Thailand Vietnam Socio Economic Panel

Data Cleaning Guidelines

2017 (7th WAVE)

Rattiya S. Lippe

Somkid Naprom

Mai Thanh Tu

PREFACE

The purpose of socio economic household surveys is to obtain correct information from respondents in the sample in order for researchers to be able to answer socially relevant questions. Researchers use the data to develop and test empirical models and hereby contribute to a better understanding of people's decisions, their strategies, their success and failures, their attitudes, their experience and their beliefs. Having this information is key to advising policy makers about which policies will work and which one will not.

Undoubtedly, the collection of household panel data in rural areas of the developing countries is a challenging task. It requires patience, participation and commitment by respondents, enumerators and supervisors. Interviews take place under conditions which can be stressful: rain, heat and storms, dogs who bark and dogs who want to bite, mosquitos, snakes and other threats. Also, enumerators must wake up early and sleep late. They sometimes must travel long distance to reach their interview places in tight mini-buses, with the risk of traffic accidents. Usually a rural household survey lasts a month during which time enumerators are away from home and must stay in tight living quarters, eat simple food, do not have drinks for fun and so on. Enumerators must be disciplined because every day they have to ask the same questions again and again using the same text from the questionnaire. In a panel survey enumerators are often confronted with respondents who are asking them: why do you come again? Didn't you ask me that already last year? What do I get from answering all these questions?

We must have great respect for the work of the enumerators. But like all of us, no matter of our level of education or rank, as human beings we make mistakes. Error is human! Although we can learn from our mistakes, we are likely to make new ones. With the growing adoption of tablet computers in conducting interviews, non- sampling errors on the one hand can be reduced, for example, through in-built plausibility checks. On the other hand, too many plausibility rules can make the tablet program slow and thus disturb the interview environment which can lead to new errors. Also, electronic questionnaires increase the probability that enumerators make entries in the wrong column when touching the screen.

This prelude should say enough why data which were collected aren't just ready for use! The so-called "raw data" must be cleaned to make them ready for use by the researchers. This is not to say however that raw data are like raw food that must first be "cooked" before it can be consumed. Rather, raw data must undergo a rigorous assessment procedure that detects inconsistencies in the answers and implausible values for quantitative information and finds reasonable replacements for missing values. This can be a tedious job again. Data cleaning usually starts a few weeks after the survey and it is normally carried out by staff who was not or not continuously involved in data collection. Data cleaners thus are often confronted with many puzzles that can arise from the answers and they have to make decisions about plausibility. Did the enumerator pick the wrong unit? Is the entry a code or a number? Does the code which was chosen by the enumerator match with the answer to the question before? And so on and so forth. Although general statistical rules for identifying outliers exist, like the "Two- Standard- Deviation-Rule",

there is a need for more specific rules and project specific guidelines when dealing with complex rural household surveys in developing countries. Such guidelines are best developed together with those who were involved in the survey and have experience with past cleaning work of the panel data.

The guidelines presented in this document is an example of project-specific guidelines. The TVSEP Database Manager together with the National Research Data Collection Managers from Thailand and Vietnam have taken up this task and have combined their past experience from cleaning the TVSEP data. I am convinced that this document is a major step forward towards making the cleaning process more efficient and more transparent. On over 100 pages the TVSEP managers have come up with an excellent blend of general data cleaning rules and formal STATA commands. The use of practical examples that stem from past data cleaning experience and the up-to-data country-specific plausibility values are extremely useful, not only for the data cleaning teams in the two TVSEP countries, but they will also be useful for the data users who get some advice on the interpretations of the variables in the panel.

I congratulate the authors for having put together so many bits and pieces and details into a coherent and easily accessible document. I am convinced that their investment will have high returns in terms of faster data cleaning and less frequent questions to answer from the data users who are the TVSEP's major customers.

Hermann Waibel   Hannover, July 2018

CONTENTS

LIST OF TABLES

I. General principles for data cleaning

1. Use the paper-based questionnaire, the cleaning guidelines, the do-files previously provided and the notes from survey to think about possible connections between variables and sections when checking.

2. Please follow the cleaning steps in the guidelines and the Master do-file. This will facilitate the national team leader and the database manager to cross-check do-files between sections and countries.

3. The variable names and their labels must be consistent with the TVSEP database and the household questionnaire. Please do not rename it unless the variable name and its label are not consistent with the questionnaire.

4. Please do not rename the data and do-files (see Table 1).

5. Auxiliary variables, those are generated during the cleaning process, shall be dropped at the end, unless they are needed for the next step.

6. Please pay attention to codes 90, 98 and 99. These could sometimes be reported values and not codes.

7. Check carefully whether zeros and missing observations are reported zero values or are non-applicable questions.

8. For Vietnam, most of the monetary values are in the unit "1000 VND", please be careful with number of digits.

9. Make notes (in the do-file) of any decision you take when checking (cleaning I & II) and replacement of implausible and missing values (cleaning II)..

10. If excel files are used during the cleaning process, save all files in the sub-folder named "excel" and refer the excel file name in the respective do-file.

II. Missing values in the dataset

Missing vales can be divided in two groups:

1. A value which is missing due to skip instructions in the questionnaire (valid missing) is reflected in the data files with a system missing ("."). Please replace it with .s in cleaning I but check the skip instructions in the household questionnaire again.

2. A value can also be missing because :
   a) enumerator did not ask although a question is enabled
   b) interview was not completed.

- We call these "invalid missing".

   for these, cases, when downloading the data from the the Survey Solutions program to STATA the following is shown:

   - the value: **-999,999,999** for numeric variables
   - the value: **##N/A##** for string variables

   These missing values should be replaced with ".m" before checking empty rows in cleaning I and cleaning II. Recode ".m" to "." and replace missing cases following the TVSEP cleaning II procedure. .

   , Replacement of "valid missing" to ".s" must be done before replacement of "invalid missing" to "." in order to avoid mixing up cases.

III. TVSEP cleaning procedure in nutshell

The cleaning procedure of the TVSEP dataset consists of four steps:

A. Data files preparation

Part A is a preparatory part, where it needs to be checked that all interviews conducted during the field survey are available in STATA files, and all variables are in the correct format and correctly labelled.

B. Cleaning I

In cleaning I, every variable is investigated in order to identify problems such as potential implausible values and missing cases. For this purpose, a so-called "dummy variable (d_var)" for each variable in the dataset is generated, that equals to 1 if the respective variable has a problem. A user-written program "flagoutlier2_prov" helps to identify outlying observations. Once all problems in every section had been discovered, the local team leader summarizes the problems for each interviewed household in an Excel table named "Plist (Problem list)". The national cleaning team leader decides whether which households will be called and produce the excel table named "Clist (Call list)" after calling back session, new information will be incorporated in the respective STATA data files. At the end of cleaning I, we have data version 1 that will be used in the next step.

C. Cleaning II

The cleaning II is aimed to generate a cleaned dataset for each section. For this purpose, a so-called "_x variable" for each variable in the dataset (except for the text-variables) is generated, in which all changes and corrections (that were identified for example in Cleaning I) are made.

Second, it is aimed to prepare the income and consumption aggregation variables. For this purpose, "imputed variables (i_ xvar)" is generated for some variables that have missing values and is linked to the income or consumption aggregation process following the replacement rules. At the end of Cleaning II, we have a new version of a dataset (version 2), and income and consumption component variables to be used in the aggregation step. Meanwhile, all observations that contain local languages, for example in the specified code 90 variables or in the variable labels shall be translated into English and merged back at the end of cleaning II

D. Income and consumption aggregates

In part D, income and consumption aggregates across different income generating activities and expenditures in different sections will be generated.

IV. Data files preparation

The preparation step is to prepare the data set according to the TVSEP data file format and to check whether all variables are correctly labelled. In order to kick off, please setup the folder tree before beginning the cleaning process. The structure of do- and data files determines the directory of necessary sub-folders during the cleaning process (see appendix A). The preparation steps are as follow:

1. Adjust and run the insheet do files (exported from Survey Solutions) to import the datasets
   1.1   Execute all rosters in the sub-folder tvsep_th/tvsep_vn
   1.2   In order to completely execute the do-files, variables names have to be corrected (i.e. changing capital letter to a small one or rename variable).
   1.3   Rename and label variables according to the previous waves.
   1.4   Drop variables with the dummy prefix
   1.5   Do-files exported from Survey Solution which has been revised will be saved in the sub-folder do_ss.
   1.6   The prepared data files from this step will be saved in the sub-folder w7_vA

2. After having data version A, all data files need to be organized following the format of the TVSEP database.
   2.1   Prepare the 2017 survey information file (SvyInfo_TH/VN) based on information from section 1 and cross check with the survey information from the 2016 panel (if necessary also other panels). This file includes important unique interview identifiers such as QID, hhid, parentid1 (interviewed), province ID, district ID, sub-district ID, village ID and ID of each household. Please also check whether all interviews are in STATA files by village and compare with the record from the field survey team.
   2.2   Some sections are split into several exported data files e.g. expenditure has 9 files. According to the TVSEP format all these 9 files shall be appended into one data file named "exp.dta"
   2.3   Check whether variables are in the correct format, rename and label the variables according to the previous wave. Please cross check with the paper-based questionnaire. If variables are not correctly named and labelled, please revise them in the insheet do-file (do_ss from step 1) and execute do_ss files again.
   2.4   Household member ID is very important for merging individual members through waves. Please cross check with the household member ID from the previous waves. The member id must be revised (if necessary) before starting cleaning I.
   2.5   Merge prepared dataset from step 2.3 with the survey information using the parenteid as the common variable for merging.
   2.6   All do-files from this step will be saved in the sub-folder do_A

2.7   The prepared data files will be saved in the folder w7_v0 which will be used for cleaning step I. Please follow the file name format from the previous waves (see table 1) and start with a small letter.

Notes

The TVSEP dataset is structured according to the tables in the household questionnaire and comparable to the previous waves. Variables from different tables can be combined by append or merge operations available in STATA. The following identifying variables are important common variables to combine data files.

1. Questionnaire ID (QID) has 10 digits for TH and 11 digits for VN. It is assembled from digit codes of province: 2 (Thailand)/3 (Vietnam); district: 2; sub-district (commune-Vietnam): 2; village: 2 and household:2. QID is generated from village ID (10004 or vill) * 100 + Id of the household (10005)
2. Household id (hhid) is a unique identifier for each household in the database. It was generated in the first panel wave and not shown in the household questionnaire. This variable is merged in all datasets during the preparation step.
3. ID variables (e.g. __42000 crop ID) is a unique identifier for each observation of each household in the dataset. This ID variable is a count variable created by the interviewer during the survey.
4. Roster Id (Id) is another unique identifier for each observation of each household generated by the survey solutions program. For example, if a household has four types of livestock, the survey solution will create roster Id that start from 0 and counts to 3.
5. "parentid1" refers to the Id attributes of the corresponding parent that records at higher level. It is automatically generated from the Survey Solutions and stored in STATA as a string variable. In the TVSEP dataset, the parentid1 contains an identifier of the household.

Table 1: (sub) sections in questionnaire and data files

Household questionnaire		Data files	Do file (cleaning I)	Do-file (cleaning II)
(sub) sections	Page			
1 Survey information	3	SvyInfo_TH SvyInfo_VN		
2.1-2.3 Household members Education Health	5-13	mem	w7_outlier_mem	w7_clean_mem
2.4 Household dynamics	15	hhdyn	w7_outlier_hhdyn	w7_clean_hhdyn
8 Household expenditures	19	exp	w7_outlier_exp	w7_clean_exp
3.1a,b Shocks	21-23	shocks	w7_outlier_shocks	w7_clean_shocks
3.1f Aspirations	29	aspir	w7_outlier_aspir	w7_clean_aspir
3.2 Risks	31	risks	w7_outlier_risks	w7_clean_risks
4.1 Land	33	land	w7_outlier_land	w7_clean_land
4.2 Agriculture	35-43	crops	w7_outlier_crops	w7_clean_crops
4.3.1 Livestock	45	livest	w7_outlier_livest	w7_clean_livest
4.3.2 Livestock products	47	lstprod	w7_outlier_lstprod	w7_clean_lstprod
4.4 Fishing, hunting…	49-51	hunting	w7_outlier_hunting	w7_clean_hunting
5 Wage employment	53	offempl	w7_outlier_offempl	w7_clean_offempl
6 Non-farm self-employment	55-57	selfempl	w7_outlier_selfempl	w7_clean_selfempl
7.1C Credit rationing	59	credrat	w7_outlier_credrat	w7_clean_credrat
7.1D Default history	59	defhist	w7_outlier_defhist	w7_clean_defhist
7.1A Borrowing	61-63	borr	w7_outlier_borr	w7_clean_borr
7.1E Savings	67	sav	w7_outlier_sav	w7_clean_sav
7.2A Public transfers	69	transf	w7_outlier_transf	w7_clean_transf
7.2B Insurances	71	insur	w7_outlier_insur	w7_clean_insur
6.2 Investment	75	invest	w7_outlier_invest	w7_clean_invest
6.2 Disinvestment	75	disinvest	w7_outlier_disinvest	w7_clean_disinvest
9.1 Household wealth	77	assets	w7_outlier_assets	w7_clean_assets
9.2 Housing conditions	81	house	w7_outlier_house	w7_clean_house
Questions at household level		hh	w7_outlier_hh	w7_clean_hh
Member Id and nucleus household membership		mem_Id[1]	w7_outlier_mem	w7_clean_mem

---

[1] An additional output data file from the member section. The responsible data cleaner for member section needs to provide this data file as earlier as possible to the other cleaners during cleaning step I and II.

V. Interview comments preparation

Interview comments are often useful for data cleaners to determine missing and outlying cases as well as implausible replacement. Interview comments can be noted from enumerators and/or conversation between enumerators and Data Checking Assistants (DCA). The interview comments is exported as tab-delimited data in a .tab file that is a system generated files from the Survey Solutions program.

In order to incorporate interview comments into each dataset during cleaning I and II, the preparation step is required. In this step, three output files of each section will be prepared and saved in the sub-folder w7_com (see file name in appendix B). The following commands (see also MasterDoCom_w7.do) use section 4.3.1 as an example:

```
version 14
set more off
capture clear all
```

1. **execute the "settingsh.do"**
   ```
   global SETTINGS "…\Folder Tree_TH2017\Cleaning_7thWave"
   ```
                                    // adjusted according to your directory

   ```
   do "$SETTINGS\settingshW7.do"
   ```

2. **Prepare full interview comments of each section**
   ```
   cd "$DO_INSHEET_W7"
   clear
   insheet using interview_comments.tab
   ```

   *** Keep corresponding roster names & non-empty comments[2]
   ```
   keep if (roster == "StockRoster4_3_1")
   keep if !missing(comment) & !missing(id1)
   ```

   *** Save all interview comments in the STATA and Excel format[3]

   ```
   cd "$RAW_W7_COM"
   save com_livest.dta, replace

   export excel using "..your path\w7_com\com_livest.xls",
   firstrow(variables) replace
   ```

---

[2] See the corresponding roster names in table 2
[3] Excel is also helpful to check all interview comments of a variable at the same time.

3. Prepare the last interview comment of each variable[4]

3.1 Keep only the latest updated comments using date and time variables
*** Create numeric dates and times to check for updated comments
```
gen date2 =date(date , "MDY")
```
   //Type - format date2 %tdMonth_dd,_CCYY - to get Human Readable Form
```
gen time2 =clock(time, "hms")
```

3.2 Check for duplicates if necessary
```
bysort interviewid roster variable id1 (date2 time2): gen chk_dup =
cond(_N==1,0,_n)
tab chk_dup, m
```

3.3 Keep only the latest updated comments
```
bysort interviewid roster variable id1 (date2 time2): keep if _n==_N
```

3.4 Create consistency among comments for later check
```
replace comment =ustrtrim(comment)
replace comment =ustrlower(comment)
```

3.5 Rename vars in "roster names" to match vars in "livest.dta"
```
tab variable, m
rename (interviewid id1) (parentid1 Id)
```

3.6 Create vars containing comments corresponding to vars in "livest.dta"
```
levelsof variable, local(vars)
 foreach var of local vars {
local name =substr("`var'", 2, .)
gen t__`name' = comment if variable == "`var'"
```
                                          //Each var contain corresponding survey comments
```
 }
*
ds t__431*
 foreach var in `r(varlist)' {
bysort parentid1 Id (`var'): replace `var'= `var'[_N]
 }
 *
collapse (lastnm) t__431*, by(parentid1 Id)
```

4. Save the last interview comment for merging with the corresponding section
```
cd "$RAW_W7_COM"
save t_livest.dta, replace
clear
```

---

[4] The last interview file can be used to merge with your dataset during the cleaning process.

Table 2: Roster Name in the tab-delimited interview comments file

Household questionnaire		TVSEP	Roster name
(sub) sections	Page	data files	
1 Survey information	3	SvyInfo_TH SvyInfo_VN	TVSEP 2017 svyInfo
2.1-2.3 Household members Education Health	5-13	mem	HHRoster2
2.4 Household dynamics	15	hhdyn	hhroster2_4
8 Household expenditures	19	exp	ExpenditureRosterCorruption ExpenditureRosterEducation ExpenditureRosterFood ExpenditureRosterHealth ExpenditureRosterNonFood ExpenditureRosterOther ExpenditureRosterSocial ExpenditureRosterTotal ExpenditureRosterTransportCommun
3.1a,b Shocks	21-23	shocks	ShockRoster3_1a ShockRoster3_1b
3.1f Aspirations	29	aspir	Roster3_1d
3.2 Risks	31	risks	FutureRiskRoster
4.1 Land	33	land	LandRoster
4.2 Agriculture	35-43	crops	AgricultureRoster
4.3.1 Livestock	45	livest	StockRoster4_3_1
4.3.2 Livestock Products	47	lstprod	LivestockRoster4_3_2
4.4 Fishing, hunting…	49-51	hunting	HuntingRoster
5 Wage employment	53	offempl	offfarmroster
6 Non-farm self-employment	55-57	selfempl	NonFarmRoster
7.1C Credit rationing	59	credrat	borrowing_roster1
7.1D Default history	59	defhist	borrowing_roster2
7.1A Borrowing	61-63	borr	borrowing_roster3
7.1E Savings	67	sav	savings_roster
7.2A Public transfers	69	transf	PublicTransferRoster
7.2B Insurances	71	insur	InsuranceRoster
6.2 Investment	75	invest	InvestmentRoster
6.2 Disinvestment	75	disinvest	DivestmentRoster
9.1 Household Wealth	77	assets	HouseholdWealthRosterA HouseholdWealthRosterB
9.2 Housing conditions	81	house	TVSEP 2017 svyInfo
Questions at household level		hh	TVSEP 2017 svyInfo

VI. Cleaning I

The purpose of cleaning I is to flag potential outliers and to identify missing cases. Please provide 1 do-file for each dataset (see table 1).

The structure of variables in the cleaning I is as follows:

- Uncleaned variables (e.g. __43107a) contain the original values or answers those exported from the Survey Solutions
- Dummy variables (e.g. d__32005) are generated to indicate missing and outlying observations in cleaning I.
- Text variables (e.g. __21018t) record specific answers of code 90 in local language (_t suffix)

```
version 14
set more off
capture clear all
```

1. **execute the "settingsh.do"**
   ```
   global SETTINGS "…\Folder Tree_TH2017\Cleaning_7thWave"
   ```
                                    // adjusted according to your directory

   ```
   do "$SETTINGS\settingshW7.do"
   ```

2. **Set up necessary files**

   2.1 Execute flag outlier program(s)
   ```
   do "$PROGRAMS\program_flagoutlier2_prov.do "
   ```

   2.2 Open log file
   ```
   cd "$LOG_1"
   capture log close
   log using livest, replace
   ```

   2.3 Open the corresponding data file
   ```
   cd "$RAW_W6_V0"
   use livest.dta, clear
   ```

   2.4 Merge data with the last interview comment file
   ```
   capture drop _merge
   count
   merge m:m parentid1 Id using "$RAW_W7_COM\t_livest.dta"
   count
   drop if _merge==2      //Drop obs appeared in "t_livest.dta" only
   drop _merge
   count
   ```

Note: interview comments are notes and conversation between enumerators and supervisors during the survey. If the last interview comment is not sufficient, please check the full interview comment file.

3. Preliminary check

If your STATA does not have the following commands, please install them first:

```
ssc install findval
ssc install valuesof
ssc install distinct
```

3.1 Encode missing value

3.1.1 Check true value 98\99 in numeric continuous variables only

```
foreach num in 98 99 {
findval `num' __41003 __41009a __41011h __41013
  foreach var in `r(vars)' {
        count if `var' == `num'
        display "Number of `num' found in `var':  " `r(N)'
  }
 }
 *
```

By using this above command set, number of observation with 98 can be detected in each variable.

Code 98 is used when respondent does not know the answer or does not want to answer. If a plausible answer for quantitative information (e.g. crop yield) is also 98 enumerators will write 98.1 instead. However, enumerators might forget to apply this rule during the interview. Therefore data cleaners should check with the survey comment whether 98 is true value or codes for do not want to answer and do not know. If it is a true value, please replace it with 98.1 before proceeding to step 3.1.2.

Code 99 is for non-applicable case. In the 2017 survey, there is only two variables i.e. 23012 and 41014. For example variable 23012 can contains code 99 when a household member did not have an occupation during the reference period. Therefore number of weeks that this person was unable pursue his main occupation because of health impairment is not applicable.

3.1.2 Encode true code 98 and 99 to '.b' \'.c'

```
ds __43*, has(type numeric)
foreach var in `r(varlist)' {
    assert `var' <= .
    }
*
mvdecode `r(varlist)', mv(98 =.b \ 99 =.c)
```

3.1.3 Dealing with 98.1 and 99.1 in data

```
ds __43*, has(type numeric)
foreach var in `r(varlist)' {
    replace `var' =98 if `var' ==float(98.1)
    replace `var' =99 if `var' ==float(99.1)
    }
*
```

3.2 Drop empty rows

Empty row is a row which has no observation in all important columns (variables) due to data entry errors or merging. The following commands can be used to drop empty rows:

Table 3: STATA commands for checking and dropping empty rows

Commands	Explanation
ds __431* , has(type numeric) foreach var in `r(varlist)' { mvdecode `var', mv(-1000000000 -999999999=.m) } *	Recode invalid missing codes from Survey Solutions to .m
egen nmiss = rowmiss(__431*)	Generates an auxiliary variable "nmiss" that counts the number of missing observations for all variables in the section
tab nmiss	Shows you the number of missing values in variable list for each observation
ds __431* scalar nvars = `: word count `r(varlist)''	Generate the number of variables (include string and numeric variables)
drop if nmiss==nvars	Drops rows that have missing values for all variables
br if nmiss==nvars-1 & __43100!=.	Browses to check whether the non-missing variable for observations with nvars - 1 is not an ID variable (__43100)
drop if nmiss==nvars - 1 & __43100!=.	Drops if the non-missing variable is only the ID variable
drop nmiss nvars	Drops the auxiliary variable

3.3 Drop duplicates

Duplicates are observations (rows) with same response for all or major variables. For example, for one household, if rice production is recorded two times with the same answer for all questions, this observation could be a duplicate observation. There should be at least one variable that differentiate the two observations e.g. household could grow 2 times rice on the same parcel but at different time periods. There are two approaches to check duplicates:

- checking duplicates based on all variables

```
duplicates tag, generate(dup)
tab dup, m
br if dup>0
duplicates drop
drop dup
```

- checking duplicates based on important variables

```
sort hhid-V
quietly by hhid-V: gen chk_dup = cond(_N==1,0,_n)
tab chk_dup, m
br if chk_dup >0
br if chk_dup>1
drop if chk_dup>1
drop chk_dup
```

The second approach is more useful to identify duplicates based on important variables when all variables are not the same, but the differentiating variable is not an important one e.g. ID variable (__43100).

3.4 Generate dummy variable for each variable to indicate outlier and missing observations

```
label define flag 1 "outlier" 99 "missing"
ds __431*, has(type numeric)
foreach var of varlist `r(varlist)' {

    generate d`var' =.
    label var d`var' "Outlier `: var lab `var''"
    label val d`var' flagoutlier
}

*
```

3.5 Generate dummy variables of 3 provinces

```
tab prov, gen(prov)
rename (prov1 prov2 prov3) (HT HU DL)or (UB BR NP)
```

4. Flag outliers and missing observations

4.1 check variable name and its label using command describe

If you find wrongly labelled variables, please correct the label and add a comment beside the revised command.

4.2 encode missing values and check implausible cases by variables or group of variables

- Encode the valid missing "." to ".s" prior to the skip instruction in the household questionnaire. Although this missing value is generated following the skipping design, it is recommended to always check with the questionnaire and conditional checking commands (each variable or group of variables) in the example do-files.

- Encode the special missing values ".m" (recoded from -999,999,999) and ##N/A## (for string variables) to "." to be checked for calling back and/or replacement in cleaning I and II, respectively. This replacement step must be done after encoding the valid missing to ".s" in order to avoid mixing up cases[5].

- Check outlier and missing observations

General notes:

- Please check skip instruction of each variable before you flag out outlying and missing observations.

- Some variables need to be cross checked with other variables in different sections. Therefore opening the relevant datasets of other sections or merging them is required. This might involve two or more sections.

- The common error patterns that occur from using the tablet questionnaire are digit error and switching column. In most cases these errors can be corrected during the cleaning II (you can guess!). Please do not flag out outliers or missing that can reasonably be corrected from surrounding context and common sense. Please note such cases in the do-file for correcting these errors in cleaning II. For example enumerators recorded height of member in variable 23006 instead of 23007.

- Please do not flag out outliers and missing observations of text variables (variables with -t suffix).

---

[5] Valid and invalid missing values shall be encoded variables by variables before checking implausible observations.

Notes by type of variable

ID variables: check whether IDs of the rows are consecutively numbers. There should not be missing or zero values. This rule does not apply to the following fixed ID variables __21001, __22001, __23001, __42004, __50003, __60003, __72101, __72204.

Table 4: Example STATA commands for checking ID variables

Commands	Explanation
`tab __43100, m`	
`label define d_value 0"wrong" 1"correct"`	Define value of dummy variable
`bysort hhid (Id): gen row_ID =_n`	Generate consecutive row ID based on hhid and roster Id
`gen d_rowID =(_43100 - row_ID)!=0`	Generate dummy variable for checking rows those are not consecutively numbered
`Label val d_rowID d_value`	
`tab d_rowID, m`	
`distinct hhid if d_rowID == 1` or `count if _43100 != row_ID`	Count the number of observations that are not consecutively numbered

Please do not flag out ID variables but note in the do-file how many observations and which hhid those non-consecutive number ID variables. You can use this information during the cleaning II.

String variables: check whether the observations are within the range of answering codes provided in the questionnaire for each question.

Table 5: Example STATA commands for checking string variables

Commands	Explanation
`tab __43102, m`	Tabulates the uncleaned variable and shows the number of missing
`br __431* if __43102>=.`	Browse to see the missing cases
`replace d__43102=1 if __43102 == .`	Flag out only the invalid missing cases which you cannot edit in cleaning II based on interview comments or other information

Numeric integer variables: (e.g. __42006 (month or year of planting period): check whether the observations are plausible.

Table 6: Example STATA commands for checking numeric integer variables

Commands	Explanation
tab __42006 , mis	Tabulates the uncleaned variable and shows the number of missing
replace d__42006=1 if __42006 == .	Replaces the variables d__42006 to 1 if the observation for the variable is missing
replace d__42006=1 if __42006<1 | (__42006>12 & __42006<2000) | (__42006>2016)	Replaces the variable d__42006 to 1 if the planting period is less than 1 month, larger than 12 months and less than 2000, or larger than 2016
tab d__42006 , mis	Tabulates the dummy variable to check if missing values and outliers are correctly flagged in d__42006

Note: In some cases, you might be able to edit the data without calling back household. For example, year of planning period was record as 20016, this errors could be from typing error. If you are sure in such cases, please do not flag out.

Numeric continuous variables

For numeric continuous variables, please use the flag outlier program (program_flag2_prov.do[6]) to detect outlying observations. If the number of outliers is more than 2 percent of total observation, the range of plausibility rules in the appendix will be used as reference to determine the final outlying observations for the plist. Please note that in some sections, you must check outliers by group of variables. For instance, in section 4.3.1 you check stock at the end together with stock at the beginning and changes over the reference period.

In general, to all of the three variable types (string, integer, and continuous variables) you can check the additional comments from the field if an explanation provided for an implausible value in the comments does make sense, the observation can be accepted and does not have to be marked as an outlier.

---

[6] The standard procedure for the identification of outliers is to calculate lower and upper bounds by adding and subtracting respectively two standard deviations from the median of any group of at least ten observations. Values below and above these thresholds are considered to be outlying.

5. Drop rows and dummy variables without errors

```
egen row = rowtotal(d_*)        // Generates an auxiliary variable
                                "row" that  sums up all the dummy
                                variables (d_*)  to identify rows
                                without problems (i.e., rowtotal==0)
tab row
drop if row == 0                // drops rows without problems
drop row                        // drops auxiliary variables
```

6. Create listing

Variable "rowno" is generated and will be used to identify specific observations of households in each section that need to be corrected by calling back households.

```
gen rowno= __43100
```

Please note that when you generate the variable "rowno" in sections that have only one row per household (i.e. hh.dta), the rowno variable should be created with missing value.

```
gen rowno= .
```

7. Replace dummy variables and code missing observations[7]

In this step, the dummy variables will be replaced with the actual implausible values from the uncleaned variables for the flagged outlier observations and coded with "-99" for missing cases.

```
drop __431*t                        // make sure t-vars are removed

foreach var of varlist __431* {
sum d`var'
if `r(N)' == 0 {
   drop d`var'        }    // Drops dummy variables without outliers
else {                         // opens the following inverse case
replaced `var' = `var'    if d`var' == 1 & `var' !=.
replaced`var' = -99 if d`var' == 1 & `var' ==.
drop `var'
rename d`var' `var'
local varskeep `varskeep' `var'
   }
}
*
keep QID hhid `varskeep' rowno prov distr subdistr vill Id
aorder
```

---

[7] Please note that commands in step 7-9 must be executed together.

8. Replace ".b\.c\.s" (code 98, 99 and valid missing) in the list with missing so that the non-response errors are not in the list of calling back

```
ds __431*, has(type numeric)
capture {
 foreach var of varlist `r(varlist)'  {
   foreach letter in b c {
        replace `var' = . if `var' == .`letter'
   }
 }
}
*
```

9. Check empty rows

```
if !_rc{
   tempvar miss
   egen `miss' =rowtotal(`varskeep'), miss
        // generate a variable "miss" that sums up all variables
        treating missing as 0 to identify rows without problems
   br if `miss' ==.
   drop if `miss' ==.
   drop `miss'
}
*
```

10. Save data in the temp folder

```
quiet compress
sort hhid
cd "$CLEAN_W7"
save w7_outlier_livest, replace
log close
```

11. Preparing problem list (plist) and call list (clist)
The local team leader uses the "w7_outlier_list_main.do" and
w7_outlier_collate.do"- files to export the implausible and missing cases into an
Excel file called "plist" and separate excel files by section. After that please import
the excel files as separate sheets into the excel file named "plist_dimension" and
modify the excel file to filter all sheets. The excel "plist_dimension" has an overview
from which household can be selected (name of household head, respondent, contact
information etc.). To prepare this file, information relating to the household shall be
extracted from the data file "hh.dta".

A team meeting (optional) shall be organized to finalize the plist_dimension file and
produce the file named "Clist[8]". The decision of including/excluding flagged out

---

[8] Clist file is the plist_dimension that only contains households those we will call back.

cases will be made by the national cleaning team leader. Only exceptional cases will be called back.

During calling process, please make sure that you call a household only ONCE! That means everybody in the cleaning team should be finished until step 10.

If the telephone number of a household is wrongly recorded, you call the village headman to reach this household. However, the first thing after reaching this HH is to ask the correct telephone number from the respondent. Then you call HH respondent directly to get the information. Not asking other people (e.g. village headman, neighbour) to interview the HH respondent for you. **Never accept information from the third party. It must be from the HH respondent.**

Major rules for finalizing the call list

In principle, we will call back to check implausible interdependent variables across different sections. The major rules to finalize the call list are as following:

I.   Variables 21014 (major occupation) and 21015 (second occupation) should match with information in section 4.2, 5 and 6; and also with education sub-section if household members are students. For example, if the household respondent reported that his daughter is currently working in a computer company in Bangkok, but job information of this person does not exist in section 5. We will call back household to ask questions in section 5 of this household member[9].

II. Each crop should be allocated to a plot of land listed for agricultural production in section land. If a household reported in section land that 3 plots were used for permanent crops (41004=4), but there is only one row for a permanent crop in section crops instead of 3 rows, such case will be clarified by calling back whether household used only 1 plot for permanent crops or information in section 4.2 is missing. In principle, information (if plots are used for growing crops) that is shown in the land section must be shown in the crop section.

III. In the stock section, you will call back for missing/implausible cases those are at enterprise level which is according the number of stocks (i.e. household owns at least 5 buffalos, beef cattle, dairy cattle, pig fattening, goats and fighting cocks; 2 piglet production; 1000 chickens, ducks and turkeys).

In this case you can call households if the stock is not balanced and you cannot edit based on the surrounding information. With regard to implausible price, additional information on size of livestock or age shall be asked to cross check whether the reported price is reasonable even it is detected as an outlier.

---

[9] If household respondent does know information of any absence household member, you may consider to directly call to this person. However, the first attempt shall be with the household respondent.

Major input cost in the stock sections shall be also cross checked with the section crop and number of animal. If household had at least 1000 chickens, and neither feed cost nor own produced feed (crop section variable 42012a) is recorded, such case has to be called back.

IV.   Investment and disinvestment can be for farm, business and personal use as well as household use. The section refers to the investment or disinvestment strategy for productive and consumptive purpose above a value of 5000 THB/1.5 Million VND. Therefore, cross checking between sections investment/disinvestment and land, livestock (purchasing or sale value of livestock at each time) or asset section is required. If any information is missing or implausible either from the investment/disinvestment section or the relevant sections, please call back households for further clarifications.

V.   Call back if income from any business is extremely high or negative. In this case, it is required that a data cleaner will calculate the total cost and revenue per business by applying the commands in cleaning II.

VI.   If household reported a case of default history but this loan does not exist in the section borrowing, we will call back to ask for this missing case.

Generally, you will ask households a set of question in order to clarify missing and outlying observations. It could be that you cannot fill in information question by question while speaking with a household respondent. However, a caller should prepare what to ask and have background information about a household[10]. Please make notes of all answers you receive and fill in the call back information in the excel call list to be used for replacement at the end of cleaning I.

12. Replacing new value

The new information from calling back session should be checked by the data cleaning members with respect to their responsible sections together with the local team leader. It could be that old information is better than the new one. After carefully checking, a local team leader creates a replace commands for all sections (only acceptable cases) and send the Clist file with replace commands to the cleaning team members.

To replace the information from calling back, please do it as step 12 in the do_outlier_*name* by opening the respective data file from the folder w7_v0:

```
set more off
capture clear all
cd "$RAW_W7_V0"
use livest.dta, clear
```

Please copy the replace commands of your responsible sections from the plist_dimenstion excel file to the respective do file and execute them. (**Close all**

---

[10] It is recommended to go through the questionnaire (TVSEP 2017 survey server) to get familiar with the household before calling.

commands from steps 2.1, 2.2, 3.1.1, 3.1.2, 3.4, 4.2[11] (only commands for checking implausible cases) and steps 5-10 before executing the replace commands.

```
replace _31012a=30 if hhid==695 & Id==29320 &  _31012a==60
…
…
…
```

13. **Save data version 1**

At the end of cleaning 1, you will have data version 1 which will be saved in the folder "w7_v1"

```
quiet compress
sort hhid
lab data "TVSEP – wave7 – 2017TH- livest_v1 - $S_DATE, $S_TIME"
cd "$RAW_W7_V1"
save livest.dta, replace
```

14. **Exporting code 90 text variables for translation**

After having data version 1, please export code 90 variables (_t* variables) including necessary variables such as QID, hhid, Id, parentid1, __10001, __10002, __10003, __10004 and __10005 to Excel files for translation. The translation shall be done parallel to cleaning II. The translated response should be recorded by creating a new variable with a _ten suffix (e.g. __43102t >>> __43102ten).

```
cd "$RAW_W7_V1"
use livest.dta, clear
sort hhid Id
count                           // note "number of observation"
export excel QID hhid Id __42003at using "directory
path\livest.xls", firstrow(variables)
```

Notes:

- Number of observation in the data version 1 and in the exported excel shall be the same.
- Translation of code 90 text variables can be done simultaneously with cleaning II.

---

[11] Please do not close the command that you encode special valid and invalid missing values.

VII. Cleaning II

The purpose of cleaning step II is to provide a cleaned dataset for each section and generate variables to be used in the income and consumption aggregation. Please follow this cleaning guideline, master do-file and the provided do-files from the previous wave to create do-files of your respective section. In addition, the do_1-files of "Cleaning I" will be also useful to identify variables that need further cleaning.

The structure of the variables in the cleaning II is as follows:

- Uncleaned variables are exported from the Survey Solutions and renamed with __prefix (i.e. __43100) please do not correct these variables.
- Text variables (e.g. __21018t and __21018ten) record specific answers of code 90 in local language (_t suffix) and in English (_ten suffix).
- Cleaned variables: _x prefix (i.e. _x32005) are generated for each variable to include all changes and corrections that will be made in cleaning step II. You do not need to generate _x variables for t-variables since t-variables do not need cleaning.
- Imputed variables: (e.g. i_x43107a) are those with missing observations and therefore imputed using replacement rules

```
version 14
set more off
capture clear all
```

1. **execute the "settingsh.do"**
   ```
   global SETTINGS "…\Folder Tree_TH2017\Cleaning_7thWave"
                             // adjusted according to your directory

   do "$SETTINGS\settingshW7.do"
   ```

2. **Set up necessary files**

   2.1   Execute flag outlier program(s)
   ```
   do "$PROGRAMS\program_flagoutlier2_prov.do "
   ```

   2.2   Open log file
   ```
   cd "$LOG_2"
   capture log close
   log using livest, replace
   ```

   2.3   Open the corresponding data file
   ```
   cd "$RAW_W7_V1"
   use livest.dta, clear
   ```

2.4 Merging interview comments

```
capture drop _merge
count
merge m:m parentid1 Id using "$RAW_W7_COM\t_livest.dta"
count
drop if _merge==2    //Drop obs appeared in "t_livest.dta" only
drop _merge
count
```

3 Preliminary check

3.1 Encode missing value

3.1.1 Check true value 98\99 in numeric continuous variables only

```
foreach num in 98 99 {
findval `num' __41003 __41009a __41011h __41013
foreach var in `r(vars)' {
     count if `var' == `num'
     display "Number of `num' found in `var':  " `r(N)'
}
 }
 *
```

By using this command set, number of 98 can be detected in each variable. Please check with the survey comment whether 98 and 99 are true value or codes for do not want to answer, do not know and not applicable. If it is a true value, please replace it with 98.1 or 99.1.

3.1.2 Encode true code 98 and 99 to '.b' \'.c'

```
ds __43*, has(type numeric)
foreach var in `r(varlist)' {
assert `var' <= .
}
*
mvdecode `r(varlist)', mv(98 =.b \ 99 =.c)
```

3.1.3 Dealing with 98.1 and 99.1 in data

```
ds __43*, has(type numeric)
foreach var in `r(varlist)' {
replace `var' =98 if `var' ==float(98.1)
replace `var' =99 if `var' ==float(99.1)
}
*
```

3.2 Drop empty rows

Empty row is a row which has no observation in all important columns (variables) and should be dropped. To do this, you can use the following commands:

```
ds __431* , has(type numeric)
foreach var in `r(varlist)' {
mvdecode `var', mv(-1000000000 -999999999=.m )
}
*
egen nmiss = rowmiss(__431*)
tab nmiss
ds __431*
scalar nvars = `: word count `r(varlist)''
drop if nmiss==nvars
br if nmiss==nvars-1 & __43100!=.
drop if nmiss==nvars - 1 & __43100!=.
drop nmiss nvars
```

3.3 Drop duplicates

Duplicates are observations (rows) with same response for all or major variables. For example, for one household, if rice production is recorded two times with the same answer for all questions, this observation could be a duplicate observation. There should be at least one variable that differentiate the two observations e.g. HH could grow 2 times rice on the same parcel but at different time periods. There are two ways of checking for duplicates:

- checking for duplicates based on all variables
```
duplicates tag, generate(dup)
tab dup, m
br if dup>0
duplicates drop
drop dup
```

- checking for duplicates based on important variables
```
sort hhid-V            // V is the important variables that
should not be duplicated
quietly by hhid-V: gen chk_dup = cond(_N==1,0,_n)
tab chk_dup, m
br if chk_dup >0
br if chk_dup>1
drop if chk_dup>1
drop chk_dup
```

The second way is more useful to identify duplicates based on important variables when all variables are not the same, but the differentiating variable is not an important one e.g. ID variable (__43100).

3.4 Generate the cleaned _x variables

```
foreach x of varlist __43* {
capt confirm numeric var `x'              // exclude string vars
if !_rc {
local name = substr("`x'",3,.)            // store root of varname
gen _x`name'=`x'                          // create copy
lab var _x`name' "`: var lab `x''"        // copy variable label
lab val _x`name' `: val lab `x''          // copy value label
  }
}
*
```

4   Cleaning and correcting implausible / missing values in the _x variables

The common error patterns that occur from using the tablet questionnaire are digit error and switching column. In most cases these errors can be corrected with commons sense and support information from survey comments (you can guess!). For example enumerators recorded height of member in variable 23006 instead of 23007. For this kind of error, instead of flagging out in cleaning I, please make note in the do-file and edit them in cleaning II. In some sections, checking outliers by group of variables is required. In this regard, the digit error or switching column errors shall be revised in the auxiliary variables (i.e. generate variables with x_prefix) during the cleaning I process and these correction commands can be used in cleaning II.

In many categorical variables those have a long code list, code 90 was often used. If the answer (see code 90 text variables) corresponds to one of the code list in the household questionnaire, please recode it. For example, variable 42003a was recorded with code 90 (Hom Putum rice), you can recode this observation from code 90 to 102: "Other fragrant rice". It is important to recode it at the beginning of cleaning II because some variables will be later used as unique group identifier to check and generate income and consumption variables. If there are many lines of command that you would have to write for recoding code 90, it can be faster to copy to excel and generate syntax automatically[12]. In case that code 90 does not match with any existing code list, please check the number of observation. If this new answer is recorded more than 200 observations, report the cases to the research database manager (some answers from code 90 were already recoded in the 2013 panel wave). Otherwise, please leave it as code 90 and data users can decide to group them later according to their research objectives. In this regard, translation of code 90 text variables is important.

Detecting outliers in cleaning II aims to determine problematic cases before replacement and imputation processes. The standard procedure of identifying outliers will be applied. The first step is to use the flag outlier program (mean +/- 2 standard deviation) to identify outlying observations (notes in the cleaning I do-files are useful). If the number of

---

[12] The excel file shall be saved in the respective excel folders of cleaning I and II, and note the excel file name in your do-file.

outlying observation is more than 2 percent of total observation, the range of plausibility rules in appendix C (Thailand) and D (Vietnam) will be used as reference to reduce the number of outliers and to judge whether replacement/imputation is needed.

In order to generate proper income and consumption variables, it is necessary to replace implausible and missing values. This also includes observations replaced by missing values if implausible values are detected and cannot be revised from calling back or during the first step of cleaning II. Two approaches will be applied to replace missing and outlying observations. The first one is to apply the written impute program (*impute2.do*) that imputes missing value depending on certain predefined variables. However, the program may not work if number of observation at the nearest possible level of sample (village, commune, etc.) is not sufficient. In that sense, replacing by mean seems to be practical. Importantly, the generated mean variable should not include outlying observations. Then you can replace missing value in the imputed variable (i_x variable) with the generated mean.

Zero observations shall be checked whether they are true zeros. If it is untrue zero answers, please correct it if possible. For variables those are used to generate income and consumption, please apply the replacement procedure if you cannot guess the value.

5. Generate income and consumption variables

Table 6 provides an overview of income source and variables those are generated at the end of cleaning II. Income variables at the household level will be saved in the sub-folder w7_v3 (see files' name in appendix B) to be used in the income aggregation.

Table 7: Income variables

Income component	Data file (.dta)	variable at individual activity	Variable at the household level
Remittances received from HH members	mem	_x21080	_x10080
Remittances received from non-HH members	hhdyn	_x24012 & _x24013	_x10081
Income from land rent	land	x41083	x10083
Income from crop	crop	x42084	x10084
Income from livestock	livest lstprod	_x43185 _x43285	_x10085a _x10085b _x10085 = _x10085a+ _x10085b
Income from natural resource extraction	hunting	_44086	_x10086
Income from wage employment	offempl	x50087	x10087
Income from non-farm self-employment	selfempl	_x60088	_x10088
Capital income from saving	sav	x71520	x10092
Public transfer received	transf	x72193	x10093
Indemnity payments received	insur	x72212	x10094
Deduction component			
Cost of land rent for agricultural purposes	land	_x41096	_x10096
Cost of land rent for business purpose	land	_x41096a	_x10096a
Cost of loan for productive assets	borr	x71197	x10097

Note: Unlike wave 1 to 5, value of owner-occupied dwelling and depreciation of productive assets are excluded from household income aggregate.

After calculating income at the individual activity, please check outliers using summary statistics (e.g. `tabstat`, `sum`, `mean` etc.) and box plot graph by provincial level and in many cases also categorized by type of activity. Negative and extreme positive income should be checked through all relevant variables (revenue and cost) by type of activity.

The sources of consumption in the household questionnaire are described in table 7. The expenditure section impacts on all consumption sub-aggregates except for housing. Expenditure for rent has been excluded since (imputed) rent equivalents from the housing section will be used. Received remittances from section 2.1 and 2.4 are not included because the questions do not differentiate between monetary and in-kind remittances. The consumption variables will be generated after income variables and are saved in the sub-folder w7_v3 for household consumption aggregate (see files' names in appendix B).

Table 8: Consumption variables

Consumption component	Sources in the questionnaire	Data file (.dta)	key variables for calculation	Variable at the household level
Food	In kind payment from rental outed land	land	_x41011h	food_410
	Crop	crops	_x42011, hcVal	food_420
	Livestock	livest	_x43107a	food_431
	Livestock products	lstprod	_x43205	food_432
	Natural extraction	hunting	_x44016a	food_440
	Non-farm self employment	selfempl	_x60032*	food_600
	In kind loan received during the reference period	borr	_x71105	food_711
	Public transfer	transf	x72103**	food_720
	Expenditures	exp	_x80003a, _x80002, x80004	food_800
Non-food	Crops	crops	_x42011, hcVal	non_food_420
	Livestock products (silk from silk worm, dung)	livest	_x43205	non_food_432
	Natural extraction	hunting	_x44016a	non_food_440
	Non-farm self employment	selfempl	x60032	non_food_600
	Premium paid for insurance	insur	x72209	non_food_722
	Expenditures	exp	x80002, x80004	non_food_800
Housing	Housing (rent equivalent)	house	x92017	rent_900
Durable goods	Assets	assets	_x91108	non_food_910
Education	Expenditures	exp	_x80004	edu_800
Health	Expenditures	exp	x80004	health_800
Rice	In kind payment from rental outed land	land	_x41011h	rice_410
	Crops	crops	_x42011, hcVal	rice_420
	Expenditures	exp	_x80003a, _x80002, x80004	rice_800
Alcohol/Tobacco	Crops	crops	_x42011, hcVal	alc_420
	Expenditures	exp	x80004	alc_800
Social	Expenditures	exp	x80004	social_800
Transport & Communication	Expenditures	exp	_x80004	transcom_800

Note * if variable _x60005a = 1, 7, 8 13 14 15 16, otherwise it is considered as non-food consumption.
    ** if variable _x72102a == 2 and 3

6. Merge English code 90 text variables

Merge data file with the English code 90 text variables from the sub-folder w7_v90en. After merging the translated code 90-variables please check whether the translation was done properly. The number of observations after merging should not be changed.

7. Clear up and save the final data files
- Please drop unnecessary auxiliary variables those are generated during the cleaning process.
- Re-check whether all variables are named and labelled correctly.
- Ordering all variables

```
order hhid Id prov distr subdistr vill _x* __*
```

- Some notes for dataset or some variables can be useful for data users.
```
note _dta:
note variable:
```

- Label dataset
```
lab data "TVSEP - wave 7(TH/VN) - 2017 – livest(v2.0) - $S_DATE,
$S_TIME"
```

- Save the final cleaned data files in the folder w7_v2
```
cd "$RAW_W7_V2"
save crops, replace
```

- Save the income variables at the household level
```
duplicates drop hhid, force
keep hhid QID _x10085a _x10431

compress
sort hhid

lab data "TVSEP - wave 7 VN - 2017 – livest_hh(v2.0) - $S_DATE,
$S_TIME"

cd "$INCOME_W7"
save livest_hh, replace
```

- Save the consumption variables at the household level

```
cd "$RAW_W7_V2"
use livest, clear
duplicates drop hhid, force
keep hhid QID food_431

compress
sort hhid

lab data "TVSEP - wave 7 VN - 2017 – hc_livest(2.0) - $S_DATE,
$S_TIME"

cd "$CONSAGG_W7"
save hc_livest.dta, replace
```

VIII. Notes by section[13] (for cleaning I and cleaning II)

The purpose of this chapter is to outline specific checking steps and decision rules of some variables and/or sections for the process of cleaning I and II. These notes are a culmination of cleaning experience from 2007 to 2016 as well as based on some relevant literatures. The way and steps of checking shall be applied in both cleaning I and II. Regarding to cleaning I this note will be used in order to decide whether to flag out implausible / missing cases for calling back households. After that the checking processes will be repeated again in the cleaning II before replacement of implausible and missing values. Example of detailed STATA commands can be found in the provided do-files from the cleaning 2016.

Each section starts with the common checking rules that can be used to decide whether you would flag out for the case of cleaning I and correct or impute for the case of cleaning II. Furthermore, generating income and consumption variables are explained in the following sections i.e. member, household dynamics, land, agriculture, livestock, natural extraction, wage employment, non-farm self-employment, borrowing, public transfer, insurance, expenditure and housing condition.

Section 2.1 / 2.2 / 2.3 Household members, Education and Health (mem.dta)

The member section contains variables on member characteristics, education and their health status. Many variables in these sections shall be consistent with the previous panel wave (i.e. wave 6). You do not flag out for implausible or missing cases those can be replaced by using information from the panel wave 6 or edit from survey comments.

Checking rules

1. Variable 21022a is the first question that we asked household respondent to state whether each member is still belonging to the household during the reference period. If this variable equal to 1 or 2, the following variables shall be recorded according to the skip instruction. If the person passed away during the reference period, we skip to question 14 of section 2.1.
2. There should be only one household head in each household. ID of respondent and household head (data file hh.dta) should be consistent with variable 21001 in section 2.
3. Variables 21014 and 21015 should match with information in section 4.2, 5 and 6; and also with education sub-section if household members are students. However, if a person has more than 2 occupations, the third one may not be recorded in 21014 and 21015 but information can be found in section 4.2, 5 or 6. Merging variables 42001, 50002a – 50031, 60002a - 60029 from sections 4, 5 and 6 is

---

[13] The sections are sequenced following the TVSEP household questionnaire 2017. That is why the numbering of sections is mixed.

recommended. For example, if the household respondent reported that his daughter who is currently working in a computer company in Bangkok, but job information of this person does not exist in section 5. Please flag out for call back households.

Following the 2017 enumerator survey guideline, if one household member (Mr. A) has a business and another household member (Ms. B) also works in this family business, two cases can be identified:

- A and B are business partners or husband and wife, and B does not receive regular salary. Please record code H = 3 in v21014. However, only one roster for this business will be recorded in section 6.
- If B received salary during the reference period that means B works as a wage-employee. Therefore, code H can be equal to 5 or 7 in v2014.

However, if one of them or both have more than 1 occupation; enumerators took the working period into account to determine the main and the second occupations.

4. Check for the duration of schooling, year started/year finished assessing plausibility and whether the member was ever absent for more than one year.
5. Cross check variables 23012 with variables 21014 and 21015.
6. In cleaning II, merging with the final member dataset from 2016 is recommended. Information from wave 6 can be used for correcting implausible or missing values in some variables but not all.
7. In this section, two variables (i.e. _x21020 and _x21021) are part of household income. Although there is no clear rules how much in kind or in cash that household would receive from migrants, nevertheless please check outliers by using summary statistics or box plot. If extreme value exists it could be because of digit errors or any other reasons e.g. children bought a house for their parent in the rural area. In this case the value of the house will not be part of income calculation.

Income variables

Remittances received from absent household members are fully considered as source of household income.

Step 1 calculating received income per household member (_x21080) for all cases that HH received transfer from absent HH members

```
gen _x21080 = _x21020
```

Step 2 Aggregate remittances per household (_x10080)

```
bysort hhid: egen _x10080 = total(_x21080)

label var _x10080 "Remittances received from absent HH members"

collapse (count) _x12100 = _x21001 (mean) _x10080 (sum) _x21022 (first)
QID , by(hhid)
```

Other output files from the member section

Member section contains several key variables that are needed for merging with other sections during the cleaning and analysis process. All these key variables shall be generated in the beginning of cleaning I and again after completion of cleaning II. Name this data file as mem_Id and save it in the sub-folder w7_v1 and w7_v2, respectively.

_x21001 (HH member Id) is the identification number of each household member and always the same during the panel waves. Due to some programming errors of the electronic questionnaire, the household member Id shall be revised at the first step before start cleaning I. Please use the revised HH member Id from section member to correct member Id that exist in section shock, wage-employment, non-farm self-employment, public transfer and insurance.

_x21022 (nucleus household membership) is a dummy variable equal to 1 for any person (including infants less than six months old) who spent at least 180 days in the household during the reference period. Any HH members who absent more than 180 days (including the HH head) is excluded from the calculation of per capita income and consumption.

_x12122 (HH size nucleus) is generated from the sum of variable _x21022. This variable will be used in calculating per capita income and consumption. Please generate it directly after checking variable 21016

Section 2.4 Household dynamics (hhdyn.dta)

Checking rules

1. Check plausibility of number of events and sum of payments do not seem to be out of place.
2. Check if any amount has not already been occurring in 2.1.In section 2.4, it should be payments for/from non-household members only!
3. Similar to the remittances from absent HH members that there is no clear rules how much households would receive money or in-kind from non-household members. However, please check the summary statistics of variables 24012 and 24013 that should not be extremely out of range. If that is the case, find out whether there is any error that can be revised in cleaning II (i.e. digit errors).

Income variables

Remittance from non-household members was asked only for the total of monetary and in-kind remittances. Therefore a distinction in remittance of in-kind and cash is not possible.

Step 1 Generate positive net remittances (_x24081)

The net income from remittance from non-household members is calculated by deducting the value of remittance the household sent (_x24013) from the value of remittance that household received (_x24012).

Step 2 Remittances at the household level (_x10081)

The remittance at the household level will be used in the household income aggregate. It is generated from summation of all individual positive net remittance by household.

```
egen _x10081=total(_x24081), missing by(hhid)
```

Section 8 Household expenditures (exp.dta)

Checking rules

Food items

1. For food items that purchased quantity is recorded, convert it into per capita quantity which requires merging variable *nucleus household size* from the data file mem_id.dta. Identify outliers (only the upper bound) by using the program flag outlier and plausibility rules.
2. Generate price of each food item from reported value divided by the quantity purchase and check outliers. Always clarify the unit for example THB/1000 VND per kilogram or per piece.
3. If there is no expenditure for rice (or other staples, cross check with section 4.1 (in-kind payment from land), section 4.2 (home consumption variable), and section 2.4 (receive rice from non-household members). Missing observations will be replaced in the consumption aggregate.
4. Please note that the average amount spent for each item is recorded per month.

Non-food items:

1. Generate per capita expenditure of each non-food item by considering only household nucleus member (any persons who spent at least 180 days in the household during the reference period and check whether values are implausible.

Transportation and communication:

1. Please note that expenditure for item 127 (maintenance for car and motorbike) and 128 (Insurance and fee for car and motor bike) is recorded in the yearly basis.

Education:

1. Cross check with section 2.2 (education) whether there is any student. How many students are in the household? And what grade are they currently enrolled in?
2. Use the flag outlier program (starts with mean+/- 2SD) to flag implausible cases in combination with the plausibility rules whether the reported educational expenditure is reasonable.
3. For education fee, pocket money and meal, convert it to daily expenditure per head e.g. divided by 200 school day and number of students in household. Compare this value with the average daily wage in that province; the daily expenditure per capita of this item should not be higher than the daily wage.
4. It can also happen that household had educational expenditures but no child is going to school reported in section 2.2. This is because we do not have a code for the pre-school grade in section education. However, this is part of household expenditure. During the survey, enumerators were instructed to record the expenditure in section 8 and make comment as "pre-school cost". In this regards, checking or merging survey comments will be useful. If there are more than 100

cases of pre-school cost, please report to the TVSEP HQ whether the new item code of expenditure shall be added in the cleaning process as well as for the next survey wave.

Health:

1. Use the flag outlier program to flag implausible value. Cross check with section 2.3 for extreme outlying cases.
2. If there is no household members who were sick during the reference, but health expenditure is recorded, please cross check with the age because they may buy some supplementary but not report any health impairment.

Social:

1. Use the flag outlier program to flag implausible value. Cross check with section shock and survey comments for outlying observation whether the implausible value is acceptable or not.

Consumption variables

The expenditure section impacts on all consumption sub-aggregates except for housing that the reported- or estimated hedonic housing value will be used. Following the enumerator guideline, if less than half of the items in each sub-group (i.e. food, non-food, transport and communication etc.) are not recorded (for example, respondent could not answer), the total expenditure of this sub-group will be asked. This implies that enumerator still records expenditure for items that household can answer. Another case is household could only answer the total expenditure of each sub-group but not for individual item. Additional survey comment "HH cannot separate expenditure by each item" will be added. Therefore, it requires a data cleaner to check whether reported sub-total or sub-total of single items are the same of different. If total expenditure of a certain expenditure group is higher than the sum of the respective sub-expenditure, total expenditure is used for consumption aggregate. Otherwise the sum of the sub-expenditures is used. Celebrations and funerals expenditure will be dropped[14] from the consumption aggregate variables (data file hc_exp[15]) but not in the expenditure dataset (data file exp.dta[16]).

---

[14] Following Deaton and Zaidi (p.32), "More complex is the case of lumpy and relatively infrequent expenditures such as marriages and dowries, births, and funerals. While almost all households incur relatively large expenditures on these at some stage, only a relatively small proportion of households are likely to make such expenditures during the reference period typically covered by the survey. Ideally, we would want to smooth these lumpy expenditures, spreading them over several years, but lacking the information to do so - which might come, for example, by incorporating multi-year reference periods for such items - we recommend leaving them out of the consumption aggregate."

[15] Consumption variables in this file will be used for aggregate.

[16] Data file of the expenditure section (exp.dta) will include all expenditure data recorded during the survey.

Steps of cleaning in expenditure do-file after preliminary check (only in cleaning II)

1. Check variable _x80001
   1.1 Check whether _x80001 is equal to variable Id roster or not (Id). If not, replace answers in _x80001 from Id Roster. If the Id Roster is code 90 91 and 92, variable _x80001 is equal to "." Replacement with answering codes from the Id roster is also required.
   1.2 If codes 90, 91 and 92 in variable _x80001 can be recoded to any existing items codes, please do it at this step. After recoding _x80001, change the value to _x80002 or _x80004 , respectively. If the value of the item is in monthly basis, divide the value by 12. Please also renumber _x80000 after recoding _x80001.
   1.3 Due to recoding codes 90-92 to existing codes, duplications occur in some households. Please sum up the quantity and expenditure value of the duplicate cases by hhid and expenditure item. Be aware that the reported value can be not the same for the same household and item. In this case, you do not drop duplicates.
   1.4 Recode code 91 and 92 in variable _x80001 to code 90.
2. Check variable _x80002, _x80003 and _x80004 by item
3. Calculate subtotals if missing
4. Calculate annual expenditures
   For 2017 data, the list of items purchased on the monthly basis is:

```
* Transform Food, Non-Food, Trans/Comm to annual values
foreach x of numlist 1/15 121 131/133 17/23 25 27/29 126 {
      // only monthly codes from Questionnaire
      *dis "Item Id `x'"
      replace _x80004 = 12 * _x80004 if _x80001 == `x' & _x80004 <
.     // times 12 months
}
*
```

5. Rectangularize dataset
6. Generate consumption aggregate components
7. Save cleaned full dataset of expenditure section in w7_v2
8. Generate consumption variables at the HH level
9. Save consumption variables (hc_exp.dta) in w7_v3

Section 3.1a and 3.1b Shocks (shock.dta)

The dataset includes reported shock events from section 3.1a and 3.1b which can be distinguished by event Id (__31001). The event from shock 3.1b starts with 4 even if the household stated less than 3 shocks in section 3.1.a.

Check variable 31005b and 31006a grouping by shock type and rank by the severity. The average of income or asset loss value of the same shock type that has high severity on HH should be higher than the lower severity.

Please be careful of overlapping extra expenditure, income and asset lost.

Examples:

1. Pickup truck was heavily damaged before the accident it was worth 300.000THB/200.000.000VND due to the damage it had to be repaired. The repair cost was 50.000THB/30.000.000VND. In spite of the repair the car was not in the same condition as before the accident and is worth 200.000THB/165.000.000VND — so in 31006a you would enter 100.000THB/70.000VND for the loss of asset and for 31005b you would enter 50.000THB/35.000VND for the repair.

2. In case of floods/ droughts of rice fields the household incurs income loss (including home consumption and sale) or extra expenditures;

   ▪ if rice is very young and the HH replanted: cost of replanting is entered in 31005b
   ▪ if rice is ready to be harvested ask respondent to estimate the difference between the expected yield and the actual yield: income loss due to shock. Enter this into 31005a

3. Storm destroys rubber tree

   ▪ replant = extra expenditures and loss of income for reference period and no asset loss;
   ▪ Household does not replant = loss of asset and loss of income

4. Job loss: income loss and extra expenditures for looking for new job

5. Respondent reports that his shop burned down over night: loss in net income (profits) under house damage shock). No extra precautions were taken; did not re-build shop in the reference period so no extra expenditure, only asset loss (value of shop + equipment).

6. You are a registered farmer selling organic products, the buyer finds out that the quality standard is not met, so the farmer has to sell at a lower price: income loss

7. The government decided to increase the tax on fertilizer. Therefore the farmer's expenditures for fertilizer increased compared to what he had expected, add these extra expenditures to 31005b

If income, expenditure and asset lost are overlapping and it can be edited in cleaning II, we will not call back households for this case. Only extremely loss will be flagged out.

Section 3.1f Aspirations (aspir.dta)

1. Cross check with section member how many there are children under the age of 16 belonging to the household.
2. If any under age 16 years old children are missing in this section, flag out for Plist but not call back. Leave missing cases in cleaning II.
3. If all under age 16 years old children are recorded but some questions have missing, please only flag out for plist and remarkable in cleaning II but not calling back households

Section 3.2 Risks (risk.dta)

1. Check variable 32003 and 32003a by grouping by type of risk event. Enumerators often mix up between impact of risk on household income and assets.
2. For variable 32015, please use the flag outlier program to detect implausible value and flag out for remarkable in cleaning II

Section 4.1 Land (land.dta)

Checking rules

41003   keep local unit during cleaning. Detect implausible value with the flag outlier program and boxplot, separated between homestead and other main land use. Often missing cases occur due to skip instruction, therefore cross-checking between variables _41004 - _41011h is required.

41005   For homestead (41004=1), answer in this variable should be code 11 -14. If it is recorded with any other codes (eg. 1 -9), flag these case for revision in cleaning II but not for calling back.

41009a   Check land price per (rai / 1000 m2) by using the program flag outlier and information from the provincial statistical office

41006 – 11h   Check missing and implausible cases following the skip instruction in the questionnaire. Only flag out problematic cases for plist and remarkable for cleaning II, but not for calling back.

In case that a household rent out a land parcel to a relative and receive no payment, enumerator record either code 1-cash or code 98 in 41011a and zero in 41011h. For example, household reported that they rented a homestead land from relatives but does not have to pay a rental fee. Enumerators put 13 (if house owned) or 14 (if house rented) in Q5 and in Q10a put code 1 or 98 and make a comment "rent from relatives and no payment" and put zero "0" in Q11h. In cleaning II replace code 1 or 98 with code 4 (no payment) in variable _x41011 and leave _x41011 as zero.

41011h   Check rental rate per (rai / 1000 m2) by using the program flag outlier and information from the plausibility table. Please note that this variable may contain zero value if household did not receive/pay any rental fee. The average rental rate per unit area shall be generated without this case.

41013   distance of each land parcel from homestead. Enumerators sometimes mix up between km. and meter, please run the descriptive of this variable to detect implausible cases. The principle rules are

•   If the parcel is located in the same village, maximum distance is 5 km. All outliers will be replaced with 2.5 km.

•   If the parcel is in the same district, maximum distance is 20 km., replace outliers with 10 for the case of same district (rural area) and with 5 for same district (urban area)

•   If the parcel is located in the same province, maximum distance is 50 km. replace outliers with 25 km for the case of same province (rural area) and 12.5 km for urban area.

•   Other location e.g. Bangkok, HCMC. We leave the reported value.

If household owns the plot but rents out, the skip instruction does not work properly. Enumerators were instructed to record as follows:

Q4 put code 2 (rented out)

Q5 put code 90, specify "HH owns plot and rents out; type of land document; how was the land obtain"

So you would ask the type of land document of the house and fill in for other, specify (Q5) and then ask the value of the land and fill in in other, specify (Q5) in the format as stated previously. In this way Q10a and Q11h will show up and enumerators can record the rented out information. If the household rents out, values for Q10a and Q11h should still be provided

In cleaning I, you put back information from the code 90 text variable into the respective uncleaned variables and replace implausible or missing cases in cleaning II if possible.

Income variables

Income from the land section is calculated for those parcels rented out (i.e. _x41005==5 and _x41004==2 | _x41003a==8)

Step 1 calculate income per parcel (_x41083)

```
egen _x41083 = rowtotal(_x41083c _x41083k _x41083ck)
```

_x41083c: in cash income from land rent by parcel

_x41083k: in-kind income from land rent by parcel

_41083ck: mixed in cash and in-kind income from land rent by parcel

Step 2 Income from all parcels that household rented out is summarized to the household level

```
egen _x10083 = total(_x41083), missing by(hhid)
```

Cost of land rent for agricultural purpose (_x41096)

Step 1 cost of land rent for agricultural purpose per parcel (_x41096) is the sum of rental payment (in cash and in-kind) that household paid to rent in the land for all agricultural purposes

Step 2 Sum up the expenditure of land rent per household

Cost of land rent for business purpose (_x41096a)

Step 1 cost of land rent for business purpose per parcel (_x41096a) is the sum of rental payment (in cash and in-kind) that household paid to rent in the land for the business purpose (41004 = 9)

Step 2 Sum up the expenditure of land rent per household

Consumption variables

In-kind rent payments from renting out land parcels are part of food, as well as of rice consumption. The question that specifies type of in-kind payment has been dropped out since 2016 survey. If variable 41010a equals to code 2 (in kind) or 3 (cash and kind), please merge the text comment variable to identify type of in-kind payment[17].

---

[17] Detailed commands can be found in the do-file named "hc_land".

Section 4.2 Agriculture (crop.dta)

Checking rules

1. In variable _x42003a (crop variety) If code 90 was used, check if the code already exists. If you see the variety is existing in the code list, please recode it; otherwise leave it as code 90.

2. Quantity unit (_x42009): the quantity unit is very important for checking quantity balance and crop yield. Kilogram or Ton is mostly preferred. Check also code 90 text variable whether it can be recoded to any existing unit in the code list. Before checking quantity balance, the new variable named "factor unit (funit)" will be generated to be used to convert original unit to kilogram.

3. Yield by crop type (_x42010/_x42005): check yield per type of crop using the flag outlier program and compare with information from the plausibility table.   If outlier is low yield, please firstly check digit errors either in variable 42010 (total production) or planted area (42005), otherwise cross check with the shock section and survey comments whether household reported any effect from shock events on low yield. If there is no shock effect, all outlying observations (if possible) will be imputed using replacement rules (mean starting at the village level or written impute program). For the case of high-yield outliers, check digit errors from variables 42010 and 42005. The revised yield variable will be used to calculate the new total production variable (_x42010).

   In Vietnam (especially in Dak Lak) yield might be low because enumerators recorded area of the whole plot for the case of intercropping. For example if coffee is planted with avocado in a 1 hectare parcel, enumerators entered 1 hectare for coffee and 1 hectare for avocado in the variable 42005. In this case please recalculate the new planted area for both crop using the share of 75 percent of the area for coffee and 25 percent for the shaded trees.

4. Variables 42015a and 42015b specify type of product/quality sold in each observation. If the product undergoes simple processing such as cutting, chipping and drying, there is always weight loss involved. As a result, taking the final product into the balance will not work out. For those cases where the transformation rate is known, (e.g. from paddy rice to milled rice the conversion is around 70 %) we will use the fresh weight equivalent to calculate the balance. If the product type are different between column 15a and 15b, the product type has to be equalize based on the lager amount either sale1 or sale2. For rice, use paddy as the base. For other product e.g. cassava and Para-rubber use the main product that has main sale volume. We calculate everything standardize base product.

Table 9: Example of equalized quantity of sale A and sale B

	Paddy	Mill
Sold quantity	100	100
Reported sale price	7	14
Sold value	700	1400
Conversion paddy to mill	1	0.7
Final quantity used in x42014 and x42014b	100	142
Price	7	1400/142 = 9.8
Final sold value	700	980
Share	0.41	0.59
Weighted average price	(7*0.41) + (9*0.59)	

5. Check quantity balance

Total production x42010 should be **equal or more than** the sum of 42011, 11b, 12, 13, 13a, 14, 14b. The quantity unit (42009) must be the same for all respective components.

If the total production 42010 is <u>less than</u> the sum of 42011, 11b, 12, 13, 13a, 14, 14b, please browse those cases to determine error patterns. Some values are clearly logically impossible due to digit error. In cleaning I these errors shall be firstly revised in the auxiliary variables of 42010 – 42014b to avoid flag out unnecessary cases for the plist. For example, you generate the cleaned variables of 42010 – 42014 b with the prefix _x and replace the correct value in the _x variables. All these commands can be later used in the cleaning II. For the sake of cleaning I, it helps you to decide that you will flag out the cases for calling back household or only for replacement in cleaning II. If digit error does not exist, the quantity difference will be deducted proportionally from each component in cleaning II. However, yield (total production divided by planted area) per type of crop should be firstly checked.

If the reported total production 42010 is <u>greater than</u> the sum of 42011, 11b, 12, 13, 13a, 14, 14b. If the sum is outside plausibility range, check for digit error. Then add the difference multiply with share to allocate the quantity difference in each component 42011, 11b, 12, 13, 13a, 14, 14b (do in the auxiliary variables for cleaning I), see example in table 9.

Table 10: Example of proportional quantity added in each component

Variable (420…)	10	11	11b	12	13	12a	13a	14	14b
Reported data	110	50	10	10	20	5	5	0	0
Share (%)		5	1	1	2	0.5	0.5	0	0
After Adding 10 kg different	110	55	11	11	22	5.5	5.5	0	0

If HH only reported the total production in variable 42010, but no information in variable 11-14b.

- Rice: check household expenditure for rice and nucleus household member size. If the rice expenditure is zero, we assume that HH produce rice for home consumption. Replace x42011 with the quantity from 42010. But if there is expenditure reported in section 8, we do not add this quantity in any variable from 11-14b (leave the column missing)
- In some cases, type of crops and total production clearly indicate that this is only for home consumption for example household harvested 2 kg of mango, in this case we add this quantity to variable x42011 and leave the other components (11b-14b) with zero
- For other crops those have no further information how HH allocate this total production quantity, we leave variable 11b-14b with missing (invalid-missing) but use the total production quantity to calculate income

6. Crop price will be checked by type of crop and quality sold. Technically the quality group of each crop type (cropXqualityA and cropxqualityB) will be generated. Therefore variable _x42015a and _x42015b shall be firstly revised if needed or replaced if missing. Unit of price must be clear for example per kilogram or per tonne etc. Apply the flag outlier program to check implausible price and use the plausibility rules as orientation if the program detects more than 2 percent of total observation. Price variables (_x42016 and _x42016b) can be missing because household did not sell their product during the reference period. However, price variables are still needed for calculating other in-kind received or consumed value. Missing price will be replaced with the generated mean price (without outlier cases) calculated at the village level (if possible, otherwise move to the nearest location level i.e. commune, sub-district).

7. Apply the flag outlier program and information from the plausibility table to check expenditure of each activity by crop type. The expenditure shall be converted into expenditure per unit area. Concerning price of fertilizer and daily wage, information from the plausibility table will be used as reference.

8. Total expenditure is the sum of expenditure from land preparation (42018), seedlings (42020), fertilizer 42023, pesticides (42025a, 42025b, 42025c), machinery cost for harvesting (42027), irrigation (42029), other activities (42036) and labour cost (42019, 42020, 42021, 42022, 42024, 42026, 42028, 42029)

   For perennial crops, we will not include the investment cost (from land preparation and seedling) those occur during the reference period and out of period in the total expenditure.

   Check total expenditure per unit area by major crop type again using the flag outlier program to detect implausible cases.

Income variables

a) Gross income from crops is the sum of home consumption value (hcVal) + give away value (gaVal) + household processing value (hpVal) + in kind payment (kpVal) + sale value (salesValA and salesValB)

b) Values of home consumption, give away, household processing and in-kind payment are calculated from the respective quantity component (_x42011, _x42011b, _x42012 and _x42013) multiplied with the weighted average sale price (saleprw).

c) The weighted average sale price (saleprw) is the share of sale price A and sale price B.

d) Net income = gross income – total expenditure (exclude investment cost).

Consumption variables

Consumption value is calculated by multiplying home consumption quantity with the weighted average sale price. The section contributes to food and non-food consumption. Special consumption value (i.e. rice consumption, alcohol and tobacco) is also generated[18].

---

[18] Detailed commands can be found in the do-file named "hc_crops".

Section 4.3.1 Livestock and aquaculture (livest.dta)

Checking rules

1. Balancing stocks (not apply for insects/fish/seafood)

Errors from switching column and digit errors are likely to occur often in this section which results in unbalance of stock. . If this is the case, number of stock and value from switching column shall be revised at the first step in cleaning II[19]. Generate a new variable named "stockatend" (stock at the end) = _x43103 + _x43104 + _x43105 - _x43106 - _x43107 - _x43108 - _x43118 and compare this variable with the reported stock at the end (_x43109). In cleaning I, please flag out if the stock does not balance.

2. Check price of each component by type of stocks using information from the plausibility table as reference.

    Price of stock at the beginning = _x43103a/_x43103
    Sale price = _x43110/_x43108
    Purchasing price = _x43105a/_x43105
    Home consumption price = _x43107a/_x43107
    Price of stock loss = _x43106a/_x43106
    End price = _x43109a/_x43109

    Principle rules to replace implausible price in cleaning II are:

    2.1 Small animal (chicken, duck and pig fattening)

- If price at the beginning is greater than the sale price, replace the beginning price with the average purchasing price at the provincial level generated from the survey data without outliers
- Not every household would sell their stocks. If there is no sale price, please check outlier of the beginning price by comparing with the statistical price from the plausi table. If the beginning price is implausible, replace it with the average purchasing price at the provincial level generated from the survey data without outliers.
- If the sale price is lower than minimum farm gate price, replace sale price with the average sale price at the provincial level generated from the survey data without outliers.
- If household has a breeding pig and it gives birth to 10 fattening pigs, but not all these 10 pigs are breeding pig, the piglets are put in variable 43105 instead of variable 43104. In cleaning II, data cleaner should switch the information from variable 43105 to 43104.

---

[19] For cleaning I, revising number of stock and value from switching column and digit errors shall be done in the auxiliary variables.

2.2 Big animal (buffalo, beef cattle, dairy cattle etc)

- If there is no change in stock (i.e. no born, purchase, sale, loss and home consumption) and end price is less than the begin price, replace the end price with the begin price.
- If there is no born and purchase, but there is sale and the end price is less than the begin price, replace end price with the begin price. There are two cases for the sale price under this condition:
  a) Sale price is equal or greater than the beginning price. This case does make sense. Replacement is not required.
  b) Sale price is less than the beginning price, replace sale price with the average farm gate price from the plausibility table.

- If end price is less than the beginning price and
  a) no change in stock, the end price can be lower than the beginning price but the end stock value shall be greater than the beginning value.
  b) no born and purchase, but there are changes in disposals and the sale price is greater than the beginning price. Only replace end price with the beginning price.
  c) Sale price is less than the opening price, replace sale price with the average farm gate price from the plausibility table.

- End price is equal to the begin price but the sale price is less than the beginning price and no born and purchase, replace sale price with the average farm gate price from the plausibility table
- Sale price is less than the begin price and there is no change in stock and the stock at the end, replace sale price with the average farm gate price from the plausibility table.

2.3 Home consumption price should not be greater than the sale price, if this is the case, replace home consumption price with the sale price.

4.4 Please note that we are checking the price of each addition and disposals component. After correction do not forget to convert back the value by multiplying with the unit of each component.

3. Generate cost per unit (only for livestock and poultry) and using the flag outlier program to check implausible cases. For extreme outlier cases, please flag out for the plist in cleaning I, otherwise replace implausible value following the standard procedure.

Income variables

The livestock gross income (revenue) is calculated as sale value plus value of home consumption, in-kind payment and give away plus value of stock at the end and minus the sum of purchasing/receiving value and value of stock at the beginning of the accounting period (FAO, 1985).

Step 1 income from livestock and aquaculture per activity

Income ($\_x43185$) = total revenue ($\_x43134$) – total cost ($\_x43135$)

Total revenue ($\_x43134$) = ($\_x43110$ + $\_x43107a$ + $\_x43109a$ + $\_x43118a^{20}$) – ($\_x43103a$ + $\_x43105a$)

Total cost ($\_x43135$) = restocking ($\_x43113$) + feeding ($\_x43114$) + veterinary treatment ($\_x43115$) + hired labor cost ($\_x43116$)

Please check summary statistics of net income per type of livestock especially if the net income is extremely negative. This occurs often for the case of hen that household only sold eggs but not the hens. In this case, the final total net income would come from the livestock product.

Step 2 generate income from livestock at the household level

```
egen _x10085a = total(_x43185), by(hhid)
```

Consumption variables

The section contributes to food consumption. Variable home consumption value (43107a) will be used to generate food consumption value from stocks[21]. Please drop animals that can be not consumed i.e. silkworms.

---

[20] $\_x43118a$ is the value of input used for own business It is calculated from unit of use ($\_x43118$) multiplied with the sale price. This is considered as input items through a business process. Therefore it is included in the calculation of total revenue from livestock. The value self-produced input will be accounted in the calculation of income from self-employment by subtracting it from the total revenue of the respective business.

[21] Detailed commands can be found in the do-file named "hc_livest".

Section 4.3.2 Livestock products (lstprod.dta)

Checking rules

1. Check for a correct balance 43204 = 43205 + 43206+43218
2. Check consumption and sale price of each livestock product whether they are realistic. Please use the information from the plausi table as reference
3. Home consumption quantity can be checked based on the number of nucleus household members whether the recorded quantity is realistic.
4. Cost of inputs shall be converted to cost per unit of livestock product.

Income variables

Step 1 net income from livestock product (_x43285)

Total revenue (_x43234) = home consumption (_x43205a) + sold value (_x43207) + input for own business (_x43218a)

Total cost (_x43235) = packaging (_x43208) + hired labor (_x43208a) + other cash cost (_x43210)

Net income (_x43285) = total revenue (_x43234) – total cost (_x43235)

Step 2 net income from livestock products at the household level

```
egen _x10085b = total(_x43285), by(hhid)
```

Consumption variables

The section contributes to food and non-food consumption. Variable 43205a is used to generate the consumption variables. Non-food livestock products are dung and silk[22].

---

[22] Detailed commands can be found in the do-file named "hc_lstprod".

Section 4.4 Natural resource extraction (hunting.dta)

Checking rules

1. The sum of quantity sold and home consumption cannot be greater than the reported total output
2. Check home consumption price per type of product extracted using the information from the plausi table as reference.

Income variables

Step 1 income per activity of natural extraction (_x44086) = total output value (_x44017a) – payment for access in cash or kind per year (_x44005) – fuel cost (_x44011) – hired labour cost (_x44012) – other costs (_x44012a)

Step 2 income from natural extraction at the household level

```
egen _x10086= total(_x44086), by(hhid)
```

Consumption variables

The section contributes to food and non-food consumption. Timber products, fire wood and gem constitute the non-food production. Consumption value is computed by multiplying the quantities used for home consumption with the plausible sold price[23].

---

[23] Detailed commands can be found in the do-file named "hc_hunting".

Section 5 Wage employment (offempl.dta)

Checking rules

1. Cross check with variables 21014 and 21015 in section member whether information of household members who work as wage employees are recorded in this section. If not, please flag out for team leader to decide whether to call back.
2. Check outlier of net wage in cash (50022) by type of job using the program flag outlier in combination with the information from the plausibility table. Please note that net wage in cash can be recorded in different time unit (50023). If the wage is extremely out of the range, cross check with other variables e.g. 50006, 50007, 50009a. If the out of range value is reasonable based on these information, we leave it. Otherwise calling back household or replacing outlier with the average wage generated from the data (without outlier) will be applied to correct the outlier.
3. Before calculating income variables, please correct the household member id using output file (mem_Id) from the member section.
4. Specific note for Vietnam: If a household member has a labour contract, the in-kind benefit could be high because it also includes the insurance payment that his/her employer pays directly to an insurance company.

Income variables

The income from wage employment is calculated by multiplying net wage in cash by the number of months worked plus in-kind benefit. Only income from nucleus household member will be included in the household income.

Step 1 calculate income from wage employment per household member (only nucleus household members are included)

_x50087c =(i_x50022*i_x50028a)*(i_x50029) if _x50023==2	daily wages
_x50087c = i_x50022*i_x50029 if _x50023==4	per month
_x50087c = i_x50022 if _x50023==5	per year
_x50087c = i_x50022 if _x50023==6	lumpsum

_x50087k = _x50031

_x50087 = _x50087c + _x50087k

Step 2 aggregate income from wage employment per household

```
egen _x10087 = total(_x50087), by(hhid)
```

Section 6 Non-farm self-employment (selfempl.dta)

Checking rules

1. Cross check with variables 21014 and 21015 in section member whether information of household members who own a business or a partner are recorded in this section. If not, please flag out for the team leader to decide whether to call back households.
2. The rate of return should be reasonable e.g. the rate of return more than 20 percent. If HH invested 10000 THB but get 50 THB, this would be outlier. The total revenue divided by total cost including fixed cost (investment/service life) if that ratio is higher than 3:1, that is outlier.
3. Generate daily wage = monthly payroll (60021) / number of employee (60016b) / average number of day worked per month (60040a). The reported wage should not exceed more than 3 times of the minimum daily wage from the plausi table.

Income variables

Step 1 income from self-employment (consider only if a business belongs to a nucleus household member)

Average monthly profit/loss in the business (_x60038) = average volume sale per month (_x60029) – average monthly payroll (_x60021) – costs for input purchase per month (_x60033) – average monthly value of self-produced input (_x60043)

Cash income per year (_x60088c) = Average monthly profit/loss in the business (_x60038) * number of month engaged in this business (_x60039)

In-kind income (_x60088k) = in-kind income (_x60088i) + home consumption (_x60088h)

Net income (_x60088) = cash income (_x60088c) + in-kind income (_x60088k)

If there is a big loss, it can be that household overestimated the cost. If variable cost is higher than the revenue then check digit error, wage. If the cost is higher than the revenue, then there must be a shock event. Otherwise, we would accept the case.

Step 2 aggregate income from self-employment at the household level

```
egen _x10088 = total(_x60088), by(hhid)
```

the cost of land rent for business purpose will be deducted during the income aggregate to obtain the net income from self-employment.

Consumption variables

The section contributes to food, non-food and rice consumption. Please use variable _x60005a to identify type of consumption. The business type namely rice mill, restaurant, foodstall operator, butchery, other small scale food processing contribute to food consumption. Only in the case of rice mill, home consumption is rice wherefore rice consumption is solely impacted by this business type. All other business types are source of non-food consumption[24].

Section 7.1A Borrowing (borr.dta)

Checking rules

1. Check reported interest rate by source of loan and compare with information from the plausibility table.
2. In this section, checking groups of variables are required according to the following conditions:
   a. If HH reported fully repaid (71133 = 1), no value shall be recorded in 71119b.
   b. If HH reported fully repaid (71133 = 1), total repayments of loan as of end 4/17 (71119a) should be equal to or greater than the amount of loan (71105)
   c. The first year that household borrow from this lender (71125) can be only equal to or less than the year that household received the loan (71111)
   d. The total amount of payments (repayment + interest) between 5/16 – 4/17 (71115a) is less than or equal to the total repayments of loan as of end 4/17 (71119a)
   e. If the interest rate equal to zero (71106 = 0) then the amount of loan (71105) will be equal to the sum of total repayment and remaining debt as of end 4/17 (_x71105==_x71119a+_x71119b)
   f. If the sum of total repayment and remaining debt as of end 4/17 (_x71105==_x71119a+_x71119b) equal to the amount of loan (71105), the interest rate is equal to zero.
   g. If the interest rate is greater than zero percent (71106>0% / cint > 0), the amount of loan (71105) is less than or equal to the sum of total repayment and remaining debt as of end 4/17 (_x71105==_x71119a+_x71119b).

      calculated annual interest rate (cint in %) = ((_x71119a+_x71119b)/_x71105)-1)/(_x71126)*100

3. Replacement for missing of interest rest: If the loan is from formal financial institutions (commercial banks, government banks), then use the average interest rates as replacements. The interest rates can be found from the Bank of Thailand or from the financial institutions

---

[24] Detailed commands can be found in the do-file named "hc_selfempl".

4. Replacement for loan duration: For each type of lender, use the village-average duration as replacement. If the number of non-missing observations in a given village is less than 5, use the sub-district average as the replacement. If the number of non-missing observations at the sub-district level is still less than 5, continue to higher administrative level

5. Note for Vietnam: Due to translation to Vietnamese, in variables 71115a, 71115b and 71119a, repayment was translated as repayment of principal whereas the amount of interest. Furthermore if household only paid back interest rate, time of repayment were not recorded. If you find these cases, replace 71119a with (Interest rate/month* Borrowed amount*Number of month they paid the interest if 71119a equal to 0.

Income variables

Step 1 generate cost of loan for productive purpose per each loan

cost of loan for productive purpose ($\_x71197$) = -i$\_x71130$ * i$\_x71129$

interest payment in reference period (-i$\_x71130$) = calculated annual interest rate (%) (i$\_x71127$) /100 * amount of loan ($\_x71105$)

Productive purpose of loan (i$\_x71129$) = 1 yes fully, = 0.5 partially, = 0 no.

** Check descriptive of variable$\_x71197$ whether it does make sense or not compared with the loan size

Step 2 aggregate cost of loan for productive purpose per household

Consumption variables

Consumption from the borrowing section originates from in-kind loans which household received during the reference period. The values of these loans are added to food consumption[25].

---

[25] Detailed commands can be found in the do-file named "hc_borr".

Section 7.2A Public transfers (trans.dta)

Checking rules

1. Check total value of transfer received by type of program. Use the information from the plausibility table as reference

Income variables

Step 1 Public transfer received per household member (consider only nucleus household member)

$$\_x72193 = \text{cash transfer } (\_x72193c) + \text{in-kind transfer } (\_x72193k)$$

Step 2 Aggregate income from public transfer received per household

Section 7.2 B Insurance (insur.dta)

Checking rules

1. Check total amount of compensation payment received by type of insurance. Use the information from the plausibility table as reference

Income variables

Step 1 income from compensation payments received per household member (consider only nucleus household member)

Step 2 aggregate income from indemnity payments received at the household level

Consumption variables

Consumption from insurance originates from insurance premier paid during the reference period. All types of insurance except livestock and crop insurance are consumption relevant[26].

---

[26] Detailed commands can be found in the do-file named "hc_transf".

Section 6.2 Investment and Disinvestment

Investment and disinvestment can be for farm, business and personal use as well as household use. The section refers to the investment or disinvestment strategy for productive and consumptive purpose above a value of 5000 THB/1.5 Million VND. Therefore cross checking with section land, livestock (purchasing or sale value of livestock at time) and asset section must be done in cleaning I. If any information is missing either from this section or the relevant sections, please flag out for calling back.

Section 9.1 Household wealth (assets.dta)

Checking rules

1. Switching column and digit errors often occurs in this section.
2. Cross check with section investment for items that is less than 1 year old and has value more than 5000 THB/1.5 million VND. It should be recorded as investment
3. In variable 91003 respondents were asked to give the value of the most recently obtained item at the time when household bought or got it. Even household received item as a gift, the value of this item shall be estimated. If it is recorded as zero or missing, please flag out for calling back during the cleaning I.
4. Replacement procedure will be applied if value of asset items cannot be clarified from calling back.

In the tablet program, if an asset is brought /obtained during the reference period, the age of this asset is less than one year. However, the program does not allow decimal in the variable 91004, so enumerators put 0 in this variable and comment, the number of months until 4/2017. Data cleaner should put the information from comment back in x91004 (reported month /12).

In addition to cleaned data set, the depreciation of productive assets and consumption value of value of durable goods are generated at the end of cleaning II.

Depreciation of productive assets

Step 1   Calculate net present value of assets (_x91003npv) = the value of the most recent obtained item (_x91003) * fcpi

fcpi is the factor of consumer price index. It takes the average of CPI every year between 1976 – 2017 (for the case of wave 7), using 2017 as the base year the fcpi is set to 1.

Step 2   Generate Depreciation rate

= -0.1 if an item is mostly for business use (91008a =1, 10% depreciation)

= -0.5 if an item is business and private use (91008a =2, 5% depreciation)

= 0 if an item is mostly private use (91008a =1)

Step 3   depreciation of productive assets (_x91098) = _x91003npv * depreciation rate * _x91002 (number of item that household own)

Step 4   Aggregate deprecation of productive assets at the household level

```
egen _x10098 = sum(_x91098), by(hhid)
```

Consumption value of durable goods

Step 1 Generate consumption rate

    = 0 if an item is mostly for business use

    = -0.05 if an item is business and private use

    = 0 if an item is mostly private use

Step 2 Consumption value of durable good per item (_x91108) = _x91003npv * consumption rate * _x91002 (number of item that household own)

Step 3 Aggregate consumption value of durable goods at the household level

```
egen _x11008 = sum(_x91108), by(hhid)
```

Section 9.2 Housing conditions (house.dta)

Checking rules

1. House value shall be converted into price per square meter and check implausible value by the flag outlier program in combination with the plausibility rules
2. The missing and outlying observations (if not possible to revise) of housing value will be estimated using the hedonic regression

Consistently with Deaton and Zaidi (1999, p. 23) rent equivalents constitute the housing part of consumption. The equivalents have been computed as 2 percent of the reported plausible housing value (92016) or by regressing hedonic model for the case of implausible /missing housing value to determine the housing value, inter alia, information from the housing conditions.

IX. Income aggregate

The literature dealing with the calculation of income aggregates from LSMS surveys is very limited. Some guidance is available from the – often patchy – documentation of existing survey data (e.g. National Statistical Office, 2008) or concrete code programmed for specific surveys (Anonymous, 1991, 1994). For our present purpose we calculate the income aggregate for the 7th wave of the Thailand and Vietnam rural household sample according to the definitions and procedures suggested by Johnson et al. (1990).

The principles used in income calculation are as follows:

a) Remittances received: Those remittances from non-household members to the household are fully considered as household income. As the questionnaire asks only for the total of monetary and in-kind remittances, a distinction in the income aggregate of in-kind and cash payments is not possible.

b) Income from rents: The income from renting out agricultural or other land in cash and in kind is calculated separately and in total. In kind payments are valued at the farm gate price obtained from the receiving household's price information in the crop section if it was available. Otherwise, the mean of farm gate prices given for the commodity by at least five households from the village, sub-district or district level was used.

c) Income from agriculture: A separate income estimate was calculated for crops and livestock. The former takes the total value of output (output x farm gate price), irrespective of home consumption.

Further assumptions underlying the calculation procedure:

a) Household income is calculated on an annual basis in local currency (in Vietnam in 1000 Dong, in Thailand in THB). TVSEP offers variables converted into 2005-based PPP $.

b) The calculation is based on the nucleus household definition as defined in the following sub-section 1.2: Incomes accruing to household members in their function as head of business, as an employee or beneficiary of insurance or government transfer payments are accounted for as income. Incomes accruing to persons not included in the nucleus household definition are ignored for calculating household income.

c) Missing values in any component of the income have been imputed by different procedures described in the previous chapter. In case it was not possible to sensibly replace a missing value by an imputed one, the component remains missing and consequently the whole aggregate is undefined.

The calculation of total household income follows a procedure involving four steps:

1) The income contribution of each table row in the questionnaire is calculated (e.g. gross margin by crop, income by off-farm employment activity). At this stage, the household membership of the respective income-earner is considered for the decision whether the returns from a certain activity belong to household income. This step is done during the cleaning II process.

2) Secondly, the rows of each table are aggregated to calculate total gross income per household by source (e.g. gross margin from crop production minus land rent paid for agricultural purposes, income from off-farm employment, gross income from non-farm self-employment minus land rent paid for business purposes).

```
rename _x10084 _x10084x              /* crop income */
rename _x10088 _x10088x              /* business income */

gen _x10084=_x10084x-_x10096         /* deduct land rent */
label var _x10084 "Income from crops "

gen _x10088=_x10088x-_x10096a        /* deduct land rent */
label var _x10088 "Income from self-employment "
```

3) Thirdly, the cost of consisting of the cost of outside capital for productive purposes (i.e. interest on loans) is assigned to income components as a deduction component[27].

```
egen _x10100 = rowtotal(_x10080 _x10081 _x10083 _x10084 _x10085
_x10086 _x10087 _x10088 _x10092 _x10093 _x10094 _x10097)

label var _x10100 "Net income deduct interest payment (excl.
housing value)"
```

4) Finally, all income components are aggregated into total annual household income (netInc2). Per capita income per day is derived from this variable by division by the number of nucleus household members and number of days during the reference period. The latter variable is a pure headcount and does not consider any other weighting by age group or scale parameters.

```
local inc netInc1[28] _x10100
foreach x of varlist `inc' {
   gen dcap_`x' = `x'/_x12122/365
   lab var dcap_`x' "per capita daily income generated from
`x'(THB/1000VND)"
   }
*
```

---

[27] For panel wave 1-5, imputed depreciation of productive assets (_x10097) is assigned to income components based on their share in absolute total gross income and deducted. Housing value (_x10082) was also included in the household income aggregate.

[28] Aggregate net income deduct interest payment includes housing value (for comparison during calculation of income aggregate).

X. Consumption aggregate

In cleaning step II the relevant consumption variables were checked for missing and outlying observations. Generally, it was tried to change as few of these observations as possible. Thus, not all observations that were identified as outliers have been treated. This procedure is in line with Deaton and Zaidi (1999, p.25) who state that it might be "unclear whether the 'outlier' is genuine or not" and that "the analyst must make a judgment that balances the desirability of keeping any reasonable number of observations against the risk of contaminating the aggregate." Concerning the replacement of missing and outlying observations the standard replacement procedure applied for the calculation of the income aggregate has been adopted: "In most cases the mean of each variable with sufficient cases and plausible information for the nearest possible level of sampling (village, commune, district and province) was used for replacement.

The consumption variables generated during the cleaning II are merged in the Stata file named "consagg.dta"[29]. Total consumption equals the sum of food, non-food, and housing related consumption. Also, consumption related to durable goods is included in the consumption aggregate[30]. Interest payments from the borrowing section have been dropped from the analysis, as well, because according to Deaton and Zaidi (1999), p.32: "…interest payments should be excluded from the consumption aggregate…" Consumption from the wage-employment section is excluded due to the nature of adjusted question since 2016 which not possible to distinguish between food and non-food consumption values.

The consumption aggregate has been restricted to nucleus household members, i.e. per capita consumption is household consumption per nucleus household member. Nucleus household members are the ones who stayed in the respective household for at least 180 days during the reference period. Consumption from self-employment and insurance is only included if the head of business and the insurant, respectively, are nucleus household members. Additionally, Adult Equivalence Scale (AES) is generated based on age and gender specific food requirements using 2100 calories as reference adult (WHO, 1985) which allows us to calculate the aggregate and sub-aggregate consumption per AES household size.

---

[29] Detailed STATA commands for merging consumption variables can be found in consagg.do

[30] This classification follows Deaton and Zaidi (1999) who also include consumption of durable goods in their guidelines for consumption aggregates.

```
* Store Calorie Groups
local scale1      0    820   820
local scale2      1    1150  1150
local scale3      2    1350  1350
local scale4      3    1550  1550
local scale5      5    1850  1750
local scale6      7    2100  1800
local scale7     10    2200  1950
local scale8     12    2400  2100
local scale9     14    2650  2150
local scale10    16    2850  2150
local scale11    18    3000  2100
local scale12    30    2900  2150
local scale13    60    2450  1950


* Create Scale
gen calorie=.                    // Set Calorie levels by gender/age group
foreach x of numlist 1/13 {
    tokenize `scale`x''
    replace calorie=`2' if _x21004>=`1' & _x21003==1     // male
    replace calorie=`3' if _x21004>=`1' & _x21003==2     // female
}
gen AES_WHO = calorie/2100
lab var AES_WHO "WHO Adult Equivalence Scale"
```

The education and health sub-aggregates are stated separately in order to allow for education and health related analysis. There is a separate rice sub-aggregate (which is part of the food sub-aggregate) because rice constitutes the main source of nutrition in Thailand and Vietnam. Also, especially rice consumption has been treated extensively during the cleaning process wherefore it was decided to generate a separate sub-aggregate. Assuming every household had positive food and rice consumption, reported food and rice consumption of zero was imputed using the information of the lowest available geographic entity.

References

Bezawit Beyene Chichaibelu, Thi Hoa Vu, Lena Hohfeld and Hermann Waibel (2013). Cleaning Guideline Household survey 5th wave 2013 Dataset description and procedure for calculating household income. Unpublished manuscript. Hannover.

Deaton, Angus and Salman Zaidi (1999). Guidelines for Constructing Consumption Aggregates For Welfare Analysis, Working Papers 217, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton.

FAO (1989). Farm management glossary. Food and Agriculture Organization of the United Nations, Rome, Italy.

Johnson, Martin, Andrew D. McKay & Jeffrey I. Round (1990): Income and expenditures in a system of household accounts - Concepts and estimation. Working Paper No. 10; The World Bank, Washington

World Health Organization (WHO) (1985). Energy and Protein Requirements, Technical Report Series 724. Geneva: World Health Organization

Appendix A: Structure of Do- and Data Files
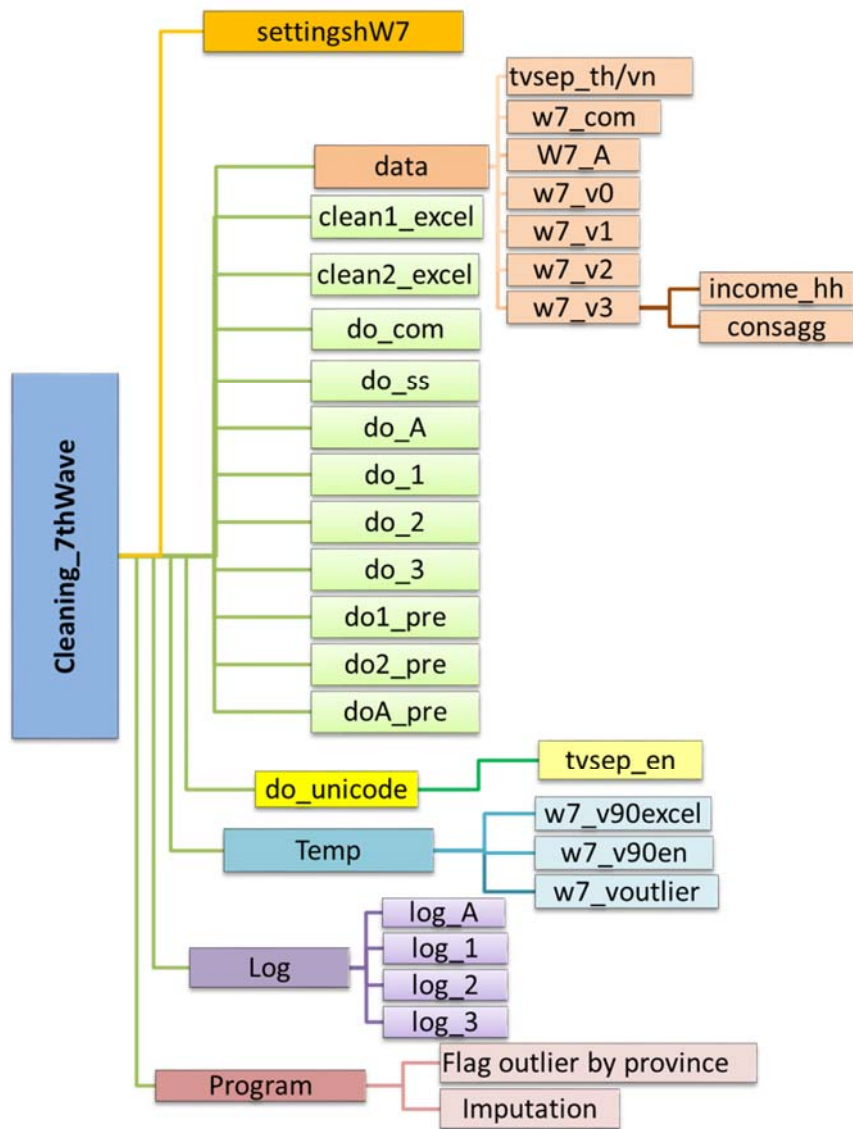
Appendix B: Name of data files during cleaning I and II

Data files	Interview comments (folder w7_com)		File of income variables at HH level (folder w7_v3/income_hh))	File of consumption variables at HH level (folder w7_v3/consagg)
	Full comment	Last comment		
mem	com_mem	t_mem	mem_hh	
hhdyn	com_hhdyn	t_hhdyn	hhdyn_hh	
exp	com_exp	t_exp		hc_exp
shock	com_shock	t_shock		
aspir	com_aspir	t_aspir		
risks	com_risk	t_risk		
land	com_land	t_land	land_hh	hc_land
crops	com_crops	t_crops	crops_hh	hc_crops
livest	com_livest	t_livest	livest_hh	hc_livest
lstprod	com_lstprod	t_lstprod	lstprod_hh	hc_lstprod
hunting	com_hunting	t_hunting	hunting_hh	hc_hunting
offempl	com_offempl	t_offempl	offempl_hh	
selfempl	com_selfempl	t_selfempl	selempl_hh	hc_selfempl
credrat	com_credrat	t_credrat		
defhist	com_defhist	t_defhist		
borr	com_borr	t_borr	borr_hh	hc_borr
sav	com_sav	t_sav	sav_hh	
transf	com_transf	t_transf	transf_hh	hc_transf
insur	com_insur	t_insur	insur_hh	hc_insur
invest	com_invest	t_invest		
disinvest	com_disinvest	t_disinvest		
assets	com_assets	t_assets	assets_hh	hc_assets
house	com_house	t_house		hc_hedonic
hh	com_hh	t_hh		

Appendix C Plausibility rules for Thailand, 2017 survey

Section 8. Household Expenditures, Prices of consumed items

Table C1: average unit consumption per person and retail price per unit by expenditure items

Exp ID	Commodity	Unit	Average (unit/person)	Price (THB/unit) Min	Price (THB/unit) Max	source of quantity	source of price	remarks
1	Rice	kg	8.5	22	43		http://www.moc.go.th/index.php/rice-iframe-8.html	Milled Jasmine Rice
121	Other staple Food*	kg	0.3	28	40		http://www.taladsimummuang.com/dmma/Portals/PriceListItem.aspx?id=050016010	*We assume its noodles; price per pack converted into kg
2	Beef/pork	kg	0.6	120	400	http://www.acfs.go.th/document/download_document/FCDT.pdf		beef (5% and pork(95%)
3	Fish	kg	0.4	50	130		http://www.moc.go.th/index.php/rice-iframe-2.html	We assume the major fish is Tilapia
4	Poultry	kg	0.4	65	85			Chicken (70%); Duck (30%)
5	Eggs	piece	20	2.4	3.7			Chicken (70%); Duck (30%)
6	Vegetable	kg	30	N/A	N/A	2016 wave		
7	Fruit	kg	35	N/A	N/A	2016 wave		
8	Food ingredients, spices (include Salt/Sugar)	kg	30	N/A	N/A	2016 wave		
9	Fermented fish	kg	0.02	37	150		http://www.taladsimummuang.com/dmma/Portals/PriceListItem.aspx?id=050016030	

Table C1: average unit consumption per person and retail price per unit by expenditure items (cont.)

Exp ID	Commodity	Unit	Average (unit/person)	Price (THB/unit)		source of quantity	source of price	remarks
				Min	Max			
10	Oil	litre	0.3	38	130		http://www.moc.go.th/index.php/rice-iframe-7.html	
11	Beverages	litre	2000	N/A	N/A	2016 wave		
131	Milk powder (0-1 year)	gram	489	0.92	3.0		http://www.central.co.th/e-shopping/milk-powders-price-check/	
12	take home and eat out	THB	800	N/A	N/A	http://www.nso.go.th/sites/2014/Pages/สำรวจ/ด้านสังคม/รายได้รายจ่ายครัวเรือน/ภาวะเศรษฐกิจและสังคมของครัวเรือน.aspx		
13	Alcohol,	THB	211	N/A	N/A	http://ubon.nso.go.th, http://buriram.old.nso.go.th http://nkphanom.old.nso.go.th		
132	cigarettes	THB	112	N/A	N/A			
133	coffee and tea	THB (n/a for kg.)	140	N/A	N/A	http://www.nso.go.th/sites/2014/DocLib13/ด้านสังคม/สาขารายได้/เศรษฐกิจสังคมครัวเรือน/44/3.ตารางสถิติในรูปแบบ%20Excel/T5.4xls.xls		
14	Other food	THB	12000	N/A	N/A	2016 wave		
15	Total Food	THB	43724	N/A	N/A	2016 wave		

Table C1: average unit consumption per person and retail price per unit by expenditure items (cont.)

Exp ID	Commodity	Unit	Average (unit/person)	Price (THB/unit)		source of quantity	source of price	remarks
17	Personal care supplies	THB	300*	N/A	N/A	http://ubon.nso.go.th http://buriram.old.nso.go.th http://nkphanom.old.nso.go.th		
18	Clothes, shoes and bags, accessories	THB	600	N/A	N/A			
19	Detergent/washing powder	THB	300*	N/A	N/A			
20	Electricity	electricity unit	275	3.2	3.8	http://www.nso.go.th/sites/2014/DocLib13/ด้านเศรษฐกิจ/สาขาพลังงาน/การใช้พลังงานของครัวเรือน/สำรวจการใช้พลังงานของครัวเรือน_2558/7.%20สรุปผลที่สำคัญ.pdf	http://www.erc.or.th/ERCWeb2/Upload/News/26512017095154เอกสารประชาสัมพันธ์ค่าไฟฟ้าและค่าFtพ.ค.-ส.ค.60.PDF	
21	Water cost	water unit	9	10.2		http://www.pwa.co.th/contents/service/table-price		
22	Liquid propane gas/charcoal	THB	170	N/A	N/A	http://www.nso.go.th/sites/2014/DocLib13/ด้านเศรษฐกิจ/สาขาพลังงาน/การใช้พลังงานของครัวเรือน/สำรวจการใช้พลังงานของครัวเรือน_2559/Pocket%20Book_พลังงาน2559_A5.pdf		Exp. of Gas used in households + exp. of Charcoal and firewood
23	Hair dresser and other beauty service	THB	5000			2016 wave		
130	Total tax payments	THB	90	N/A	N/A	http://www.nso.go.th/sites/2014/Pages/สำรวจ/ด้านสังคม/รายได้รายจ่ายครัวเรือน/ภาวะเศรษฐกิจและสังคมของครัว		

							รือน.aspx		
25	Total Non-food	THB	43724	N/A		N/A	2016 wave		

Table C1: average unit consumption per person and retail price per unit by expenditure items (cont.)

Exp ID	Commodity	รายการ	Unit	Average (unit/person)	Price (THB/unit)		source of quantity	source of price	Remarks
126	Fuel for car and motorbike	ค่าเชื้อเพลิงสำหรับรถยนต์และ/หรือมอเตอร์ไซค์	THB	1000	N/A	N/A	http://www.nso.go.th/sites/2014/DocLib13/ด้านเศรษฐกิจ/สาขาพลังงาน/การใช้พลังงานของครัวเรือน/สำรวจการใช้พลังงานของครัวเรือน_2559/Pocket%20Book_พลังงาน2559_A5.pdf		
27	Public transportation	ค่าเดินทางโดยใช้บริการรถสาธารณะ	THB	1400	N/A	N/A	https://www.m-society.go.th/article_attach/19286/20656.pdf	http://apps.dlt.go.th/ltmb/index.php?zone_id=2	
28	Telecommunication (include mobile credit)	ค่าการติดต่อสื่อสาร(โทรศัพท์บ้านและมือถือ/ไปรษณีย์)	THB	565	N/A	N/A	http://www.nso.go.th/sites/2014/Pages/สำรวจ/ด้านสังคม/รายได้รายจ่ายครัวเรือน/ภาวะเศรษฐกิจและสังคมของครัวเรือน.aspx		
29	Other transportation, communication	ค่าใช้จ่ายอื่นๆที่เกี่ยวกับการเดินทางและการสื่อสาร	THB	67000	N/A	N/A	2016 wave		
127	Maintenance for car	ค่าซ่อมบำรุงรัก	THB	3000	N/A	N/A	http://www.manager.co.		

Exp ID	Commodity	รายการ	Unit	Average (unit/person)	Price (THB/unit)		source of quantity	source of price	Remarks
	and motorbike	หารถยนต์และมอเตอร์ไซค์					th/daily/viewnews.aspx?NewsID=9560000034316		
129	Fee for car and motorbike	ค่าธรรมเนียมสำหรับรถยนต์และรถจักรยานยนต์	THB	3000	N/A	N/A			
30	Total Transportation & Communication	รวมค่าเดินทางและการติดต่อสื่อสาร	THB		N/A	N/A			

Table C1: average unit consumption per person and retail price per unit by expenditure items (cont.)

Exp ID	Commodity	รายการ	Unit	Average (unit/person)	Price (THB/unit)		source of quantity	source of price	Remarks
31	Education fees, books	ค่าเทอม	THB/year	60000	N/A	N/A			
32	Student's dress/uniform	ค่าเครื่องแบบ/แต่งกาย	THB/year	19000	N/A	N/A			
33	Tuition fee	ค่าเรียนพิเศษ	THB/year	80000	N/A	N/A			
34	Rental fee(Dorm, apartment)	ค่าเช่าหอหรืออพาร์ทเมนท์	THB/year	60000	N/A	N/A	http://www.moe.go.th/moe/upload/news18/FileUpload/38879-8038.pdf		
35	Other costs of schooling	ค่าใช้จ่ายอื่นๆที่เกี่ยวกับการศึกษา	THB/year	10000	N/A	N/A			
36	School bus	ค่ารถโรงเรียน	THB/year	60000	N/A	N/A			
37	Pocket money and meal	เงินที่ให้บุตรหลานไปโรงเรียน รวมถึงค่าอาหารก	THB/year	10000	N/A	N/A			

		ลางวัน							
39	Total education	รวมค่าใช้จ่ายด้าน การศึกษา	THB		N/A	N/A	http://www.nso.go.th/sites/2014/Pages/สำรวจ/ด้านสังคม/รายได้รายจ่ายครัวเรือน/ภาวะเศรษฐกิจและสังคมของครัวเรือน.aspx		
40	Medicine (Purchase in Pharmacy only)	ค่ายา(เฉพาะที่ซื้อจากร้านขายยา)	THB	85	N/A	N/A			
41	Doctor fee	ค่าหมอ	THB	140	N/A	N/A			
42	Other health	ค่าใช้จ่ายอื่นๆ ที่เกี่ยวกับสุขภาพ	THB		N/A	N/A			
45	Total Health(incl. Health expenditures later refunded by insurance)	ค่าใช้จ่ายรวมที่เกี่ยวกับสุขภาพ (รวมค่าใช้จ่ายด้านสุขภาพซึ่งหักลบในส่วนที่ประกันชีวิต/ประกันสุขภาพออกให้)	THB		N/A	N/A			

Table C1: average unit consumption per person and retail price per unit by expenditure items (cont.)

Exp ID	Commodity	รายการ	Unit	Average unit/person	Price (THB/unit)		source of quantity	source of price	remarks
46	Celebrations and funerals in own household	ค่าใช้จ่ายในการจัดงานรื่นเริง/งานพิธีต่างๆ ของบุคคลในครัวเรือน (เช่น งานแต่งงาน งานบุญ งานศพ งานสู่ขวัญ)	THB	230	N/A	N/A	http://www.nso.go.th/sites/2014/Pages/สำรวจ/ด้านสังคม/รายได้รายจ่ายครัวเรือน/ภาวะเศรษฐกิจและสังคมของครัวเรือน.aspx		
47	Donations (to temples, social organizations,	การบริจาค/ทำบุญ (ให้วัด ให้องค์กรทางสังคม/ก	THB	260	N/A	N/A			

#	Item (EN)	Item (TH)	unit	min			source		
	schools)	ารกุศล ให้ทุนการศึกษา)							
48	Recreation and entertainment	สันทนาการและความบันเทิง	THB	310	N/A	N/A			
49	Lottery	ล็อตเตอรี่/หวย	THB	3840	N/A	N/A	http://www.gambling study-th.org/issues_topic_1/285/1/1/%E0%B8%B4Lottery-consumers/		
56	Other gambling expenditures	ค่าใช้จ่ายๆเกี่ยวกับการพนันอื่นๆ	THB		N/A	N/A			
60	Bribery/corruption-police	การติดสินบน/คอรัปชั่น – ตำรวจ	THB		N/A	N/A			
61	Bribery/corruption-government officer	การติดสินบน/คอรัปชั่น – เจ้าหน้าที่รัฐฯ	THB		N/A	N/A			
62	Bribery/corruption-business partner	การติดสินบน/คอรัปชั่น – หุ้นส่วนทางธุรกิจ	THB		N/A	N/A			
55	Total social	รวมค่าใช้จ่ายด้านสังคม	THB		N/A	N/A			

Section 4.1 Land

Table C2: land price per square metre and rental payment rate per rai

Item	unit	min	max	source
Land price	THB/square metre	50	110000	http://research.terrabkk.com/th/article/detail/138548
Rental payment rate for doing rice (North-eastern area)	THB/rai/year	800	1200	http://www.cabinet.soc.go.th/doc_image/2558/993141705.pdf

Section 4.2 Crops

Table C3: Yield, total cost and farm gate price

Crops	Yield (kg/rai)			Total Cost (THB/kg)	Price (THB/kg)	remark	source
	min	average	max	average	average		
13 – Glutinous rice (ข้าวเหนียว)	150	400	1400	7.0	10.0		1) https://www.thairath.co.th/content/364848
101 - Jasmine rice (ข้าวเจ้า)	250	600	1200				2) http://www.oae.go.th/download/download_journal/2560/commodity59.pdf
6-Cassava (มันสำปะหลัง)	2500	3600	10000	1.5	2.0		3) http://www.arda.or.th/kasetinfo/south/para/controller/01-02.php
701-Sugar cane (อ้อย)	9000	12000	25000	0.7	0.9		4) https://www.bot.or.th/thai/attachment/commodityreport-august.pdf
10-Para rubber (ยางพารา)	150	240	350	65.0	49.0		5) http://www.oae.go.th/ewtadmin/ewt/oae_baer/download/article/article_20141013102903.pdf
25-Oil palm (ปาล์มน้ำมัน)	2300	3100	6200	2.0	5.0		
903-Garlic (กระเทียม)	700	1000	2000	25.0	77.0		
3-Peanuts (ถั่วลิสง)	200	270	550	19.0	45.0		
21-Fodder maize(ข้าวโพดเลียงสัตว์)	500	750	1500	6.0	8.0		
4-Soybeans (ถั่วเหลือง)	200	300	600	15.0	15.0		
5-Mungbeans (ถั่วเขียว)	50	120	250	20.0	32.0		
8-Kenaf (ปอ)	100	150	300		7.0		

Table C3: Yield, total cost and farm gate price (cont.)

Crops	Yield (kg/rai)			Total Cost (THB/kg)	Price (THB/kg)	remark	source
	min	average	max	average	average		
422-Mulberry (ต้นหม่อน)	3000	5000	10000	3.0	10.0		1) http://www.oae.go.th/download/download_journal/2560/com
1301-Banana (กล้วย)	2500	3500	7000	5.0	13.0	*กล้วยหอม	

Crops	Yield (kg/rai) min	Yield average	Yield max	Total Cost (THB/kg) average	Price (THB/kg) average	remark	source
1311-Coconut (มะพร้าว)	500	800	1600	4.5	11.0		modity59.pdf
1315-Mango (มะม่วง)	900	1300	2600	6.0	36.0		2) http://www.arda.or.th/kasetinfo/south/para/controller/01-02.php
1319-Longan (ลำไย)	600	825	1600	13.0	38.0		
1320-Orange (ส้มเขียวหวาน)	1400	1900	3800	13.0	47.0		
1321-Pomelo (ส้มโอ)	1000	1400	2800	9.0	36.0		3) https://www.bot.or.th/thai/attachment/commodityreport-august.pdf
1323-Rambutan (เงาะ)	800	1150	2300	12.0	34.0		
11-Coffee (กาแฟ)	90	120	240	60.0	68.0		
412-Pepper (พริกไทย)	300	500	1000	96.0	360.0		4) http://www.oae.go.th/ewtadmin/ewt/oae_baer/download/article/article_20141013102903.pdf
414-Tea (ชา)	500	730	1500	34.0	70.0	2 types of tea	
901-Shallot (หอมแดง))	1500	2000	4000	15.0	27.0		
902-Onion (หอมหัวใหญ่)	3000	4300	9000	7.0	13.0		
908-Cucumber (แตงกวา)	3500	5000	10000	8.5	10.0		
914-Yard long bean (ถั่วฝักยาว)	1000	1500	3000	9.5	30.0		
33-Tobacco(ยาสูบ)	150	255	500	10.0	54.0		
26-sweet potato (มันเทศ)	2500	3500	7000	9.0	12.0		
1304-Dragon fruit (แก้วมังกร)	8200	11000	22000	3.0	45.0		
1314-Tamarind (มะขาม)	2200	3000	6000	9.0	35.0		
932-Angled gourd (บวบ)	1500	2000	4000	9.0	25.0		
946-Bitter gourd (มะระ)	1800	2500	5000	6.0	60.0		
905-Pumpkin (ฟักทอง)	3500	5000	10000	3.0	8.0		
915-Waxgourd (ฟักเขียว/แฟง)	3000	4000	8000	4.0	5.0		

Table C3: Yield, total cost and farm gate price (cont.)

Crops	Yield (kg/rai)			Total Cost (THB/kg)	Price (THB/kg)	remark	source
	min	average	max	average	average		1) http://www.oae.go.th/download/download_journal/2560/commodity59.pdf,
920-Cabbage (กะหล่ำปลี)	1000	1500	3000	5.0	9.0		
921-Chinese chive (กุยช่าย/ผักแป้น)	2000	3000	6000	4.0	20.0		
941-Chilli (พริก)	1500	2500	5000	8.0	300.0		2) http://www.arda.or.th/kasetinfo/

904-Galgant (ข่า)	1500	4000	8000	30.0	60.0	*case study	south/para/controller/01-02.php
942-Eggplant/Aubergine (มะเขือเปราะ/ยาว)	5000	7000	14000	7.0	150.0		3) https://www.bot.or.th/thai/attachment/commodityreport-august.pdf,
1307-Watermelon (แตงโม)	1500	2500	5000	4.0	30.0		
909-Sweet basil (ใบโหระพา)	1500	2000	4000	12.0	15.0		4) http://www.oae.go.th/ewtadmin/ewt/oae_baer/download/article/article_20141013102903.pdf,
912-Chinese mustard (กวางตุ้ง)	1500	2500	5000	2.5	10.0		
913-Chinese kale (คะน้า)	1500	2500	5000	2.5	10.0		
924-Lemongrass (ตะไคร้)	2000	3000	6000	2.0	6.0		
936-Coriander (ผักชี)	1000	1500	3000	5.0	100.0		
938-Morning glory (ผักบุ้ง)	2000	3000	6000	3.0	26.0		

Section 4.3.1 Livestock

Table C4: Purchasing price, production cost and sale price by type of livestock

Livestock	Purchasing price (THB/unit)	Production cost (THB/kg)	Sale price				source
			THB/kg		THB/1 unit		
	average	average	min	max	min	max	
							1) http://extension.dld.go.th/th1

							source
Buffalo	-	-	80	100	25000	40000	/images/stories/economic/costprice20171020.xls
Beef cattle	-	-	90	100	20000	36000	
(bull) weaning – 5 yrs	-	-	40	65			2) http://breeding.dld.go.th/dairy/index.php/dairy-price
(bull) > 5 yrs	-	-	30	45			
(cow) weaning – 8 yrs	-	-	45	75			
(cow) > 8 yrs	-	-	30	45			3) http://breeding.dld.go.th/dairy/dairy_web/law_of_dairy/images/price_for_sell.pdf
Dairy cattle	-	-					
(dairy bull) every age	-	-	25	70			
(dairy cow) weanig – before pregnant	-	-	40	70			4) http://www.oae.go.th/download/download_journal/2560/commodity59.pdf
(dairy cow) pregnant	-	-	-	-	15000	28000	
(dairy cow) give milk <15 kg./day	-	-	-	-	14000	28000	
(dairy cow) give milk > 15 kg./day	-	-	-	-	18000	35000	5) http://www.oae.go.th/download/index_table/Table3.XLS
Pig(fattening)	1900	60	54	70			
Pig(piglet production)	1900	-	45	70			
Goat	-	-	100	140			6) http://www.oae.go.th/economicdata/retailani.html
Chicken	15	35	30	45			
Fighting cock	450	-	500	3000			7) https://www.gaichon.com/price.html
Duck	25	-	60	80	120	150	

Section 4.3.2 livestock product

Table C5: Production cost and sale price by type of livestock product

Livestock product (unit)	Production cost(THB/unit)	Sale price (THB/unit)	source

	minimum	average	maximum	minimum	average	maximum	
1 Milk(litre)		15			18		http://extension.dld.go.th/th1/i
4 Chicken Eggs (pieces)		-			3		mages/stories/economic/costpr
5 Duck eggs (pieces)		-			4		ice20171020.xls

Table C6: Production cost and sale price by type of livestock product

Livestock product (unit)	Production cost (THB/unit)			Sale price (THB/unit)		
	minimum	average	maximum	minimum	average	maximum
4 - Chicken egg (pieces)	0	0.1	1	1	3	5
5 - Duck eggs (pieces)	0	0.05	0.1	3	4	6
6 - Silk (kg)	0	0	0	1300	1500	2500
8 - Silkworm pupae (kg)	0	3	40	60	100	130
26 - Dung (kg)	0	0.1	2	0.05	1	3

Source: TVSEP 2016 Data

Section 4.4 Natural resources extraction

Table C7: General retail price (THB/kg) by items

Item	Price (THB/kg)		source
	minimum	maximum	
catfish	70	90	http://www.moc.go.th/index.php/rice-iframe-3.html
tilapia	60	80	

snakehead fish	90	150	
mushroom	35	350	https://www.kasetprice.com/ราคา/เห็ด/วันนี

Table C8: General retail price (THB/kg) by items

Item	Price (THB/kg)		
	minimum	average	maximum
1 - Tilapia	45	75	100
2 - Catfish	10	80	200
3 - Snakehead fish	20	80	200
4 - Shrimp	80	90	100
5 - Mudfish	20	50	100
10 - Other marine fish/ sea fish	50	80	100
12 - Timber products	1	5	20
13 - Fire wood	1	5	110
15 - Mushrooms	10	150	600
16 - Wild plants	5	30	200
17 - Firewood for making charcoal	1	8	120
100 - Frog and toad	20	105	200
101 - Rat	120	140	160
102 - Mollusk	10	15	20

Table C8:   general retail price (THB/kg) by items (cont.)

Item	THB/kg		
	minimum	average	maximum
103 - Red ants eggs	20	200	300
106 - Fruits		50	
107 - Crab	10	30	50

110 - Bird	30	135	240
111 - Carp	20	50	100
112 - Vegetable and bamboo	6	30	120
113 - Other insects	50	175	300
116 - White fish	5	50	150
117 - Other fish	2	60	215

Source: TVSEP 2016 Data

Section 5 Wage employment

Table C9: Wage by gender, educational level and type of economic activity

items	Wage (THB / Month)		source
	min	max	

By gender	Male	10000	11500	
	Female	11000	13000	
By education	Vocational or equivalent	9100		
	Higher vocational or equivalent	9900		
	Bachelor degree or equivalent	12000		
	Master degree or equivalent	34000		
By Economic activity	Mining and quarrying	9200	25000	
	Manufacturing	9000	22000	
	Electricity, gas and water supply	11000	24000	
	Water supply; sewerage, waste management and remediation activities	8500	22500	
	Construction	9200	24000	http://service.nso.go.th/nso/web/statseries/tables/40000_Northeastern_Region/2.7.xls
	Wholesale and retail trade; repair of motor vehicles and motorcycles	9000	23000	
	Land transport, warehousing and support activities for transportation	8500	20000	
	Accommodation and food service activities	9100	25000	
	Information and communication	9600	16700	
	Financial and insurance activities	9200	20000	
	Real estate activities	9700	25000	
	Professional, scientific and technical activities	9000	21500	
	Administrative and support service activities	9150	21500	
	Private hospital activities	9000	57000	
	Arts, entertainment and recreation	9300	26500	
	Other service activities	8400	21000	

Section 7.1. Borrowing and Lending

Table C10: Interest rate by banks from May 2016 – April 2017 (% per year)

Bank ธนาคาร/แหล่งเงินกู้	Deposit Rates		Loan Rates		remarks	source
	min	max	min	max		

Bank					remarks	source
51 Government Housing Bank ธนาคารอาคารสงเคราะห์	0.75	2.00	6.25	7.25		http://www2.bot.or.th/statistics/ReportPage.aspx?reportID=223&language=th
52 Bank for Agriculture and Agricultural Cooperatives (BAAC) ธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร (ธ.ก.ส.)	0.50	1.75	5.00	5.00		
53 Government Savings Bank ธนาคารออมสิน	0.50	2.25	6.50	7.10		
55 Export-Import Bank of Thailand or Business Promotion Office at Department of Export Promotion ธนาคารเพื่อการส่งออกและนำเข้า แห่งประเทศไทย (ธสน.) หรือ สำนักส่งเสริม ธุรกิจบริการของกรมส่งเสริมการส่งออก	0.25	0.90	6.25	18.00		
56 Urban Community Development Organization สำนักงานพัฒนาชุมชนเมือง การเคหะแห่งชาติ	n/a	n/a	1.00	18.00		http://www.villagefund.or.th/uploads/gallery/image_big_5a83bccf24cd7.pdf
57 Agricultural cooperatives สหกรณ์การเกษตร	n/a	n/a	8.50	10.00		1)http://www.coopthai.com/sumrong/ 2)http://www.coopthai.com/ubon/news_detail.php?NID=7
60 Village Fund/Community Fund(Taksin village fund) กองทุนหมู่บ้าน 1 ล้านบาท	n/a	n/a	1.00	12.00		https://www.parliament.go.th/ewtadmin/ewt/parbudget/ewt_dl_link.php?nid=105
63 pawnshop โรงรับจำนำ	n/a	n/a	0.25*	1.25*	* %/month	http://www.pawn.co.th/question_a.php
64 Commercial bank ธนาคารพาณิชย์	0.05	2.50	6.03	23.10		http://www2.bot.or.th/statistics/ReportPage.aspx?reportID=223&language=th

Section 7.1. Borrowing and Lending

Table C10: Interest rate by banks from May 2016 – April 2017 (% per year) (cont.)

Bank ธนาคาร/แหล่งเงินกู้	Deposit Rates	Loan Rates	remarks	source

Project						source
65 Credit companies (e.g. Easy Buy, Quick Cash, AEON etc.) บริษัทสินเชื่อ (เช่น Easy Buy, Quick Cash, อีออน ฯลฯ)	n/a	n/a	24.00	28.00		1)http://www.aeon.co.th/aeon/services-loans 2)http://www.sabaikrapao.com/สินเชื่อเงินสดควิกแคช/
79 poverty eradication project โครงการแก้ไขปัญหาความยากจน	n/a	n/a	0.00	0.00	no interest rate	http://chiangrai.cdd.go.th/services/kkkj
80 student loan fund กองทุนเงินกู้ยืมเพื่อการศึกษา	n/a	n/a	1.00	1.00		https://www.studentloan.or.th/index.php/highlight/2/6/46
82 saving cooperative and credit union สหกรณ์ออมทรัพย์และเครดิตยูเนียน	0.35	3.30	4.25	6.25		1)http://www.cultthai.coop/new/images/interrest/2559/int_loan_230759.pdf 2)http://www.cultthai.coop/new/images/interrest/2559/dep230759.pdf

Section 7.2 Public transfer

Table C11: Government support project and support system

Project	Support system	Money support	THB/unit term	source

study for free 15 years (นโยบายเรียนฟรี 15 ปี อย่างมีคุณภาพ)	1) tuition fee	a. formal education
		1) kindergarten
		2) primary level
		3) secondary
		4) high school
		5) vocational
		  (industrial technician)
		  (commercial)
		  (food)
		  (art)
		  (general Agricultural)
		  (applied Agricultural)
		b. Non-formal education
		1) primary level
		2) secondary
		3) high school
		4) vocational
			1700
			1900
			3500
			3800
			
			6500
			4900
			5500
			6200
			5900
			11900
			
			1100
			2300
			2300
			4240
				THB/person/year	http://www.moe.go.th/moe/upload/news20/FileUpload/8646-9873.pdf

Table C11: Government support project and support system (cont.)

Project	Support system		Money support	THB/unit term	source
study for free 15 years (นโยบายเรียนฟรี 15 ปี	2) books	a. formal education - kindergarten - P1	 200 483.2	THB/person/year	

		- P2	347.2		
อย่างมีคุณภาพ)		- P3	365.6		
		- P4	580.0		
		- P5	424.0		
		- P6	496.0		
		- M1	739.2		
		- M2	564.8		
		- M3	560.0		
		- M4	1160.8		
		- M5	805.6		
		- M6	763.2		
		- vocational	1000.0		
		b. Non-formal education			
		1) primary level	290		
		2) secondary	360		
		3) high school	400		
		4) vocational	500		
	3) other educational materials	1) kindergarten	100		
		2) primary level	195		
		3) secondary	210	THB/person/semester	
		4) high school	230		
		5) vocational	230		

Table C11: Government support project and support system (cont.)

Project		Support system	Money support	THB/unit term	source
study for free 15 years (นโยบายเรียนฟรี 15 ปี	4) uniform	1) kindergarten	300		
		2) primary level	360	THB/person/year	
		3) secondary	450		

อย่างมีคุณภาพ)		4) high school 5) vocational	500 900		
	5) academy activity	a. Non-formal education 1) kindergarten 2) primary level 3) secondary 4) high school 5) vocational b. Non-formal education 1) primary level 2) secondary 3) high school 4) vocational	 215 240 440 475 475 140 290 290 530	THB/person/semester	
The aging people support เบี้ยยังชีพผู้สูงอายุ		60 – 69 70 – 79 80 – 89 90 up	600 700 800 1000	THB/person/month	http://www.dop.go.th/download/formdownload/th1511850275-804_0.pdf
The handicapped people support เบี้ยยังชีพผู้พิการ			800	THB/person/month	http://dep.go.th/sites/default/files/files/news/ขั้นตอนรับเบี้ยความพิการ.pdf

Table C11: Government support project and support system (cont.)

Project	Support system		Money support	THB/unit term	source
The low income people support	from 2017 to	*General in household exp. support*			http://www.epayment.

โครงการช่วยเหลือผู้มีรายได้น้อย	present	Income <30000 THB/year 30000 >= income <= 100000 THB/year fuel for cooking exp. support *transportation exp. support* for bus/BTS expenditure for public bus expenditure for train expenditure	500 (300) 300 (200) 45* 500 500 500	Present (previous) Present (previous) *THB/3 month THB/person/month THB/person/month THB/person/month (not cumulate to next month)	go.th/home/app/media/uploads/files/project_4_doc_18-09-2017.pdf
The rice harvesting cost and quality improving support เงินช่วยเหลือค่าเก็บเกี่ยวและปรับปรุงคุณภาพข้าว	2017/2018 crop year	1200 THB/rai but not much more than 10 rai	1200 – 12000		http://www.kaset1009.com/th/articles/103625-ช่วยค่าเก็บเกี่ยว-ปรับปรุง คุณภาพข้าวไร่ละ1,200บาท
The Infant and Young Child Care support เงินอุดหนุนเพื่อการเลี้ยงดูเด็กแรกเกิด	From 2015 to present	1) for a child born in between 1 October 2015 – 30 September 2016 and registered within 30 September 2016 2) for a child born in between 1 October 2016 – 30 September 2017 and registered within 30 September 2017 3) for a child born in between 1 October 2016 – 30 September 2016 and registered after 30 September 2016	400 THB/month* 600 THB/month** 600 THB/month*** 600 THB/month****	*from month birth up to September 2016 **from October 2016 up to child be 3 years old ***from month birth to child be 3 years old ****from month registered to child be 3 years old	http://csg.dcy.go.th/index.php/en/2011-11-25-09-29-27/149-5
Support household faced flooding ผลกระทบจากอุทกภัยเนื่องจากพายุตาลัสและพายุเซินกา			3000	THB/household	https://www.thairath.co.th/content/1031767

Section 7.2.2. Insurance

Table C12: example of payment and compensation system

Project	Payment system	Compensation system	source
Cremation member of BAAC ฌาปนกิจสงเคราะห์ ธกส.	7 THB/member person died	50000 THB/member died	https://www.facebook.com/prasit.thre/posts/1928527987370898
Reward and pension บำเหน็จ บำนาญ		-for reward pay one time (Last salary * no. of working year) -for pension pay monthly (Last salary * no. of working year)/50	http://www.thailocalmeet.com/bbs/PDF/legislation/ส่วนที่%201%20สิทธิประโยชน์ด้านบำเหน็จบำนาญข้าราชการฯ.pdf

Section 9.2 Housing conditions

Table C13: Average house building cost and rental price per square metre

item	THB/square metre		source
	min	max	
House building cost	8500	11000	https://thaiappblog.files.wordpress.com/2018/01/used_cost_thai2561.pdf
Rental price	45	600	https://flatfy.in.th/บ้านให้เช่า-ชนบท-พร้อมเฟอร์นิเจอร์

Appendix D Plausibility rules for Vietnam, 2017 survey

Section 8. Household Expenditures, Prices of consumed items

Table D1: Amount and retail price of food items

(A) Commodity	(B) Unit	(C) min	(D) Amount consume in an average month (unit/person) Average	(E) max	(F) Min	(G) Max	Information source of column (C),(D) and (E)	Information source of column (F) and (G)	Remarks
Rice	kg		11.3	16.7	5.0	20.0	https://gain.fas.usda.gov/Recent%20GAIN%20Publications/Grain%20and%20Feed%20Annual_Hanoi_Vietnam_4-13-2017.pdf	http://www.vietfood.org.vn/thi-truong/gia-noi-dia.html?start=36 http://gaosachonline.com/bang-gia-gao-hom-nay-thang-9-2017/	Non-glutinous rice
Beef/pork	kg		3.3		63.0	185.0	https://www.statista.com/statistics/756932/vietnam-meat-consumption-per-capita-by-type/	https://gain.fas.usda.gov/Recent%20GAIN%20Publications/Grain%20and%20Feed%20Update%20Quarterly_Hanoi_Vietnam_10-12-2017.pdf	beef (25%) and pork(75%)
Fish	kg		2.2	3	40.0	210.0	http://www.helgilibrary.com/indicators/fish-consumption-per-capita/vietnam/		The expensive kind of fish is cá bớp, a kind of common fish in the Central
Poultry	kg		13.7		20.0	100.0	https://www.statista.com/statistics/756932/vietnam-meat-consumption-per-capita-by-type/	http://www.vietdvm.com/tin-tuc/viet-nam/gia-ca-thi-truong/gia-ca-thi-truong-mien-bac-tuan-38-nam-2017.html	
Eggs	piece		7.4		1.7	3.5	https://nld.com.vn/kinh-te/qua-trung-cay-dang-20170614161526187.htm	http://www.vietdvm.com/tin-tuc/viet-nam/gia-ca-thi-truong/gia-ca-thi-truong-mien-bac-tuan-38-nam-2017.html	
Tea and coffee	kg		0.1				www.gso.gov.vn		
Vegetable	kg		1.				www.gso.gov.vn		
Fruit			0.8				www.gso.gov.vn		
Alcohol	Litre		1				www.gso.gov.vn		

Table D2: Household expenditure for food, non-food and transport and communication sub-group

ID			Item	Amount per month (unit per capita)		Value per month (1000 VND per capita)	
				Min	Maximum	Min	Maximum
	8A - Food	1	Rice (kg)	0.2	50	3	600
		121	other staple food (kg per capita)	0.2	34	1.25	300
		2	beef/pork (kg per capita)	0.25	15	1.5	1500
		3	Fish (kg per capita)	0	20	0.8	1000
		4	Poultry (kg per capita)	0	5	2	750
		5	Eggs (piece)	0.2	60	0.6	180
		6	Vegetable (kg)	0.1	50	1	900
		7	Fruit (kg)	0	34	0.3	500
		8	Food ingredients, spices (includ. Salt/Sugar)			1.5	250
		9	Fermented fish/ fermented paste (kg)	0	7.5	0.8	300
		10	Cooking oil (litre)	0	4	0.8	140
		11	Beverages (litre)	0	60	1	750
		131	Milk powder for baby (age < 5 years)			0	
		12	Take home and eat out			1	1000
		13	Alcohol			1	1075
		132	Cigarettes				
		133	Coffee and Tea				
		14	Other food			3	800
	8B - Non-Food	17	Personal care supplies			0.7	500
		18	Clothes, shoes and bags, accessories			1.2	1500
		19	Detergent/washing powder			1	400
		20	Electricity			1.5	700

Table D2: Household expenditure for food, non-food and transport and communication sub-group (cont.)

ID			Item	Amount per month (unit per capita)		Value per month (1000 VND per capita)	
				Min	Maximum	Min	Maximum
		21	Water cost			0.75	375
		22	Liquid propane gas/Charcoal			2	150
		23	Hair dresser and other beauty service				
		130	Total tax payments				
	8C - Transport + communication	126	Fuel for car and motorbike			1.5	900
		27	Public transportation			1	500
		28	Telecommunication (includes mobile credit)			0.5	500
		29	Other transportation, Communication			1.8	900
		127	Maintenance for car and motorbike			0.5	425
		129	Fee for car and motorbike				

Source: 2016 TVSEP data

Table D3: Household expenditure for education, health and social group

			Item	Expenditure in last 12 months Education: Per number of people who currently enrol in school Health: per sick member Social: Per HH	
0			**1**	**4**	
ID				Min	Max
	8D - Education	31	Education fees, books	6	7000
		32	Student's dress/uniform	20	2000
		33	Tuition fee	7	200,000
		34	Rental fee (Dorm, apartment)	65	18,000
		35	Other costs of schooling	12	16,500
		36	School bus	10	3000
		37	Pocket money and meal	4	36,000
	8E - Health	40	Medicine (Purchase in Pharmacy only)	3	90,000
		41	Doctor fee	5	15,7000
		42	Other health	2	84,000
		47	Donations (to temples, social organizations, schools)		10,000
		48	Recreation and entertainment		60,000
		49	Lottery		12,000
		56	Other gambling expenditures		20,000
	8F(b) - Social	60	Bribery / corruption - police		40,000
		61	Bribery / corruption - government officer		25,000
		62	Bribery / corruption - business partner		60,000

Source: 2016 TVSEP data

Table D4: Tuition fee at public school in Vietnam (1000 VND/student/year)

Level	Areas	Min	Max
Pre-school and primary, secondary, high school	Urban	540	2,700
	Rural	270	1,800
	Mountainous areas	72	540
Professional / vocational training			31,500
Graduate			39,600
Post-graduate			59,400

Source: Decree No. 86/2015/ND-CP dated October 2, 2015 of the Government on mechanism for collection and management of tuition fees applicable to educational institution in the national education system and policies on tuition fee exemption and reduction and financial support from academic year 2015-2016 to 2020 - 2021.
http://vbpl.vn/bogiaoducdaotao/Pages/vbpq-toanvan.aspx?ItemID=84950

Section 4.1

Table D5: Land price at the provincial level (1000VND/m2)

	Residential/ business land		Annual crop		Perennial crop		Source
	Min	Max	Min	Max	Min	Max	
Ha Tinh	40	9000	16.7	55.4	16.7	55.4	https://thuvienphapluat.vn/van-ban/Bat-dong-san/Quyet-dinh-01-2017-QD-UBND-dieu-chinh-94-2014-QD-UBND-Bang-gia-dat-Ha-Tinh-337869.aspx
Hue	40	5,000	11.4	23.3	11.4	23.3	https://www.thuathienhue.gov.vn/vi-vn/Tra-cuu-gia-dat https://thuvienphapluat.vn/van-ban/Bat-dong-san/Quyet-dinh-75-2014-QD-UBND-Bang-gia-dat-Thua-Thien-Hue-2015-2019-262617.aspx
Dak Lak	70	5,000	8	37	13	45	http://tnmt.daklak.gov.vn/Default.aspx?TabID=98&AspxAutoDetectCookieSupport=1

Table D6: Rental land payment rate at the provincial level for each type of main land use(1000VND/m2)(1,5% of land price)

	Residential/ business land		Annual crop		Perennial crop		Source
	Min	Max	Min	Max	Min	Max	
Ha Tinh	0.6	135	0.25	0.83	0.25	135	https://thuvienphapluat.vn/van-ban/Bat-dong-san/Quyet-dinh-01-2017-QD-UBND-dieu-chinh-94-2014-QD-UBND-Bang-gia-dat-Ha-Tinh-337869.aspx
Hue	0.6	75	0.17	0.35	0.17	75	https://www.thuathienhue.gov.vn/vi-vn/Tra-cuu-gia-dat https://thuvienphapluat.vn/van-ban/Bat-dong-san/Quyet-dinh-75-2014-QD-UBND-Bang-gia-dat-Thua-Thien-Hue-2015-2019-262617.aspx
Dak Lak	1,05	75	0.12	0.56	0.19	75	http://tnmt.daklak.gov.vn/Default.aspx?TabID=98&AspxAutoDetectCookieSupport=1

Section 4.2

Table D7: Average yield (tonnes/ha) per crop type per province and by season

(A)	(B)	(C)	(D)		Source of column C	Source of column D
	Minimum	Maximum	Average	Source of column B		
Rice	1	7	+ Ha Tinh:5,1 + Hue: 6,0 + Dak Lak: 5,8	TVSEP 2016	http://nongnghiep.vn/5-giong-lua-nang-suat-cao-post203683.html	http://gso.gov.vn/default.aspx?tabid=717
Cassava	3	50	+ Ha Tinh:14.0 + Hue:18,6 + Dak Lak:20.3	TVSEP 2016	http://nhanong.netnews.vn/gioi-thieu-2-giong-san-cao-san-mid-1-2-0-5084.html	
Coffee (bean) Average for whole cycle	0.7	7	2.3	TVSEP 2016	http://www.dalatcoffee.com.vn/viewproduct/2749_cac-dong-ca-phe-robusta-chat-luong-cao-o-viet-nam.htm	Dak Lak Statistical Yearbook
Year 1-3		3				
Peak period (Year 4-Year 10)		8				
After year 10		4				
Pepper	1.5	11	2.4– 3.2	TVSEP 2016	https://www.lafarmeakmat.com/tin-tuc/nhung-giong-ho-tieu-nang-suat-cao-tai-thi-truong-viet-nam	Dak Lak Statistical Yearbook

Note: Conversion: 4.5 coffee cherry = 1 coffee bean

Table D8: Monthly farm gate price by product type and product (1000VND/kg)

	Rice	Cassava	Coffee	Pepper
January 2017	5.1-5.8	1.2 – 1.4	46.3 – 46.5	130
February 2017	5.2-5.8	1.2 – 1.4	45,4	115
March 2017	5.2 – 5.8	1.3 – 1.6	46,3	110
April 2017	5.1 – 5.5	1.4 -1.6	43,6	97
May 2016	5.0 – 5.5	1.8- 1.9	35,4 – 36.0	
June 2016		1.5 – 1.6	37.0 – 37,4	
July 2016		1.6-1.7	38,5 – 38,9	
August 2016		1.5-1.6	38,2 – 38,7	
September 2016	4.8-5.2	1.2-1.4	41,4 – 42,2	
October 2016	4.7-5.4	1.3–1.4	41,5 – 42,3	138
November 2016		1.4 – 1.5	44,5	135
December 2016	5.0 – 5.4	1.3 – 1.5	44	140

Sources:
- Rice: http://www.vietfood.org.vn
- Cassava: http://hiephoisanvietnam.org.vn
- Coffee, Pepper: http://www.caphevietnam.com

Table D9: Unitary standard budget per crop type

	Budget (1000 VND per ha)			Source
	Min	Average	Max	
Rice		30,000		https://baomoi.com/ha-tinh-trong-lua-ca-vu-khong-mua-duoc-3-bat-pho-dan-dua-nhau-tra-ruong/c/24172000.epi
Rice		13,183	72,666	TVSEP 2016
Cassava		4,651	25,000	TVSEP 2016
Coffee		27,077	119,384	TVSEP 2016
Pepper		36,040	1,500,000	TVSEP 2016

Section 4.3.1

Table D10: Price of each livestock by size for example chick (purchasing price) and chicken (sale price) (1000 VND)

Livestock	Min price	Average	Max price	Source
Chick		8-9	20	http://hoichannuoi.vn
Chicken	60		150	http://www.vigova.vn/sanpham/vit-giong-1-ngay-tuoi.html
Young cow/buffalo	5,000	20,000		https://www.trieuphunongdan.com/gia-mua-ban-bo-giong-bo-thit/
Adult cow/buffalo	15,000		50,000	https://www.trieuphunongdan.com/gia-mua-ban-bo-giong-bo-thit/
Mothersow	2000		7000	http://hoichannuoi.vn
Piglet (20kg)	300		600	http://hoichannuoi.vn
Fattening pig	1200		2200	http://hoichannuoi.vn
Small duck			50	http://www.vigova.vn/sanpham/vit-giong-1-ngay-tuoi.html
Duck			200	http://www.vigova.vn/sanpham/vit-giong-1-ngay-tuoi.html

Table D11: Cash cost (1000 VND) per unit by type of livestock

Livestock	Min	Average	Max
Beef	2,700	11,000	80,500
Pig	0	2,500	9,000
Chicken	5	100	500

Source: TVSEP 2016

Section 4.3.2

Table D12: Price of livestock products i.e. egg

Livestock product	Min price (1000VND)	Max price (1000VND)	Source
Chicken egg	0.5	2	http://hoichannuoi.vn
Duck egg	0.5	3	http://www.vigova.vn/sanpham/vit-giong-1-ngay-tuoi.html

Section 4.4

Table D13: Price of some collecting and hunting products

Product	Min price (1000VND)	Max price (1000VND)
Fish (kg)	0.5	200
Charcoal (m3)	1	5
Timber(m3)	7000	12,000
Honey (litre)	4	500
Mushroom (kg)	1	500

Section 5

Table D14: Monthly Income from different sources of by provinces in 2016 (1000 VND)

	Total	Salary & wage	Agriculture, forestry & fishery	Non-agriculture, forestry & fishery	Others
Ha Tinh	2,074.0	787.0	531.0	410.0	346.0
Thua Thien-Hue	2,593.0	1,262.0	268.0	734.0	329.0
Dak Lak	2,321.0	796.0	870.0	513.0	142.0
Red River Delta	3,610.0	1,965.0	360.0	847.0	438.0
Ha Noi	5,057.0	3,120.0	302.0	1,060.0	575.0
Northern Central area and Central coastal area	2,432.0	1,138.0	433.0	571.0	290.0
Central Highlands	2,562.0	851.0	972.0	600.0	139.0
Ho Chi Minh city	5,481.0	3,249.0	61.0	1,521.0	650.0
Mekong River Delta	2,798.0	978.0	772.0	660.0	388.0

Source: gso.gov.vn

Table D15: Minimum daily wage in 2016 and 2017

Category (IV is the most remoted)	Year 2016 (1000 VND/day)		Year 2017 (1000 VND/day)	
	Unskilled	Skilled (plus 7%)	Unskilled	Skilled (plus 7%)
Region I	110	117	125	133
Region II	100	107	110	117
Region III	90	96	97	103
Region IV	80	85	86	92

Source: Decree 153/2016/ND-CP

Section 7.1. Borrowing and Lending

Table D16: Interest rate by banks from May 2016 – April 2017 (% per year)

Bank	Borrowing		Saving		Source
	Min	Max	Min	Max	
AgriBank	6	8.5	No duration: 0.2 Fixed duration: 4	7	http://www.agribank.com.vn
Bank of social policy	3	8.25	No duration: 0.2 Fixed duration: 4	7	http://vbsp.org.vn/
Other bank	5	10	No duration: 0.1 Less than one month: 0.4 From one month and more: 4	8	http://research.lienvietpostbank.com.vn/nhin-lai-lai-suat-nam-2016-va-du-bao-xu-huong-nam-2017

Section 7.2. Public Transfer and Insurance

Table D17: List of public programs and payments

Public programs	Payment(VND)/month	Source
Contingency fund for pre-harvest starvation and disaster relief	Deliver rice, drinking water, rice, instant noodle and cash for households affected by disaster. Estimated payment: 200,000 VND/time	http://qmt.vn/
Hunger eradication and poverty reduction (HERP) program	Support with crop and livestock production, working abroad for the poor	Decision No. 1722 QD-TTg
Allowances for war veterans and matyrers	Prioritized scheme for social and health insurance, retirement pensions - For people who participate in the war before 1945: 1,363 thousand VND/month + 230 thousand VND/ working year 2,315 thousand VND/month - After 1945: 1,262 thousand VND/month	Resolution No. 157/2016/ND-CP & 150/2006/ ND-CP Resolution No. 20/2015/ND-CP Resolution No.101/2013/ND-CP
Payment / compensation for agent orange victims	- Veterans who are agent orange victims and lose at least 81% health: 612 thousand VND/month - Veterans who are agent orange victims and lose at least 81% health and having serious diseases: 1,220 thousand VND/month - Relative who take care of agent orange victims: 1,220 thousand VND/month - Children of agent orange veteran (lose 61-80% of health): 732 thousand VND/month - Children of agent orange veteran (lose more than 81% health): 1,220 thousand VND/month	Resolution No.101/2013/ND-CP
Retirement pensions	Depending on years of working, years of paying social insurance. For people who participate in social insurance before 1995, monthly payment = Total based salary for social insurance in the last 5 working years/60	

99

Table D17: List of public programs and payments (cont.)

Public programs	Payment(VND)/month	Source
Work accidents and industrial disease	- Fault of the employee: receive 40% of + 0.6 monthly salary if losing 5-10% health, 0.16 increasing in salary for every 1% health lost. + 30 months of salary if the employee die or lose more than 81% health - Fault of the employers: + 1.5 monthly salary if losing 5-10% health, 0.4 increasing in salary for every 1% health lost + 30 months of salary if the employees die or lose more than 81% health - Allowance for 5-10 days, each day equalling to 25-40% basic salary - If the employee lose at least 31% health and cannot work anymore: 100% minimum salary - If the employee die, the household receive support for funeral and 50% minimum salary for each dependents - For those lose under 5% health: no allowance - For those lose 5-30% health: One time allowance, equalling to 5 months of minimum salary + 0.5% of minimum salary for each 1% of health loss (from 5% onward). - For those lose more than 30% health: Receive monthly the amount equalling to 30% of minimum salary + 2% minimum salary for each 15 health lost (from 30% onward) - Besides, if the employee participate in social insurance for at least 1 year, they will receive 50% of minimum wage + 50% of minimum wage for each year of paying social insurance	Labour code; Resolution No. 04/2015/ TT-BLĐTBXH

Table D17: List of public programs and payments (cont.)

Public programs	Payment(VND)/month	Source
Sickness benefits	- 100% health diagnostic and treatment fee for + Children under 6 year old + People who contribute to the war + People working in police and military + People who participate in government health insurance for 5 years continuously + At commune level health care center - 95% health diagnostic and treatment fee for: + Those who are receiving retirement pension and allowance for disability + Poor household + Ethnic minorities in remote areas + Relatives of invalid veterans or matyrers - 80% health diagnostic and treatment fee for the rest who participate in government health care insurance	Health insurance law 2014
maternity leave benefits	2.420 thousand VND 6 months leaves with salary	*Resolution* No. 47/2017/ND-CP
Survivor benefits	- Allowance for relatives of pass-away invalid veterans who lost 61% health: 684 thousand VND/month - Parents or children of pass-away invalid veterans: 976 thousand VND/month - Relative of matyrers: + 1 matyrer: 1,220 thousand VND/month + 2 matyrers: 2,440 thousand VND/month + 3 matyers and above: 3,660 thousand VND/month - Parents or children of matyrers: 976 thousand VND/month - Remarried wife/husband of matyrers: 1,220 thousand VND/month - Mother of matyers who got title "Vietnamese Heroic Mother" will get 1,023 thousand VND more - People who take care of Vietnamese Heroic Mothers: 1,220 thousand VND/month	Resolution No.101/2013/ND-CP

Table D17: List of public programs and payments (cont.)

Public programs	Payment(VND)/month	Source
disability benefits (children and adults)	- People with serious disabilities: 540 thousand VND/month - People with serious disabilities and are children/elderly: 675 thousand VND/month. - People with disabilities: 270 thousand VND/month	Resolution No. 136/2013/ND-CP
support for the elderly	- Elderly more than 80 year old and does not receive retirement pension: 270 thousand VND/month - Elderly from 60 -80 years old in poor family with no labour: 405 thousand VND/month - Elderly more than 80 years old in poor family with no labour: 540 thousand VND/month - Elderly more than 60 years old in poor family who live alone, no one to take care: 810 thousand VND/month	Resolution No. 136/2013/ND-CP
allowance for government official	- Tet holiday: 1,500 – 5,000 thousand VND - Western New Year: 500 thousand VND - Hung King: 500 thousand VND - Labour Day: 600 thousand VND - Independence Day: 600 thousand VND	
Social security cash benefit for children	- Children less than 4 years old who are orphan, abandoned or parents do not have the capacity to take care of: 675 thousand VND - Children 4-16 years old, or 16 -22 years old and are studying: 405 thousand VND - Children < 4 years old with HIV from poor family: 675 thousand VND - Children >4 years old with HIV from poor family: 540 thousand VND	Resolution No. 136/2013/ND-CP

7.2.2. Insurance for employee with work contracts

Table D18: Percentage of salary for insurance

Type of insurance	Employer	Employee	Total
Social insurance	17.5%	8%	25.5%
Health insurance	3%	1.5%	4.5%
Unemployment insurance	1%	1%	2%

Source: Decision No. 595/QD-BHXH

Table D19: Minimum monthly salary to use as basis to calculate insurance (1000 VND/month)

Region	Unskilled employee	Skilled employee
Region 1	3,750	4,012
Region 2	3,320	3,552
Region 3	2,900	3,103
Region 4	2,580	2,760

Notes:
a) Maximum monthly salary to use as basis to calculate social insurance (1000VND/month): 24.200
b) Maximum monthly salary to use as basis to calculate unemployment insurance: 20 * minimum salary

Table D20: Maximum and minimum insurance amount (1000 VND/month)

Type of insurance	Employer		Employee	
	Min	Max	Min	Max
Social insurance	450	4,235	200	1,936
Health insurance	77	726	38	363
Unemployment insurance	25	802	25	802

Section 9.2

Table D21: Average residential area per capita by provinces in 2016 (m2)

Province	Total	Permanent house	Semi-Permanent house	Less-temporary house	Simple permanent house
Ha Tinh	23.1	23.9	19.9	17.4	15.4
Thua Thien-Hue	23.5	26.3	20.0	16.3	8.2
Dak Lak	19.4	20.7	19.6	10.8	7.6

Source: gso.gov.vn

Notes:
a) Housing price in the rural areas (Price of material and construction per 1m2 of house): 500 – 7000 thousand VND/m2 (Source: Provincial Department of Construction)
b) Rental price for a house in the village: Maximum 300 thousand VND/m2/month (Source: batdongsan.com.vn)