RUME-lenders.dta</code>
ID:	<code>HHID2010, creditid</code>
Unit:	one observation is one lender

---

This database list all the money lenders for the past 5 years. Each line corresponds to one money lender, identified by the variable `creditid`.

This database provides information on the type of money lender (`mltype`), the frequency or the strengths and weaknesses of each money lenders.

