HAL open science

NEEMSIS-1 Survey Manual

Christophe Jalil Nordman, Isabelle Guérin, Govindan Venkatasubramanian, Sébastien Michiels, Youna Lanos, Santosh Kumar, A. Antony Raj, Anne Hilger, Arnaud Natal

To cite this version:

Christophe Jalil Nordman, Isabelle Guérin, Govindan Venkatasubramanian, Sébastien Michiels, Youna Lanos, et al.. NEEMSIS-1 Survey Manual. [Technical Report] French Institute of Pondicherry; IRD - Institut de recherche pour le developpement. 2017, 42 p. hal-03751302v3

HAL Id: hal-03751302

https://hal.science/hal-03751302v3

Submitted on 21 Oct 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 Attribution 4.0 International License

NEEMSIS

NETWORK, EMPLOYMENT, DEBT, MOBILITIES AND SKILLS IN INDIA SURVEY

https://neemsis.hypotheses.org

FIRST WAVE

Survey Manual

2016-17

NEEMSIS-1 core team:

Dr Christophe Jalil Nordman (IRD, LEDa-DIAL and IFP), Dr Isabelle Guérin (IRD, CESSMA and IFP), Dr Govindan Venkatasubramanian (IFP), Dr Sébastien Michiels (IFP), Youna Lanos (IRD, IFP), Santosh Kumar (IFP), Antoni Raj (IFP), Anne Hilger (PSE, LEDa-DIAL).

Survey coordinator: Dr Christophe Jalil Nordman (jalil.nordman@ird.fr)

neemsis survey

network, employment, debt, mobility & skills in South India

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

odriis
observatory
rural dynamics &
inequalities in South India

https://odriis.hypotheses.org/

PART I: INTRODUCTION OF THE NETWORKS, EMPLOYMENT, DEBT, MOBILITIES AND SKILLS IN INDIA SURVEY (NEEMSIS) 7

PART II: SURVEY METHODOLOGY 8

2.1. Coverage and sampling 8
2.2. Data collection tool 9
2.3. Breadth of topics and in-depth contextual measures 10

PART III: FILES DESCRIPTION 11

3.1. Content 11

3.1.1. Household data 11
3.1.1. Individual data 11

3.2. Merging files 11
3.3. Data format 11

PART IV: FILE CONTENT AND VARIABLES 13

Household data 13

1. General information 13

1.1. Household member module 13
1.2. "Permanent" migrants job module 14
1.3. "Permanent" migrants employment networks module 14

2. Education 14
3. Employment 15
4. Migrations 15
5. Remittances and gifts 15

5.1. Remittances received 15
5.2. Gifts 15
5.3. Remittances sent 16

6. Financial practices 16

6.1. Loans 17
6.2. Mains loans 17
6.3. Other household incomes 17
6.4. Lending money 17
6.5-9. Recommendation for loans, chit funds, savings, gold, insurance 17

7. Agriculture 19

7.1. Land 19
7.2. Cropping 19
7.3. Livestock 19
7.4. Farm equipment 19

8. Consumption and assets 20

8.1- Expenses 20
8.2- Good details 20
9. Marriage 20
10. Housing and facilities 21
11. Schemes 21
Occupations (Household questionnaire) 22
1. Self-employment 22
1.1. Business investment 22
1.2. Business labourers 23
2. Salaried jobs 23
Migration (Household questionnaire) 24
Loans - Main Loans (Household questionnaire) 26
Ego (Individual questionnaire) 27
1- Employment variables 27
1.1- Labor force participation variables 27
1.2 - Main occupation variables 29
2- Social networks variables 31
2.1- Formal social capital variables 31
2.2 - Informal social capital variables 31
2.3- Other network variables 33
3- Personality and behavior variables 33
4-5- Literacy and numeracy tests and Raven test 33
6- General perception questions about demonetisation policy 33
Alter (Individual questionnaire) 34

PART V: FIELDWORK PROCEDURE 36
5.1. Tasks for Enumerators 36
5.1.1. Working attitude 36
5.1.2. Working schedule 37
5.1.3. Interview principles 37
5.2. Tasks for Team leaders 39
5.2.1. Contact the enumeration area in advance. 39
5.2.2. Overseeing enumeration area 40

PART I: INTRODUCTION OF THE NETWORKS, EMPLOYMENT, DEBT, MOBILITIES AND SKILLS IN INDIA SURVEY (NEEMSIS)

The Networks, Employment, dEbt, Mobilities and Skills in India Survey (NEEMSIS), conducted in 2016-17, is the second wave of a multi-topic survey covering 2696 individuals, 492 households and 10 villages in-between Cuddalore and Villupuram districts in rural Tamil Nadu (South India). Most of these households had been interviewed for the Rural Employment and Microfinance Survey (RUME) conducted in 2010 within the French Institute of Pondicherry, a project led by Isabelle Guérin (IRD, CESSMA), and a team of researchers and enumerators at the IFP (led by G. Venkatasubramanian). Hundred additional households within the same 10 villages have been interviewed for NEEMSIS, thereby constituting a two-year panel of households and individuals.

One of NEEMSIS's specificity is its timing: an external shock, a nation-wide demonetisation policy, announced by Modi government in November 2016, directly affected livelihoods in the course of the survey. NEEMSIS took advantage of this context to try to address its effects on rural households. Three-hour interviews in each household covered employment, migration, education, debt, financial practices, agriculture, marriage and schemes. Two-hour individual interviews in each household covered labour force participation, social networks, cognitive and non-cognitive questions.

NEEMSIS was conducted by a team of DIAL research unit (Développement, Institutions et Mondialisation, DIAL, Paris), under the French National Research Institute for Sustainable Development (IRD), coordinated by Dr. Christophe Jalil Nordman (IRD, DIAL) and Isabelle Guérin (IRD, CESSMA). This project was co-funded by Nopoor, a five year European project (2012-2017), the IRD, and the IMTFI of the University of California. It was implemented within the French Institute of Pondicherry (IFP) in cooperation with the IRD, and is part of the "LAbour, sKills, Social networks and Mobilities in India (LAKSMI)" programme coordinated by C.J. Nordman and the "Labour, Finance and Social Dynamics" programme coordinated by I. Guérin, both constituting Axis 3. "Household Vulnerabilities, Labour and Socioeconomic Dynamics" of the IFP Social Sciences Department research agenda.

NEEMSIS survey aims at understanding the linkages between household and individuals' labour, skills, social networks and social and spatial mobilities. This includes the investigation of various forces at play, spanning from the role of social structure (norms and institutions), the development and use of social networks, to the formation of cognitive and non-cognitive skills.

PART II: SURVEY METHODOLOGY

2.1. Coverage and sampling

NEEMSIS consists of 492 households and 2696 individuals. Out of 492 households, 388 households a are RUME survey recovered households (**1.1.1**) and 104 are new households (**1.1.2.**). NEEMSIS includes a new data collection unit: the individual level (**1.1.3.**). Part of the sample was surveyed after the demonetisation policy announced in November 2016 (**1.1.4.**).

2.1.1. Panel data. A longitudinal study is particularly interesting in the moving context of rural India which experiences both fast changes, but also keeps some rigid social structure. One objective of this quantitative survey was thus to recover all the 2010 households, including households who are seasonally migrating for work and who moved to other close cities.

The sampling method follows that of the 2010 RUME survey. RUME and NEEMSIS take place in 10 rural villages in Tamil Nadu, located at the border between Villupuram and Cuddalore districts (Manappakam, Semakottai, Manamthavizhthaputhur, Natham, Korattore, Karumbur, Oraiyur, Govulapuram, Elamthampattu, Kuvagam). Villages were selected depending on land characteristics - half are irrigated villages, the other half have dry lands- and in terms of accessibility and distance to main roads and to small towns (Panruti, Villupuram, Cuddalore) in the area.
Inside villages, caste is the main criterion to choose households: since Dalits (lower castes) are more numerous than middle and upper castes households, more upper castes and middle castes households have been selected in order to better observe cross castes dynamics. Therefore, the number of households selected per caste is not exactly proportional to their actual number in the villages.

NEEMSIS survey is conducted on the same RUME 405 households in order to obtain panel data and undertake longitudinal analysis. 388 households and 2168 individuals were recovered, i.e. *4.2 percent attrition rate*. While most households could still be found in their 2010 houses, part of them "migrate" seasonally for work and part of them have even migrated permanently to their work place. Enumerators have followed a "tracking" methodology: meeting labour intermediaries ("maistries"), finding employers and the migration place and being allowed by employers to interview these households at their workplace. Most of them are usually accomodated around brick kilns industries in Chennai surroundings (Chengalpattu).
NEEMSIS has focused on recovering migrant households and couldn't afford following individual migrants, i.e. individuals who moved from their original residential place in RUME survey area between the two survey waves. However, NEEMSIS assessed all 2010 household members and basic information about individual migrants are available (migration place, reason, length, etc).

2.1.2. Refresher sample. In addition to the 388 households feeding a panel dataset, 10 households by village were added to the final sample in order to increase the sampled population. Since NEEMSIS attrition rate of the panel households between the two waves is low, new families were not selected in order to have the same characteristics as attritioned households. In each village, five households were selected in "Colony" area where mostly lower castes live, and five households in "UR" where middle and upper castes live.

Households in each village area were randomly chosen: every five houses, enumerators asked to answer the questionnaires. In one village, Oraiyur 4 more households were interviewed (by mistake) but we kept them (3 in "Colony", 1 in "UR"). In total the new sample consists of 528 individuals.

2.1.3. An additional survey unit: the individual level. The household questionnaire is answered by one household member, usually the head, about all household members so we have information each household member for all these modules. Only agriculture, consumption and some schemes are household level questions. NEEMSIS added a new survey unit: the individual level or "Egos" level. Two household members are directly addressed individual questionnaires: the respondent of the household questionnaire (called “EGO 1”) and one younger household member (“EGO 2”). NEEMSIS has 953 "Egos".
See Individual data section

2.1.4. Pre/post demonetisation samples. One third of households, 142 households and 732 individuals, had been interviewed after the demonetisation policy announced in November 2016, between mid-January 2017 and early May 2017.

2.1.5 Overview of NEEMSIS sample. The following table summarizes NEEMSIS household and individual total sample and sub samples:

	Pre demonetisation			Post demonetisation			Total		
	HH	Individuals	Egos	HH	Individuals	Egos	HH	Individuals	Egos
Panel	317	1808	622	71	360	133	388	2168	755
New	33	156	57	71	372	141	104	528	198
Total	350	1964	679	142	732	274	492	2696	953

dummydemonetisation. Dummy variable that takes 1 if the household was interviewed after January 2017.
dummynewHH. Dummy variable that takes one if the household is a new 2016-2017 selected household, i.e. not a RUME-NEEMSIS panel household.

It is important to point out that priority was given to panel households all along the survey, hence the high share of panel households interviewed in the pre-demonetisation sample compared to the post-demonetisation sample. New households were selected in a second stage.

2.2. Data collection tool

NEEMSIS used tablets for data collection, relying on the Survey CTO software. This tool allowed to increase the quality of the data collected, because it is meant to check quality at each stage of the data entry process (missing observations, constraints on answers), and also to reduce the cost, time and errors associated with data entry as this is done instantaneously on the field.

2.3. Breadth of topics and in-depth contextual measures

2.3.1. Household questionnaire. NEEMSIS focuses on all aspects of rural life, including in-depth modules on the households' financial practices, migrations, remittances, labour, agricultural activities. NEEMSIS questionnaire includes all 2010 RUME household questionnaire modules on employment, migration and remittances, financial practices (loans, savings, lending practices, gold), agriculture, consumption and housing. NEEMSIS kept the same variables in order to observe their variation between the two waves but has also supplemented these modules with new questions in order to have better measures and economic indicators. For instance, occupation module has more detailed questions about business outputs and costs to better calculate business profits. It also includes more variables to better define the degree of informality of occupations, etc.
New modules have also been added in the household questionnaire about individual migrations, education, marriage and government schemes. 2010 questionnaire had a few questions about education but NEEMSIS used them to implement a full module on education since measuring networks and skills is one of NEEMSIS's main objective. Marriage module is more complete as well: marriage is indeed one of the most important networks' component. Finally, a "schemes" module was brought in the household questionnaire: public schemes represent an important share of rural households income, especially in Tamil Nadu where social welfare has expanded a lot over the last decades.

2.3.2. Individual questionnaires. Individual questionnaires provide a range of information on labour force participation and outcomes (including wages and earnings), social networks related to community involvement, interpersonal employment networks, perception of community level employment networks and cognitive and non-cognitive skills assessments (numeracy, literacy and Raven's test; personality and behaviour questions).

2.3.3. "Demonetisation variables". In December 2016, NEEMSIS team worked at improving the questionnaires by adding questions in order to tackle the impact of the demonetisation policy, announced by Modi government in early November 2016, on labour, financial practices, and networks formation and usage. These variables have the prefix **demo-** and concern only households interviewed after January 2017, i.e. one third of the final household sample.

PART III: FILES DESCRIPTION

3.1. Content

3.1.1. Household data

NEEMSIS household questionnaire consists of 4 files:

- NEEMSIS-HH.dta
- NEEMSIS-occupations.dta
- NEEMSIS-migrations.dta
- NEEMSIS-loans_mainloans.dta

The first file is the main file that includes all household questionnaire variables except occupations, migrations, remittances and loans modules, which are sub-modules and kept in separate dta files since they contain more than one observation by individual. HHID (household identifier) and INDID (individual identifier) combined are not unique identifiers in occupations, migrations, remittances and loans sub-modules but occupationid, migrationjobid and loanid are respectively unique identifiers.

3.1.1. Individual data

NEEMSIS individual questionnaire includes:

- NEEMSIS-ego.dta
- NEEMSIS-alters.dta

NEEMSIS-HH, NEEMSIS-ego and NEEMSIS-alter have the same core IDs: *villageID, villageareaID, HHID, INDID, egoid*

3.2. Merging files

All files can be merged together using *HHID* and *INDID*.
Egoid informs if household members are Ego 1 or Ego 2 or none of them.

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 (lists). 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). Irrelevant questions are skipped on the tablet (used to appear as "missing", recoded as "66").
"77" Other
For categorical variables with "Other" option, new categories were added when needed without deleting information contained in the next question "If "Other", please specifiy".
"88" Don't know. The respondent doesn't have the answer to the question.
"99" No response. The respondent doesn't want to reply.

Calculated variables. Either calculation, or summary variables of two variables. Calculated variables don't appear in the first column of the pdf questionnaire. ID and name variables visible in the second column of pdf questionnaire (ex: *migrantid migrantname*) and in italic in **Part IV** disappeared with merging of modules into the household database. ID and name variables equal to **INDID** and **name.**

In Part IV, names in blue and bold correspond to group of questions in the pdf questionnaire.

PART IV: FILE CONTENT AND VARIABLES

Household data

File: NEEMSIS-HH.dta

ID: villageid, villageareaid, HHID, INDID, egoid

Household database is the main database of NEEMSIS since it includes most variables of the household questionnaire. One observation corresponds to one household member, identified with **HHID** and **INDID**. Other household questionnaire modules (occupations, migrations and loans) are part of separate databases where **HHID** and **INDID** are no more the unique identifiers since individuals have several occupations, migration activities, and loans. In those sub modules, **occupationid, migrationjobid** or **loanid**, combined with **HHID** and **INDID** are unique identifiers.

Household identification. Panel households have the same village name and village area as in 2010 ("Colony"/"UR") even if their address (and the interview place) has changed. **householdid** summarizes 2010 household information: household head name, address and household identifier (**HHID2010**). **householdid** can be "new household" for households selected within the 10 villages in 2016 for the refresher sample (**dummynewHH**).

1. General information

dummydemonetisation. Dummy variable that identifies the sample affected by the shock.
interviewplace. Variable informs about the interview place only for the post-demonetisation sample.
address. Key variable in order to i) identify the interview place of pre-demonetisation households not informed in **interviewplace** (households address is also the interview place) ii) draw the household migration trajectory at least in two points in time (2010/2016).

caste and **religion** are the same as in 2010, but **comefrom** tells about the household questionnaire respondent's birth place. The respondend is not necessarily the household head in the Household member module (**relationshiptohead**).

numfamily. Number of household members: includes all 2010 household members even though they don't live at home anymore or died since 2010, but also new household members who have joined the household between the two waves (children, in-laws, parents, other family members, etc).

1.1. Household member module

All 2010 and new 2016-2017 household members are listed in this module. Each have an ID (**INDID**) and **name.**

livinghome. Key variable of NEEMSIS since tells us about each members status in the household, especially what 2010 household members became and about the situation of newcomers. Individuals in the first category "Yes, share meals or/and income" are considered as living home because they are coming back home daily and sharing meals. The household reference place is where the 2010 elder

and still alive household members live in 2016-17. It can be the same place as in 2010, the household migration place, or any new address. NEEMSIS attempts to quantify individual migration using two other categories for **livinghome**: "temporary" migration and "permanent" migration.
Temporary migrants' home is the same as other household members, they mostly left home temporarily for work or studies and share systematically their income with household members who stayed back home. Their income is pooled with other household members' income. For that reason, they are still considered as living home.
"Permanent" migrants, on the other hand, got married and settled in order to start a new household or have left home for more than one year and don't plan to come back living home in the short run. Their income is not pooled with other household members' income. Permanent migrants information are not asked in Occupation and Migration module but in General Information module only ("Permanent" migrants job). Their monetary contribution to the household are accounted as remittances.

lefthomedurationlessoneyear. **lefthomedurationmoreoneyear**. In months if left less than one year ago ("temporary" migrants), in years if left more than one year ago ("permanent" migrants).

lefthomedestination. lefthomereason. Individual migrants', whether temporary or permanent), destination and reason for migrating.

relationshiptohead. Household questionnaire respondents is not necessarily the household head. The head might be different from 2010 head (for instance if 2010 head died or is much older). 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.

1.2. "Permanent" migrants job module

Questions asked only if is a "permanent" migrant (**livinghome**=3) and if the household respondent says he/she has been working over the past 12 months (dummypermanentmigrantwork=1).
Several occupations for this migrant can be recorded (**permanentmigrantoccupnber** and **permanentmigrantoccupname**)

1.3. "Permanent" migrants employment networks module

Module opens only if the "permanent" migrant found this occupation through its social network (**permanentmigrantfindjob**=1, 2 or 3). He/she may have found this work through a relative, a friend or other kind of interpersonnal network.
All variable with prefix **pmsnfindjob-** provide with information about characteristics of the migrants network. Throughout this module one has to keep in mind that the household respondent's knowledge about the migrant's life is partial and incomplete and knowing more about individual migrants would require direct interviews.

2. Education

Education questions are repeated for each household member, except 2010 household members who died and children below 5 years old are excluded. If individuals have ever attended school (**everattendedschool**=1), detailed questions follows about education level and the kind of education. If they haven't, we ask why (**reasonneverattendedschool**). If they are not currently at school (**currentlyatschool**=0), we also ask why (**reasondropping**); if they are (**currentlyatschool**=1), we want to know about their education expenses (total, school fees, materials, transportation).

3. Employment

dummyworkedpastyear. Variable repeated for each household member who is living at home (livinghome=1 or 2) 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 Occupations database.

stoppedworking. If hasn't worked over the previous year, informs if it is permanent because of health issues/accident. Variable existing in the first wave and particularly relevant for NEEMSIS context.

4. Migrations

dummymigration. "Yes" if at least one household member currently migrates for work (household level question). The job has to be outside the village and the household member stays outside more than one day. If the answer is "Yes", **migrantlist** gives the list of migrants' ID corresponding to **INDID** *(migrantid migrantname).* The list of potential migrants excludes "permanent" migrants since Household members and Permanent migrant job already provided this information. It includes only members living home/temporary migrating (**livinghome**=1 or 2).

See Migration database for detailed variables about migration activities of household members listed in **migrantlist**.

5. Remittances and gifts

5.1. Remittances received

dummyremreceived. "Yes" if at least one household member received money as a remittance over the previous year (household level question).
remreceivedlist. List of household members' IDs, corresponding to **INDID (***remittancesreceivedid remittancesreceivedname*), who received remittances over the previous year.

Remittances questions are repeated for household members who have received remittances and for each remittance source, which can be identified looking at the variables' prefix 1, 2 (the data shows maximum two sources), replacing previous ID *remreceivedsourceid.*

remreceivedhhsource1-2 informs about the source of the remittance: outside the household ("Someone else" and name specified in **remreceivedothersource1-2**)) or a "permanent" migrant household member (${name+"household member's position in the roster"}).
remreceivedsourcename1-2. Summary variable of **remreceivedhhsource1-2** and **remreceivedothersource1-2.** Name of the sender of the remittance.
If the sender is an household member (**livinghome**=3), **remreceivedsourcerelation1-2** and **remreceivedsourceplace1-2** are skipped since this information are already provided in General information. Other variables apply for all kind of senders.

Three variables are related to the effect of demonetisation: **demoremreceived1-2, demoremreceivedamount1-2, demoremreceivedform1-2**.

5.2. Gifts

The aim is to account for informal ways of securing savings besides Gold and Marriage gifts, through other festivals/occasions. Gifts included hare are monetary and non-monetary. Since this module was added after the demonetisation policy, we can not compare gifts received by pre and post demonetisation samples and give evidence about this policy's impact on interpersonal exchanges/solidarity. However this module provides additional information about household members interpersonal network.

dummygiftsreceived. "Yes" if at least one household member received gifts since demonetisation, i.e. after November 2016 (household level question).
giftsrecipientlist. List of household members' IDs, corresponding to **INDID (***giftsrecipientid giftsrecipientname*), who received gifts.

giftsource. List of type of gift senders (employer, friends, relatives living outside household, etc). Gifts sender ID corresponds to *giftsourceid* in the household questionnaire but *giftsourceid* was removed from the database and gifts variables recoded with the type of sender as a suffix: **_WKP, _rel, _emp, _friends**, respectively well known people, relatives, employer and friends.
Depending on the relation with the gift sender, variables include: the occasion when the gift was offered (**giftoccasion_**), the kind of gift (**gifttype_**), their number (**giftsourcenb_**) and total monetary value (**giftamount_**).

Only 6 percent of households said that at least one household member received gifts (**dummygiftsreceived**=1). Gifts received are most probably under declared because: i) respondents might be reluctant to provide information about gifts received and about the person who offered the gift, ii) respondents might be reluctant to provide for this kind of information in front of other household members. iii) respondents don't know about the gifts received by other household members.

5.3. Remittances sent

dummyremsent. "Yes" if at least one household member sent money as a remittance over the previous year (household level question).
remsenderlist. List of household members' IDs, corresponding to **INDID (***remsenderid remsendername*), who sent remittances over the previous year.

Remittances questions are repeated for household members who have sent remittances and for each remittance recipient, which can be identified looking at the variables' prefix 1, 2 (the data shows maximum two recipients), replacing previous ID *remsentrecipientid.*
remsenthhrecipient1-2 informs about the recipient of the remittance: outside the household ("Someone else" and name specified in **remsentotherrecipient**)) or a "permanent" migrant household member (${name+"household member's position in the roster"}).
remsentrecipientname1-2. Summary variable of **remsenthhrecipient1-2** and **remsentotherrecipient1-2 .** Name of the recipient of the remittance.
If the recipient is an household member (**livinghome**=3), **remsentrelation1-2** and **remsentplace1-2** are skipped since this information are already provided in General information. Other variables apply for all kind of recipients.
Three variables are related to the effect of demonetisation: **demoremsentamount1-2, demoremsentform1-2.**

6. Financial practices

Most of financial practices questions build on 2010 RUME survey. They include a wide range of semi-formal and informal financial transactions, beyond formal transactions.

6.1. Loans

dummyloans. "Yes" if at least one household member has taken a loan since January 2015 or still has outstanding loans taken before January 2015 (household level question). In RUME survey, similarly, loans taken from January 2009 were recorded, plus outstanding loans taken before.
borrowerlist. List of household members' IDs, corresponding to **INDID (***borrowerid borrowername*), who took loans.
nbloansbyborrower. Number of loans contracted for each borrower.

See loans variables in Loans-Main Loans database.

sumhhloans. Calculate variable: sum of **nbloansbyborrower**. Total number of loans contracted in the household over the period defined above.
loandetails1-80. Catalog of all members' loans in Loans-Main Loans database. Variable built in order to select the three household main loans.

6.2. Mains loans

threemainloans. See Loans-Main Loans database.

6.3. Other household incomes

This module asks for other household incomes outside work, land, housing and schemes. It includes incomes from interest, dividends share sale or capital gains (**dummyincomeassets, incomeassets**).
We assumed that demonetisation policy has encouraged people to diversify even more their income sources, hence new variables: **demootherinc, demootherincamount, demoexchangeasked, demoexchangeaccepted, democommissionamount.**

6.4. Lending money

dummylendingmoney. "Yes" if at least one household member lent money over the previous year (household level question).
hhlenderlist. List of household members' IDs, corresponding to **INDID (***lenderid lendername*), who lent money over the previous year.

Variables were added after demonetisation policy: **datelendingmoney** (to make the difference between loans contracted before/after demonetisation), **demolendingkind, demotermslending, demodummyrepaylending, demorepaytermslending, demointerestlending**

6.5-9. Recommendation for loans, chit funds, savings, gold, insurance

Each of these five modules start with a dummy variable to know if at least one household member recommended someone (**dummyrecommendgiven**), belongs to a chit fund (**dummychitfund**), has a

bank account (**dummysavingaccount**), has gold (**dummygold**) or has an formal insurance (**dummyinsurance**).
The second variable lists household members concerned by the module using their **INDID**: **recommendgivenlist, chitfundbelongerlist, savingsownerlist, goldownerlist, insuranceownerlist.**

INDID and **name** are found under the following names in the household questionnaire:
- *recommendgivenid recommendgivenname*
- *chitfundbelongerid chitfundbelongername*
- *savingsownerid savingsownername*
- *goldownerid goldownername*
- *insuranceownerid insuranceownername*

Members can belong to several chit funds, have several bank accounts or several insurances (**nbchitfunds, nbsavingaccounts, nbinsurance**). In that case, variables are duplicated with suffixes 1 to 4 in order to have the same information for each chit fund, bank account and insurance.

Some variables were added after demonetisation:
- in Savings module:
i) to account for the number of bank accounts opened and applications for cards after demonetisation: **savingsaccountdate1-3, datedebitcard1-4, datecreditcard1**
ii) to address the impact of debit card and credit card use after demonetisation: **demousedebitcard1-2, demousecreditcard1-2**
- in Gold module to account for the use of gold as a safety net in a context of external shock: **demogoldquanti, demogoldreasonbuy, demogoldpledged, demogoldreasonpledge**

7. Agriculture

Agriculture questions mostly build on 2010 RUME survey. Agriculture module are household level questions (in **NEEMSIS-HH**, same values will appear for all household members).

7.1. Land

dummyeverhadland. Answer is "Yes" if the household owns or leases land from someone else now or/and over the last 5 years. In that case, following sets of questions, each starting with a dummy, are asked if the dummy is 1:

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

7.2. Cropping

This module opens if the household own or lease land from someone else at present (**dummyeverhadland**=1) and if the household cultivates at least one crop (**productlist** answer is not "None")

productlist. List of crops. Crops' ID and name are *productid* and *productname* in the household questionnaire but *productid* and *productname* were removed from the database and crops variables recoded by type of crop using a suffix: **_paddy, _ragi, _millets, _tapioca, _cotton, _sugarca, _savukku, _guave, _groundnut**.

Besides incomes from cropping, this module also deals with cultivation costs (production cost and labour cost). Since Occupation module didn't ask those questions when occupation is "Agricultural activity on own household farm" (**kindofwork**=1), NEEMSIS focuses on them here.

Three variables are related to the effect of demonetisation: **demonbagriworkers, demoagriactivity, demoagrikindofjob.**

7.3. Livestock

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

livestocklist. List of animals. Livestock' ID and name are *livestocktid* and *livestockname* in the household questionnaire but *livestockid* and *livestockname* were removed from the database and livestock variables recoded by kind of livestock using a suffix: **_cow, _goat, _chicken, _bullock.**

To account for the flow of livestock over the last 3 years, NEEMSIS has **Details cattle loss** and **Details cattle sold** questions.

7.4. Farm equipment

It contains owned and borrowed farm equipment.

equipmentlist. equipmentborrowedlist. List of farm equipment. Equipment' ID and name are *equipmentid, equipmentid2* and *equipmentname, equipmentname2* in the household questionnaire but were removed from the database and equipment variables recoded by kind of equipment using a suffix: **_tractor, _bullockcart _ploughmach.** Equipment owned variables have the prefix **equiown-** and equipment borrowed variables the prefix **equilent-.**

8. Consumption and assets

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

8.1- Expenses

Food expenses are measured on a weekly basis on average. Health, festivals and death expenses are recorded on a annual basis. Festivals expenses include festivals for household members but also festivals outside household, and exclude marriage expenses accounted for in Marriage.
The effect of demonetisation policy on consumption is measured through several variables: **demoexpenses, democonsoless, democonsomore, democonsosame, democonsopractices, democonsoplace.**

8.2- Good details

listgoods. List of potential consumer durables owned by the household. Goods' ID and name, *goodid* and *goodname* in the household questionnaire, were removed from the database and consumer durables variables recoded by kind of consumer durables using a suffix: **_car, _cookgas, _computer, _antenna, _bike, _fridge, _furniture, _tailormach, _phone, _landline, _DVD, _camera.**
Goods 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).

9. Marriage

dummymarriage. "Yes" if at least one household member got married over the last six year (household level question). The objective is to know who got married in the household since 2010 RUME survey.
marriedlist. List of household members' IDs, corresponding to **INDID (***marriedid marriedname*), who got married. This list includes household members who left home (**livinghome**=3) after marriage (mainly daughters, but also sons and daughters who started a new household elsewhere), who joined the household after marriage (mainly daugthers-in-law) or who stayed in the household even after marriage (mainly sons).

Marriage module consists of three sets of questions:
- Marriage cost: **marriage dowry, marriagetotalcost, marriageexpense**
- Marriage finance: **howpaymarriage**
- **Marriage loans details**

marriageloansource. List of lenders for marriage. Question asked if the marriage was at least party financed with loans (**howpaymarriage**=1 or 3). Lender's ID and name (*marriagesourceid* and *marriagesourcename* in the household questionnaire) were removed from the database and recoded by kind of lender using a suffix: **_wellknown, _banks, _coopbanks, _relatives, _employer, _maistry, _colleagues, _pawnbroker, _monlender, _friends.**
Those prefix are added on **marriageloannb** and **marriageloanamount** variables.
- Marriage gifts.
marriagegiftsource. List of gift senders for marriage. Gift sender's ID and name (*marriagegiftsourceid* and *marriagegiftsourcename* in the household questionnaire) were removed from the database and recoded by kind of gift sender using a suffix: **_wellknown, _shg, _relatives, _employer, _maistry, _colleagues, _shopkeeper, _friends.**

Those prefix are added on **marriagegiftnb** and **marriagegifttype marriagegiftamount** and **marriagegoldamount** variables.

10. Housing and facilities

This module accounts for the household's housing status (owner, tenant) at the date of the interview and over the past 5 years (since the first wave of the survey), and the house facilities (water, electricity).

11. Schemes

Most of rural households in Tamil Nadu benefit from public schemes and pensions: only 2 percent of RUME/NEEMSIS area is not benefiting at all from any of the schemes listed in **schemeslist**.

Schemes and pensions" ID and name are *schemeid* and *schemename* in the household questionnaire but were removed from the database and schemes variables were recoded by kind of schemes and pension using a suffix: **_freemachine, _cashmarriage, _goldmarriage, _cashfunerals, _animals, _freegas, _educ, _farmequi.**
These prefixes add on following variables: **schemerecipient, schemeyear, schemeamount, pensionrecipient, pensionamount.**
Ration card, housing and land schemes variables have the prefix **rationcard-, housingscheme-, landscheme-**.
For NREGS (National Rural Employment Guarantee Schemes), **nreganberdaysworked** and **nregaincome** give values for each household members' listed in **nregarecipientlist.**
One variable accounts for the effect of demonetisation on schemes: **demoscheme.**

Occupations (Household questionnaire)

File: NEEMSIS-occupations.dta

ID: HHID, INDID, occupationid, businessloanid, businesslabourerid

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**), an estimated number of days a month **daysamonth** and hours a day worked **hoursaday** (the number of hours a year calculated **hoursayear**) and an estimated annual income **annualincome**, the date they took up the job and the impact of demonetisation policy on each occupation (post-demonetisation sample). It includes unpaid workers. Individuals who haven't worked the past year answer questions about the reasons they haven't worked (**reasonnotworkpastyear, stoppedworking**). However these individuals may have worked before. The Individual questionnaire provides further information about labour force participation (see Ego database and **NEEMSIS-ego.dta).**
demooccup and **datestartoccup** aim at assessing the effect of demonetisation on occupations in general.

The observation unit is the occupation identifier (**occupationid**), the business kind of loan (**businessloanid**) and the business labourers identifier (**businesslabourerid**). In other words, **businessloanid** and **businesslabourerid** are in long within **occupationid** and there can be multiple observations by occupation for self-business but not for other kind of jobs (**kindofwork**).

kindofwork. This variable is the main variable of this module. Depending its value, specific questions are asked about self-business and wage jobs:

> **1** "Agricultural activity on own household farm". See 7.2.Cropping for details about own agricultural activity (**NEEMSIS-HH.dta**)
> **2** "Self-employed, own account worker" > **1. Self-employment**
> **3** "Agricultural salaried job", **4** "Non agricultural salaried job", **5** "Unpaid worker in family business (non-agricultural)", **6** "Unpaid worker in other business (non-agricultural)",
7 "Unpaid worker in own farm" or **8** " Unpaid worker in another farm" > **2. Salaried job**

1. Self-employment

These questions exclude farm businesses since there is a module in household questionnaire dealing specifically with agricultural occupations. See 7.2.Cropping and **NEEMSIS-HH.dta.**

1.1. Business investment

businesssourceinvestment. For self-employed, we asked for sources of investment into their business. Options 1, 2 and 3 are loans (loans from relatives, bank loans, informal loans). The other options are own capital, savings or the absence of investment sources.

businessloanid. businessloanname Refers to the kind of business loans: relatives, bank/formal loans and informal loans.

numberbusinessloan. Whenever options 1, 2, 3 are selected in **businesssourceinvest** variable, we ask the number of each type of loan.

Following variables provide more information about the business loan lender and can be duplicated up to 3 times since one business can have more than one loan from each type of lender. For instance, a business can have two informal loans from two different lenders (a finance company and a friend). These variables are **namebusinesslender1-3, addressbusinesslender1-3, relationbusinesslender1-3, castebusinesslender1-3, occupbusinesslender1-3.**

1.2. Business labourers

If the business involves several people's labour (**dummybusinesslabourers=1**), **nbbusinesslabourers** gives the number of labourers and **businesslabourerid** is each business labourer identifier. 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.

dummyhhmember. Business labourer is or not a household member. If the answer is "Yes", following variables are skipped because we already have information about this person in household information and occupation modules. This variable was added after the pilot survey, hence missing values. For these cases, **relationshipbusinesslabourer** variable informs if it is a household member or not. There are several variables about labourers' socioeconomic characteristics and job's (address, relation, caste, type of job and salary).
businesslabourerdate. This variable added after demonetisation shock in order to know if labourers joined the business after demonetisation.

Five other variables are related to the effect of demonetisation on business: **demobusinessloss, demobusiness, demobusinessactivity, demobusinesskindofjob, demobusinessactivityother**

2. Salaried jobs

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 include farm and non-farm jobs and the term "salaried" job here doesn't mean we talk about formal employment.
Wage jobs questions are related to the workplace, the employer's characteristics and the terms of the employment.

casteemployer. Other castes categories added after the survey
17 "Nayar" **18** "Odayar" **19** "Parvatha kulam" **20** "Raja kulam" **21** "Yadhavar" **22** "Malayali"
Wage workers may have had employers from different castes for the same occupation. When several castes, we don't know which caste was mentioned first by the respondent.

Three other variables are related to the effect of demonetisation on wage jobs: **datestartoccup, demojobtype, demowage, demowagetype, demosalariedjob, demosalariedjobother**

Migration (Household questionnaire)

File: NEEMSIS-migration.dta

ID: HHID, INDID, migrationjobid

The list of household migrants for work (**migrantlist)** is given in **NEEMSIS-HH.dta** (4.1. Migration) when at least one household member migrates for work, except "permanent" migrants (**dummymigration**).
Several questions were then asked about migrants' migration jobs. These variables are in **NEEMSIS-migration.dta**. Each line corresponds to one kind of migration job (brick kiln, sugarcane cutting, farm and non-farm coolie, construction work, work in a private company, self-employment), identified by the variable **migrationjobid (**and *migrationjoblist* in pdf questionnaire**)**. This variable counts migration jobs at the level of the individual identified by **INDID** and **name** *(migrantid borrowername).*
Migration jobs recorded in this module are also supposed to appear in Occupation database as occupation of respective household members (**INDID**), except for children below 10 years old.

migrationtype. Individual or familial migration. Familial means with household members listed in the Household members module. It doesn't include family members outside the household roster.

dummyadvance. Seasonal advance on migrants' salary that they have to pay back, at best at the end of the season (can take many years to repay the employer).

> If the migrant received an advance in this migration job, several questions are asked: **migrationadvanceamount, dummyadvancebalance.**
If the advance is still in balance (**dummyadvancebalance=1),** we asked about its amount, potential issues faced by the migrant (**advanceamountbalance, advancebalanceproblem, advancebalanceperception**).

Following variables are related to the effect of demonetisation on advance: **demopreviousadvance, demoadvancetransfer, demoadvancevar**
demopreviousadvance. Missing values if migrants didn't get any advance or who got an advance but were interviewed before demonetisation
demoadvancevar. Impact of demonetisation on the advance and if migrants accepted it or declined it. The reasons why migrants accepted or declined this change of the advance are asked according to the answer to **demoadvancevar: demoacceptadvanceinc** if "1", **demorefusedadvanceinc** if "2", **demoperceptionadvancedec** if "3".

> If the migrant didn't received any advance in this migration job, additional questions were asked only after demonetisation: **demodeclineadvance, demoadvancereasonrefused.**

Missing values concern only pre demonetisation respondents in this database, there no missing values for post demonetisation respondents.

migrationjobtype. Degree of stability of migration jobs but permanent and fixed term jobs doesn't mean signed contract (often implicit contract/oral agreement)

migrationsalary. Migration season's total wage including potential allowances on the migration place but excluding the advance. It should match with **annualincome** variable in Occupation database.
migrationpension, migrationbonus, migrationinsurance. Should similarly match with **salariedjobpension, salariedjobbonus, salariedjobinsurance** in Occupation database.

migrationmainoccup. Workers' perception about the importance of this migration job compared to other jobs, if any over the last year.

Loans - Main Loans (Household questionnaire)

File: NEEMSIS-loans_mainloans.dta

IDs: HHID, INDID, loanid

Two sets of variables are available to document the borrowing practices of the households (6.1. Loans and 6.2.Main loans modules). Both sets are combined in **NEEMSIS-loans_mainloans.dta**.

In a first stage, the list of borrowers and the number of loans contracted (**borrowerlist** and **nbloansbyborrower**) are recorded in **NEEMSIS-HH.dta** (6.1. Loans).
Several questions were then asked for each loan contracted by each member. These variables are in **NEEMSIS-loans_mainloans.dta**. Each line corresponds to one loan, identified by the variable **loanid**. This variable counts the loans at the level of the individual identified by **INDID** and **name** *(borrowerid borrowername).* For instance, if one individual has contracted three loans, the corresponding variables will thus be recorded on three lines, with **loanid** taking respectively the value 1, 2, 3 for each line.
Variables: **loanamount** to **loaninfo**

Three variables are related to the effect of demonetisation, and were asked only for the loans contracted after 2016-11-01 (**loandate**): **demoloanreason, demoSHG, demoSHGother.**
loandate. Added after demonetisation as well to make the difference between loans contracted before/after demonetisation.
loaninfo. Concanates three variables for each loan: the borrower's name (**name**), the amount of the loan (**loanamount**) and the reason for borrowing (**loanreasongiven**). Thanks to this variable, all members' loans are cataloged in **NEEMSIS-HH.dta** (**loandetails1-80**, see 6.1. Loans)

In a second stage, three loans per household (**threemainloans**) were selected among the pool of all members' loans (**loandetails1-80**) in **NEEMSIS-HH.dta** (6.2.Main 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 were asked and are included in **NEEMSIS-loans_mainloans.dta**. The corresponding variables are recorded on the lines of the loans chosen as main loans (ID and names are **loanid loaninfo** instead of *mainloanid mainloanname*), and are consequently missing for the other loans.
Variables: **lendersex** to **guaranteetype**.
Three variables are related to the effect of demonetisation, and were asked only for the loans contracted after 2016-11-01: **demodummyrepay, demorepayterms, demointerest.**

loanreasongiven. In pilot stage of NEEMSIS, this variable was in multiple choice and it was changed later to a single choice variable. In only 16 cases, several options were selected in the pilot versions, hence two other variables: loanreasongiven2, loanreasongiven3
It corresponds to the reason borrowers gave to the lender to borrow money
loaneffectivereason. How the borrower actually used this money once he/she gets it. The reason can differ from the reason initially given and can be multiple.

Ego (Individual questionnaire)

File: NEEMSIS-ego.dta

ID: villageid, villageareaid, HHID, INDID, egoid

NEEMSIS includes one additional unit of analysis, the individual, using a specific questionnaire *explicitly addressed to two household members*. The two household members interviewed with individual questionnaires are called '*egos'*.
Ego 1 is the household questionnaire respondent, except in few instances where he/she can't answer the individual questionnaire (in 2 percent cases according to **dummyhhrespondentego1** variable).
Ego 2 is randomly selected on a criterion of age since NEEMSIS aims at collecting more detailed information about young people. A young member between 18 and 25 in the household roster is first selected and interviewed if available. If there is no such individual in this age range or this person is not available, we look for another young member between 26 and 35. Once again, if there is no one in that age range or not available to answer, NEEMSIS interview a household member above 35.

NEEMSIS has data for 953 *egos*. There are two different individuals IDs in this database: **INDID** refers to the respondent position in the household roster (same ID used in the household database) and **egoid** informs us if the respondent is Ego 1 or Ego 2.

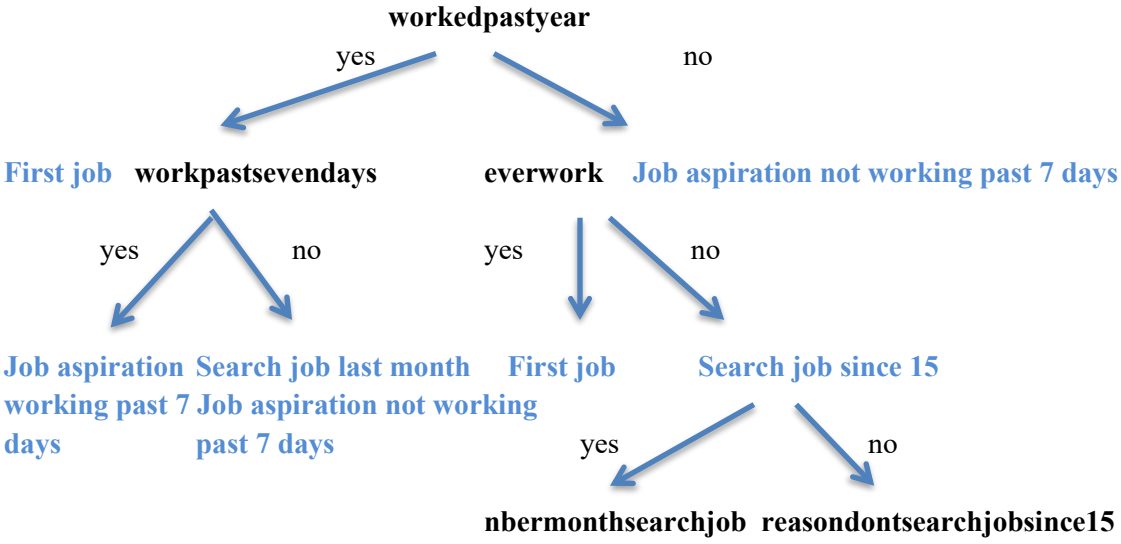
NEEMSIS addresses exactly the same questions to Ego 1 and Ego 2. A diversity of approaches and tools were used to design the individual questionnaire: it includes modules on labour force participation and professional aspiration (1), formal and informal social network modules (2) and cognitive skills and personality traits inspired by psychology and cognitive sciences (3).

1- Employment variables

1.1- Labor force participation variables

Labor force participation variables complement household questionnaire employment module. While household questionnaire employment questions focus on employment over the previous year, the individual questionnaire goes further into the interviewee's employment history (previous jobs, first job, method for searching jobs) and asks about job aspirations.

workedpastyear. In the household questionnaire this dummy variable reports if household members have worked or not during the year before the survey. Labor force participation questions depend on this variable.



everwork. If ego respondent hasn't worked over the past year, this variable captures if she/he has ever worked, including unpaid jobs and excluding housework.

First job. Addressed to Egos who declared working the previous year or who have ever worked (cf above). Includes unpaid first jobs except housework.
Respondents whose first job was unpaid (**kindofworkfirstjob**>=5) are also asked the age when they started their first paid job (**agestartworkingpaidjob)** in order to account for the time between the entrance in the labor force and the access to paid work. However, part of respondents may never have had access to paid work such as those who remained unpaid workers or/and have stopped working. So this variable should be interpreted with caution and crossed with other labour force participation variables.
"66", "99" values correspond to "not applicable" and "no response"

Search job last month. Respondents have worked the previous year but not the past seven days. They are questioned about job search/business initiative over the last month and the reason why they didn't work.

Search job since 15. Respondents have never worked, including as unpaid workers but they can be doing housework. They are asked about job search/ attempt to set up a business since 15 years old and the reason why they couldn't work.

Job aspirations not working past 7 days. Respondents have already worked before but not the last 7 days or have never worked. Questions about availability to work and under what conditions and their job preferences.
Job aspirations working past 7 days. Accounts for the gap between current job conditions and outcomes (salary, number of hours worked) and workers' aspirations.

1.2 - Main occupation variables

Details about occupations are now asked directly to Ego 1 and Ego 2 whereas only one household member (usually the head) addressed occupation questions about each household member in the household questionnaire (**NEEMSIS-occupations)**.

Following modules are consequently more precise and detailed. They focus on the respondent's "main occupation" when he/she is reported as working over the previous year in the household questionnaire (**workedpastyear**=1). For this purpose and using household questionnaire information, NEEMSIS defined "main occupation" as the respondent's most time consuming job over the past year, i.e the occupation with the maximum of hours a year (**maxhoursayear**) since many interviewees have more than one occupation.

hhmainoccupID. Calculated variable. Occupation ID of Ego's most time consuming job over the past year
hhmainoccupname. Calculated variable. Occupation name of Ego's most time consuming job over the past year

hhmainoccupID and **hhmainoccupname** correspond to **occupationid** and **occupationname** in **NEEMSIS-occupations** database.

Since that definition of "main occupation" may not be satisfactory, NEEMSIS asks the respondent if this calculated main occupation is really his/her main occupation (**dummymainoccup**) and in case it is not, the respondent informs about which occupation he/she feels is the main one (**othermainoccup**).

mainoccup. Summary variable of **occupationname** and **othermainoccup.**

dummyregularmainoccup. Perception variable. Regular in the sense a more or less fixed number of days a month over the year and hours a day. In the household questionnaire (**NEEMSIS-occupations file**), **daysamonth** and **hoursaday** variables gives the *average* number of days a month and hours a day spent in an occupation but don't provide any information about the variability of days worked depending on months and of hours worked depending on days.

inddaysayear2. Effective working days a year if dummyregularmainoccup=1
indhoursaday2. Effective working hours a day if dummyregularmainoccup=1
indhoursayear2. Calculated variable. Effective working hours a year (**inddaysayear2* indhoursaday2)**

Main occupation schedule over the past year.

If the respondent feels it is not regular (if **dummyregularmainoccup=0)**, he/she can make a calendar of this "main occupation" over the previous year
The initial objective of this module was 1/ to have a more detailed variables than **monthsayear** (**NEEMSIS-occupations file**) that could inform about the months worked during the year in order to capture the jobs seasonality (**calendar_months**). 2/ to have the number of days a month (**inddaysamonth1)** and the number of hours a day (**indhoursaday1**) worked over the previous year *depending on each month*.

indhoursamonth1. Calculated variable: inddaysamonth1* indhoursaday1
indhoursayear1. Calculated variable. Sum of indhoursamonth1

However only 6 Egos (or 0.7 percent of Ego sample) declared their "main occupation" is not regular so we decided to this module from the database, which consisted of **inddaysamonth1, indhoursaday1** and **indhoursamonth1** variables. We kept **indhoursayear1** in order to build the summary variable:

indhoursayear. Summary variable of indhoursayear1 and indhoursayear2

Characteristics of main occupation in the past 12 months

beforemainoccup. Kind of job of the previous "main occupation". What is defined as a main occupation is left here to the appreciation of the respondent. Depending on the kind of job of the current main occupation (**mainoccuptype**), more details are asked about the previous and the current main occupation:

- If the current main occupation is "self-employed" (**mainoccuptype**=2):

- dummypreviouswagejob, previousjobcontract, reasonstoppedwagejob

- Self-employment as [**mainoccup**], including:
  - questions about unpaid (including household members) and employed workers in the business when it was set up and on the date of the interview.
  - **Details about paid business workers** (number of paid workers, frequency of payments, total labour cost)
  - **Details about business income and expenses** (gross receipt, fixed costs, production costs, social security)
  - Details payment in kind

These questions exclude agriculture on own/lease land (**mainoccuptype**=1) since there is a module in household questionnaire dealing specifically with agricultural occupations (**NEEMSIS-HH file**)

- If the current main occupation is a salaried job or an unpaid job (**mainoccuptype**>=3):

- Wage job as [**mainoccup**], including:
  - employers' status
  - **Details terms of wage job** (size employing structure, contract, job type, social security)
  - Details salary (net payment)
  - Details payment in kind

Most questions target only Egos who are salaried as a main occupation (**mainoccuptype**=3 ou **mainoccuptype**=4). A few variables also include unpaid workers (**mainoccuptype**>=5) such as **wagejobnbworkers** (only unpaid worker outside own farm or business), **wagejobpaymentinkind, wagejobpaymentinkindlist** and **Details payment in kind.**

2- Social networks variables

2.1- Formal social capital variables

NEEMSIS measures *formal social capital* (section 2.1. individual questionnaire), i.e. individuals' involvement into public institutions, such as associations, political parties, trade unions, Self-Help Groups, youth unions, farmer unions, village councils (panchayat), religious groups, etc. (**associationlist).** Variables in this module are numbered from 1 to 3 correponding to the first, second and third association/institution Egos belong to (not to be confused with values in **associationlist**). In other words, Egos have maximum 3 associations.

associationid1-3.. The value in data corresponds to the kind of association/institution selected in **associationlist.**
associationname1-3. Name of association/institution. For instance the name of the political party.

Further details are asked about each association**,** such as the degree of participation to that association, its size, the role of that association to find a job or to support Ego's business.

dummyassorecommendation1-3. This variable opens the name generator module if the answer is "Yes". See next section about this variable.

2.2 - Informal social capital variables

In NEEMSIS, *informal social capital* data collection follows the name generator sociological method (section 2.2. individual questionnaire). Name generators consist of several questions designed to invite respondents (i.e. Egos) to recall and elicit people ("Alters") with whom they maintain certain types of direct relationships, to delineate the core members of their network (Marsden, 2005).

In order to identify Ego's relationships, NEEMSIS' individual questionnaire has chosen the criterion of interdependency or regular interaction among individuals involved in social relations, thus conveying resources needed in different dimensions of labour activities but also conveying resources needed because of the demonetisation policy (Campbell and Lee, 1991, Crossley et al., 2015). The criterion chosen has the advantage of being clear and unequivocal, as it is less likely to be interpreted differently across respondents and we neither *a priori* define nor exclude specific social categories. The network composition, effective and potential, is the consequence of the nature of the network support that we observe.

In **"NEEMSIS-ego",** the following questions open the name generator module:

1) Has someone recommended you to join this association? **dummyassorecommendation1-3**
2) If you need to borrow tools/machinery for your business, would you ask somebody you are directly in contact with? **sntechnicalhelp**
3) If you have to go away from your business a couple of days, would you entrust your business to someone you are directly in contact with? **snentrustbusiness**
4) If you need to recruit a worker, would you ask someone you are directly in contact with to help you to find him/her ? **snrecruitworker**

5) If you need to find new suppliers or customers, would you ask someone you are directly in contact with to help you to find them ? **snfindsuppliers**
6) Did you lend tools or machinery within the past 12 months from someone you are directly in contact with ? **snlendtools**
7) If you needed to find a new job, is there someone you are directly in contact with you would feel comfortable to ask for help? **snfindjob**
8) In the last 12 months, did you recommend someone you are directly in contact with for a job opening that you knew of ? snrecommendforjob
9) Was anybody you are directly in contact with hired based on your recommendation ? **snrecojobsuccess**
10) Did you ask help to someone due to demonetisation ? **sndemoasked**
11) Did you provide help to someone due to demonetisation ? **sndemogiven**

Among Egos, who is addressed these questions ?

2) 3) 4) 5) 6) only target Egos whose "main occupation" (**mainoccuptype** in previous module) is self-employed or agriculture on own/lease land.
1) 7) 8) 9) 10) 11) are asked to all Egos, no matter of their main occupation and if they have a job or not.
10) 11) concern only households interviewed after January 2017, i.e. around a third of the final household sample.

Effective and potential networks

1) 6) 8) 9) 10) 11) invite Egos to mention their effective network, i.e. people from whom they have effectively asked help or helped.
2) 3) 4) 5) 7) concerns potential network

When the answer to any of these questions is "Yes", Egos are invited to elicit "Alters" (household members are supposed to be excluded). Variables ending with the suffix **-namelist** account for Alters.

- ➔ If **-namelist** value is "74" and Ego is interviewed before demonetisation (**dummydemonetisation**=0) or **-namelist** value is "88" and Ego is interviewed after demonetisation (**dummydemonetisation**=1), it means that Ego mentions one or several "New person" (or new Alters) for the corresponding network.
- ➔ If **-namelist** displays another value, it means that Egos selected an Alter he/she already mentioned in a previous network name generator question.

Since NEEMSIS was using tablets for data collection, this coding prevents the respondent from giving once again all details about an Alter when he/she already give them previously.

One Alter can thus be mentioned several times (for several reasons) across social network modules. Since **-namelist** variables are not directly usable and intuitive for a data user, a specific database deals with Alters data: see Alter data and **NEEMSIS-alters.dta.**

2.3- Other network variables

NEEMSIS also includes following network information:

- *network size* (section 2.3. individual questionnaire): **nbcontactphone nbpersonevent**

- *resources of network* (section 2.4. individual questionnaire): contact with people who occupy specific positions or are leaders of large organizations and whom Ego can contact easily (**contactlist nbcontact dummycontactleaders contactleaders)**

- *network perception* (section 2.5. individual questionnaire) :
-trust in the community:
trustemployees, trustneighborhood, networkpeoplehelping, networkhelpkinmember & the impact of demonetisation policy on trust variables (demotrustemployees, demotrustneighborhood, demonetworkpeoplehelping, demonetworkhelpkinmember)
- trust in institutions after demonetisation: **demotrustbank**

Variables with prefix **demo-** concern only households interviewed after January 2017, i.e. around a third of the final household sample.

3- Personality and behavior variables

The non-cognitive skills assessment relies on the "Big Five Inventory" questionnaire (41 questions), and further includes the socio-emotional skill 'grit' (the tendency to sustain interest in long term goals and persistence). The Big Five factor model is usually attributed to Allport and Odbert (1936). The Big Five taxonomy has since been replicated across cultures (John and Srivastava, 1999) and developmental stages of the life course (Soto et al., 2008).

As far as possible these questions were addressed to the respondent alone with less people around. When it wasn't possible to interview Ego alone (relatives in the house), enumerators warned them not to influence or comment Ego's answers.

4-5- Literacy and numeracy tests and Raven test

The cognitive skills assessment includes the '*Raven Coloured Progressive Matrices*', which have been previously used in cognitive skills assessment in low-literacy populations in developing countries, a literacy test (four questions), and a numeracy test (four questions).

Raven's Coloured Progressive Matrices are a nonverbal test used to measure abstract reasoning. It consists of identifying the missing element from a set of answer possibilities that logically completes a pattern (Raven, Raven, Court, 1956).

6- General perception questions about demonetisation policy

Alter (Individual questionnaire)

File: NEEMSIS-alter.dta

ID: villageID, villageareaID, HHID, INDID, egoid, alterid

egoid corresponds to the same variable in **NEEMSIS-ego.** Ego (Ego 1 or Ego 2) can mention several Alters referred as **alterid**. This database consists of 508 observations, i.e 508 Alters. Whenever Ego chose to declare a "new person "(see **-namelist** variables) in **NEEMSIS-ego,** this person is filed with **alterid** and **altername** in **NEEMSIS-alter.**

Alters can be mentioned several times across social network modules, hence a new variable we constructed here (**networkpurpose1-5**) that is more intuitive than **-namelist** variables. For instance, **networkpurpose1** is the network name generator question that have seen Alter mentioned by Ego for the first time (see questions 1 to 11 in 3- social network variables). **networkpurpose2** is the second reason why Alter is mentioned, etc.

alterid_byNP1. Alternative ID variable for Alter which is unique for each value of **networkpurpose1** (*HHID, INDID, egoid, networkpurpose1)* instead of being unique for each Ego (*HHID, INDID, egoid)*.

nb_ntwother. Constructed variable that indicates the number of times Alter is mentioned *again* (for other network purpose). "0" value means that Alter is mentioned only once across name generator modules. Alter is mentioned five times maximum in NEEMSIS data.

Name generators (**alterid** and **altername**) are followed by questions called "name interpreters" that gather information on Alters' attributes such as caste, age, gender, occupation, education, place of residence, economic situation, and about the relationship between the Egos and Alters (when and how they met, how often they interact, their level of intimacy).

All variables in green in the word file individual questionnaire are name interpreters variables. They have the following suffixes (green) that correspond to following variables (black) in **NEEMSIS-alter.dta**:

-name: altername
-relationship: relationship
-sex: sex
-age: age
-castes: caste
-educ: education
-occup: occupation
-occupother: occupother
-living: place
-compared: ecocondition
-duration: durationknown
-meet: howmet

-meetother: howmetother
-meetfrequency: meetfrequency
-invite: egoinvite
-reciprocity1: alterinvite
-intimacy: intimacy

Alter database lso includes the following variables:

associationid. Specifies the kind of association (**associationlist** in **NEEMSIS-ego**) for which Ego has received a recommandation in order to enter that association/institution. Provides an additional information about Alter who recommended.

Two variables, added after demonetisation policy, contain information about the kind of help and counterpart provided between connected Egos and Alters (**demohelpasked,**
democounterpartasked).

PART V: FIELDWORK PROCEDURE

5.1. Tasks for Enumerators

The enumerator is the one who directly communicates and collects information from households (Household and EGO 1 individual questionnaire) and another individual in the household (EGO 2). The role of enumerator, therefore, is critically important for the data set's quality and integrity. NEEMSIS enumerators' team is composed of 8 members: two fieldwork supervisors, Venkatasubramanian, Santosh Kumar and six enumerators: Antoni Raj, Sithanandam, Pazhani, Annamalai, Mayan and Vivek. The supervisors and five enumerators took part to the 2010 RUME survey data collection so they already have an excellent knowledge of the fieldwork.

Before conducting an interview, enumerators need to verify the list of households interviewed in 2010, their address, caste and basic characteristics in order to make sure they interview the same households. In the identification module of the questionnaire, the tablet asks for the village name and the village area ("Colony"/"UR") where the interview takes place. According to that preliminary selection, the list of households surveyed in 2010, with their 2010 household identifier displays. Enumerators then can start meeting available households in the list. The General information module allows to check if the household characteristics match with 2010 information. If enumerators meet a new household, i.e outside the 2010 list, he can choose the "new household" option in the household ID list, and then fill the new household information in the General Information module.

Enumerators sometimes have to meet a household several times either to complete the household questionnaire, Ego 1 individual questionnaire or to meet the second household member (Ego 2) depending on their availability. Enumerators, therefore, have to actively plan for appointment confirmation with each household and "Egos", who should be interviewed separately, and make use of time in communicating with them in order to meet the fieldwork schedule and fulfill the number of questionnaires assigned.

The six enumerators working on NEEMSIS survey were trained on the questionnaire and the tablet before the beginning of the pilot survey and the team regularly planned meeting with them in order to improve the questionnaire and clarify questions raised by the enumerators during fieldwork and thus to collect high quality data.

5.1.1. Working attitude

- Enumerator should maintain polite attitude with everyone (the respondents, their family and their friends, the team leader, the survey team members and other involved people). Interviewer's behaviour greatly influences local people's opinion as well as the survey results.
- Enumerator should be well-dressed (neat, polite), showing professional style. Dressing style recommended is:
    - Male: shirt, trousers
    - Avoid fancy dress, creating distance from the interviewees.
    - NO alcohol/ beer or smoking during work
- Enumerator should create a friendly atmosphere to encourage communication and information sharing from the respondent. Enumerator, at the mean time, still has to comply with interview procedure and questionnaire contents as required in Section IV.

- During the interview, the enumerator should not make any judgment, comment on respondent's answers or behaviors that might affect respondent's psychology and impose enumerator's self opinions.

5.1.2. Working schedule

- In the initial stage of the survey, enumerators are divided into 3 teams. Each team consists of 2 members, including at least one of them who already has fieldwork experience (5 out of 6 enumerators were already enumerators for the 2010 wave). Teams of two members enable them to get used to the new survey tool: with one tablet per team, one can focus more on the technical part and the other one can keep track of the interview. After the initial stage, enumerators will start to go alone to interview households, i.e. they will have one tablet each.
- The team and then enumerators alone will complete 1 household plus the two individual questionnaires (when both individuals are available) within a day or two days. Once they will be interviewing alone, they will complete 2010 RUME panel households available in one village within 10 days.
- Once all 2010 panel households in the 10 villages are completed, the enumerators will start tracking migrant households and to complete the 100 new households in the 10 villages.
- Enumerators, after receiving the survey plan, should understand the enumeration areas and the list of households, and should meet and interview households or establish contacts prior to the interview to ensure to meet targeted respondents and obtain the best quality of interview.
- Enumerators should strictly follow working procedure and schedule assigned. Interview and data input should follow process taught in training course. Enumerators should ensure no overlap or missed interview of any respondent and no missing or wrong information recorded on the tablet.
- Enumerators are requested to finalize questionnaires everyday on the tablet and to do fieldwork report to the coordinating team twice a week. The coordinating team can then send and save the finalized questionnaires on the Survey CTO server.

Important notice:

- Remember to "save" the file and data whenever "Save the form" note appears on the tablet and after completing the questionnaire to avoid data loss.
- Take note carefully during interview if there are additional elements not included in the questionnaire.
- Check completed questionnaire, ensure the information's integrity before finalizing it.
- Maintain and preserve survey tools and Tablet PC, avoiding loss or misplacement.
- Report any incident and make a request to coordinating team, especially in situations that the enumerator cannot solve himself/herself..
- Collect valid invoices and document as required.

5.1.3. Interview principles

Enumerator should carefully follow instruction in this manual, details as follows:

(i) Ask the question exactly as stated in the questionnaire. The questions were carefully designed to obtain the correct information for later analysis, pre-tested many times at the enumeration area and carefully translated in Tamil on the tablet so that there is no need for in live translation on the field.

Enumerator should read the full text of the question. After the first clearly reading out loud, enumerator should wait for an answer. If the respondent does not provide an answer within a certain period of time, they might: 1) not hear the question clearly; or 2) not understand the question; or 3) do not know how to answer. For all cases, an enumerator should repeat the question. If the respondent is still not able to answer the question, enumerator should ask whether respondent understands the question or not. If not, the enumerator should explain the question's content in another way.
In principle, you will almost never read the list of the coded answers to the respondent even if he or she has difficulties to answer because very often he/she will agree with your proposal. Nevertheless, you must do it in some cases that this handbook will specify (for some "questions" ending with "...", the possible answers have to be enumerated as examples).

(ii). Trying not to receive "I don't know" answer by helping the respondent to estimate or answer or find the answer most closely resembling their point. In case figures are required (income, value of purchase…) and the respondent seems to be reluctant to answer, ask for a range of value (about how much?), but without suggesting any amount. In case basic information is known, fill in the right information into the blank. If information is unclear or forecasted, enumerator should ask the respondent.

(iii). If there are really no answer, the enumerator can choose "No response" option or "99" if numerical variable (when don't want to give any answer), "Don't know" or "88" (when don't know a specific information about a household member for instance), or "Not applicable" or "66" (appears rarely because irrelevant questions are usually automatically skipped). The distinction between these three options is crucial for data analysis later.

(iv). Maintain the interview' pace. Enumerator has to administer the interview, yet listen carefully to respondent, and avoid displeasing the respondent. To do so, enumerator should refrain from lengthy discussion with the respondent; if respondents answer is inappropriate or too complex, enumerator should not interrupt suddenly but cleverly get respondent back towards the original question. Enumerator should avoid suggesting answer to respondent.

(v). Keep a completely neutral attitude regarding the topic of the interview. Enumerator should not express their surprise, agreement or disagreement with the given answers. If the respondent seeks for an answer, the enumerator should not reveal his/her opinion, only provide an explanation of interview purpose is to gather the opinions of the respondents on this issue. Enumerator should not discuss his/her views with the respondent or hint an answer until the interview is completed.

(vi). If you don't understand any question or process, read this manual first, then ask for more detail if necessary. The enumerator should inform the coordinating team about these concerns who wil gather this information in the report.

(vii). Make sure that the respondent would not think that you are an inspector (local tax authority, customs organization…)
A good way to make the interviewee more confident could be: "the information collected here is strictly confidential, will be used only for research purpose and won't be transmitted to government administration, according to the law of confidentiality of statistical surveys". Indeed, we ask questions about forms of labour, indebtedness, which can be very sensitive for households, that's why he/she needs to be sure that it is not for police control or finance companies.

(viii). Interview's Confidentiality.

All collected data is kept confidential. Any of information revealed to un-incharged person is considered as violation to interview's regulation. This principle is very important and fundamental to all statistical work.

In principle, the presence of a third person can cause embarrassment and affect answers, confidentiality of information. For that reason, the interview should preferably take place inside the house. However, it is often difficult to prevent the presence of a third person during the interview. In such cases, the enumerator should suggest the respondent to persuade the third person to go elsewhere; or explain the importance of interview's privacy and confidentiality in a proper manner. For individual questionnaires particularly, it is better to be alone (as much as possible) with the respondent because of personality and behaviour questions and cognitive tests (Raven's test). On the contrary, for the household questionnaire, the presence of other household members can be profitable since the enumerators can obtain more precise information about all members.

(ix). Answer sincerely to the questions of the respondent

Before accepting to be interviewed, the respondent can ask you some questions about the survey or the way he/she was selected to be interviewed. Be direct and friendly when you are answering. The respondent can also be worried by the duration of the survey. Tell him/her about the average duration of NEEMSIS interview, which is about half a day to one day for household plus the two individual questionnaires, and do not try to get it shorter. Always have on you your identity cards and your badge from the French Institute of Pondicherry (IFP) and do not hesitate to present on a simple request of your interlocutors.

5.2. Tasks for Team leaders

Out of 6 enumerators, one is considered as a team leader, Antoni Raj, responsible for managing team's activities. The team leader, therefore, besides acts as an enumerator, should have leading and communicating skills with other team members. This team leader works closely with two other team leaders, who are not enumerators but who have strong experience of fieldwork in the NEEMSIS area.

5.2.1. Contact the enumeration area in advance.

The team leaders are responsible for contacting households, and in some cases head of villages or village administrative officers (VAO) to confirm the list of households to be interviewed and make working appointments for the team.

There are cases where the household address is not found, nobody at home, household relocated, new household residents in the address. The team leaders should ask their contacts to households' neighbors or relatives and track these households if possible. Otherwise, fieldwork team leaders should contact with coordinating team for final decision on the sample.

Notice for team leaders to communicate with the village administrative officers (VAO):

- *Survey's purpose:* Collecting updated data in each of the 10 villages, so as to inform for policy makers to formulate policy adjustments about labour and more generally rural livelihoods.
- *Team survey composition:* number of members, task assigned.
- *Schedule:* number of panel households to interview, including migrants, plus 10 new households by village (around 20 days by village over 6 months)

- *Survey procedure:* Interview duration is from 3 to 4 hours for the household questionnaire and around 2 hours for individual quesionnaires depending on households.
- Announce survey schedule to the village administrative officer (VAO) and request for cooperation

5.2.2. Overseeing enumeration area

- Team leaders play the role of enumeration area managers, closely communicates and works with enumerators to prevent possible errors in the early days, encourage the team to work at their best.
- Team leaders should summarize, share the working experience among the team and report to coordinating team.
- Re-check in detail all filled questionnaire forms for data fulfillment and integrity.
- For any missing or errors found, team leaders and coordinating team should discuss with enumerators to find solutions. The team has to come back to household if neccessary.