\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*

\*\* PROGRAMME: prep\_hhdata\_06Uganda\_2009\_tables\_A1\_to\_A5\_vXX

\*\* AUTHOR: Dédé Houeto

\*\* OBJECTIVE: Create standard tables A1 to A5 for UNHS 2009/2010

\*\*

\*\* DATA IN: HSEC1.dta

\*\* HSEC2.dta

\*\* HSEC3.dta

\*\* HSEC10A\_CLN.dta

\*\* HSEC10AA.dta

\*\* HSEC10B\_CLN.dta

\*\*

\*\* DATA OUT: hhdata.dta (table A1)

\*\* indata.dta (table A2)

\*\* calperg.dta (table A3)

\*\* cons\_cod\_trans.dta (table A4)

\*\* cons\_cod.dta (table A5)

\*\*

\*\* NOTES:

\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

clear

set logtype text

capture log close

set more off

global path\_ug "D:\Dedevi\Projet\_UNU WIDER Growth and Poverty Project\GAPP\_Uganda"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table A1: Household Characteristics and interview details

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

use "$path\_ug\in\2009\HSEC1.dta"

\*keep

\*\*\*------- Primary Sampling Unit

\* The primary sampling unit for the 2009 UNHS is the enumeration area.

codebook ea

/\* There are 711 unique values, the UNHS report mentions 712 \*/

rename ea psu

label variable psu "Primary Sampling Unit"

\*\*\*------- Interview Quarter

\* For the Mozambique Data, instead of using the regular quarter definition for their "surquar" variable they defined the quarters

\* relative to the time period covered by the survey: the survey ran from Sept 2008 to August 2009 so they defined the quarters

\* as follows: Sept-Nov 08, Dec08-Feb09, Mar-May 08 and June-Aug 08.

\* I will use the same framework

tab month Year,m

/\* According to the repartition of months and year, the survey ran from May 2009 to April 2010. I will define the

\* the quarters accordingly. It should be noted that no data was collected in August 2009 so the 2nd quarter I am

\* defining only has two months's worth of survey data \*/

gen float survquar=1 if month>=5 & month<=7 & Year==2009

replace survquar=2 if month>=8 & month<=10 & Year==2009

replace survquar=3 if (month>=11 & month<=12 & Year==2009) | (month==1 & Year==2010)

replace survquar=4 if month>=2 & month<=4 & Year==2010

label define lsurvquar 1 "May-Jul 09" 2 "Sept-Oct 09" 3 "Nov09-Jan10" 4 "Feb-Apr 10"

label values survquar lsurvquar

label variable survquar "Sequential Survey Quarter May-Jun 09=1)"

\*\*\*------- Sequential Interview month

\* Following the Mozambique file, I am creating a survey month variable rather than an interview date variable as per the excel sheet

\* Number 1 corresponds to the first month of the survey and not to January.

gen float survmon=1 if month==5 & Year==2009

replace survmon=2 if month==6 & Year==2009

replace survmon=3 if month==7 & Year==2009

replace survmon=4 if month==8 & Year==2009

replace survmon=5 if month==9 & Year==2009

replace survmon=6 if month==10 & Year==2009

replace survmon=7 if month==11 & Year==2009

replace survmon=8 if month==12 & Year==2009

replace survmon=9 if month==1 & Year==2010

replace survmon=10 if month==2 & Year==2010

replace survmon=11 if month==3 & Year==2010

replace survmon=12 if month==4 & Year==2010

#delim;

label define lsurvmon 1 "May 09" 2 "Jun 09" 3 "Jul 09" 4 "Aug 09" 5 " Sep 09" 6 "Oct 09"

7 "Nov 09" 8 "Dec 09" 9 "Jan 10" 10 "Feb 10" 11 "Mar 10" 12 "Apr 10";

#delim cr

label values survmon lsurvmon

tab survmon month,m

label variable survmon "Sequential Survey Month (May 2009=1)"

\*\*\*------- Household Sample Weight

rename hmult hhweight

label variable hhweight "Household sample weight"

\*\*\*------- Household id

codebook hh /\*there are 6775 different values and 6775 observations in the dataset so the variable "hh" uniquely identifies the observations \*/

rename hh hhid

label variable hhid "Household ID"

\*\*\*------- Household Size

rename hsize hhsize

label variable hhsize "Household Size"

\*\*\*------- Geographical Stratification during sampling

\* As indicated in the survey report, EAS were drawn from a geographical grouping into regions and rural-urban location.

\* A variable in the file is already identified as the strata, and there are indeed 10 of them as indicated in the report

tab sregion,m

rename sregion strata

label variable strata "Strata"

\*\*\*------- Rural-Urban Location

\* I checked and the code is the opposite of the one in the Mozambique data file: 1=rural 0=urban. I will create a new variable

\* that has the same coding

tab urban,m

gen float rural=(urban==0)

replace rural=0 if urban==1

tab rural,m

label define lrural 0 "Urban" 1 "Rural"

label values rural lrural

label variable rural "Rural/Urban Location"

\*\*\*------ Spatial domains (each with its own poverty line)

\* In the Arndt & Simler 2010 paper the spatial domains are a combinaison of regions and rural/urban delimitations + the capital

\* city as a separate domain (confirmed in the Mozambique data file)

\*--> We'll use the same categories for Uganda, without singling out Kampala (can't find it in the hh data file, if needed we might

\* dig deeper)

rename regurb spdomain

label variable spdomain "Spatial domains: each with own poverty line"

\*\*\*------ (not important) Other administrative geographical boundries where survey is representative

\*\*\* LEFT TO DO:

\*\*\*------ Regions used for temporal price index calculations

\* In the Mozambique data file, they have grouped the country into North/Center/South + Rural/Urban for the tpi

\* ??? The rationale for this choice is not explained.

\*--> Without a rational for another grouping, only the region + urban/rural grouping makes sense to me, but then it will be

\* the same grouping as the spatial domains...

\*--> In the poverty software, the STATA program "020\_in\_2\_work\_folder" mentions these regions and suggests that they are to be created

\* according to the four geographical directions: North, South, East, West or a combinaison of them

\*--> !!! Need to discuss this

sort hhid

save "$path\_ug\out\hhdata\_6\_2009.dta",replace

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table A2: Individual characteristics - demographics

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

clear

use "$path\_ug\in\2009\HSEC2.dta"

\*\*\*----- Household members

/\* According to the enumerator manual, Usual and Regular household members are defined as follows:

Usual members are defined as those persons who have been living in the household for 6 months or

more during the last 12 months. However, members who have come to stay in the household permanently

are to be included as usual members, even though they have lived in this household for less than 6

months. Furthermore, children born to usual members on any date during the last 12 months will be taken

as usual members. Both these categories will be given code "1" or "2" depending upon whether they are

present or absent on the date of the interview.

Regular members refer to those persons who would have been usual members of this household, but

have been away for more than six months during the last 12 months, for education purposes, search of

employment, business transactions etc. and living in boarding schools, lodging houses or hostels etc.

These categories will be given code "3" or "4" depending upon presence or absence on the date of the

interview. \*/

/\*

\* For the purposes of the calculation of a poverty line we'll exclude from the household the members who have left

the household permanently or died

\* We'll keep the members away for more than 6 months but present on the day of the interview

--> We'll remove some of these members later on depending on the expenditure aggregates being computed

\*/

rename h2q5 resident

drop if resident==7

\*\*\*------ Household ID

rename hh hhid

label variable hhid "Household ID"

\*\*\*------ Individual ID

gen double indid=hhid\*100 + h2q1

codebook indid

\* There are 35,945 different values, and there are 35,945 observations in the dataset so indid uniquely identifies

\* the individuals

label variable indid "Individual ID"

\*\*\*------ Sex

rename h2q3 sex

label variable sex "Sex"

\*\*\*------ Age

rename h2q8 age

label variable age "Age in years completed"

\* In order to have the information on whether the mother resides in the house or not we need the file "HSEC3"

sort indid

save "$path\_ug\work\2009\temp\_A2\_1.dta",replace

clear

use "$path\_ug\in\2009\HSEC3.dta"

\*\*\*------ Individual ID

rename hh hhid

destring hhid,replace

gen double indid=hhid\*100 + h3q1

label variable indid "Individual ID"

codebook indid

\* There are 35,945 different values, and there are 35,945 observations in the dataset so indid uniquely identifies

\* the individuals

\*\*\*----- Mother lives in household?

tab h3q3,m

\*!!!! There are 14,795 (ie 42.47 %) missing responses !!!!\*

\*!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!\*

\* They have four categories: 1: Yes 2: No, Alive 3: No, Dead 4:No, Don't know. I will group them

gen motherhh=1 if h3q3==1

replace motherhh=0 if h3q3==2 | h3q3==3 | h3q3==4

label variable motherhh "Mother lives in hh"

label define lmoth 0 "No" 1 "Yes"

label values motherhh lmoth

keep indid motherhh

sort indid

merge 1:1 indid using "$path\_ug\work\2009\temp\_A2\_1.dta"

tab \_merge

\* There are some obs coming only from the using data. The explanation is that Section 3 of the questionnaire

\* is administered only to usual and regular household members, as is confirmed by the cross tab below

tab resident \_merge

\* We leave the variable as is, with additional missing values for the variable "motherhh".

drop \_merge

keep hhid indid sex age motherhh

save "$path\_ug\out\inddata\_6\_2009.dta",replace

\*\*\*\*\*\*\* Clean up

erase "$path\_ug\work\2009\temp\_A2\_1.dta"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table A3: Calorie content of food items

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* I compiled the calorie content and edible portion in an excel file then converted that file into a STATA file

\* The excel file (with more detailed information, inculding the sources) is in the "in" folder

\* Note: in order to have the description of the food items in a separable variable (in the STATA file they are a label of the

\* variable "produc"), I created an excel file called fooditems\_2009\_excel that I used to compile the calorie content.

clear

use "$path\_ug\in\2009\foodcomp\_uganda\_hhsurvey2009\_v5.dta", replace

\* Since they did not include edible portions in the file I assume that the calorie per gram is only for the edible portion.

\* I will therefore compute calperg that way

gen double calperg=((kcal\_100g\*edible)/100)/100

keep product descript calperg

label variable product "Food product code: numerical"

label variable descript "Product Description: incl. product code in the beginning"

label variable calperg "Calorie content of food product: calories per gram"

save "$path\_ug\out\calperg\_6\_2009.dta",replace

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table A4: Amount and Quantity of food transactoin - Transaction level

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Food expenditure information is recorded in section 10 of the questionnaire, there are two files for this section:

\* - One file with expenditure information

\* - One with the head count of the household, split into male:/female, children/adult, member/not member

\* I will merge both files so that I can compare my calculations with theirs

clear

use "$path\_ug\in\2009\HSEC10A\_CLN.dta"

rename hh hhid

label variable hhid "Household ID"

sort hhid

save "$path\_ug\work\2009\temp\_A4\_1.dta",replace

clear

use "$path\_ug\in\2009\HSEC10AA.dta"

rename hh hhid

destring hhid,replace

label variable hhid "Household ID"

sort hhid

merge 1:m hhid using "$path\_ug\work\2009\temp\_A4\_1.dta"

drop \_merge

save "$path\_ug\work\2009\temp\_A4\_2.dta",replace

\*\*\* In order to look for (and find) the adequate unit conversion factors I will create a file with only the item codes and the units

collapse (count) h10aaq4, by ( itmcd untcd)

rename itmcd product

rename h10aaq4 nobs

save "$path\_ug\work\2009\temp\_A4\_3\_units.dta",replace

sort product

merge m:1 product using "$path\_ug\in\2009\foodlist\_2009\_excel.dta"

drop \_merge

sort untcd

merge m:1 untcd using "$path\_ug\in\2009\unitlist\_2009.dta"

drop if \_merge==2

drop \_merge

sort product untcd

saveold "$path\_ug\work\2009\temp\_A4\_4.dta",replace

\*!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

\* NOTE: the unit conversion factors used are from the Report sent by James Muwonge to Bjorn

\*--> It would be better to have them by region (we had them for the project with Lisa Smith)

\* I will work with what I have for now and will add this as part of the questions I will

\* ask the researchers who worked on the latest poverty figures for Uganda

\*!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

\*\*\* Merge expenditure file with unitconversion factors

clear

use "$path\_ug\in\2009\ucf\_uganda\_unsh2009\_v3.dta"

sort product untcd

save, replace

clear

use "$path\_ug\work\2009\temp\_A4\_2.dta"

rename itmcd product

sort product untcd

merge m:1 product untcd using "$path\_ug\in\2009\ucf\_uganda\_unsh2009\_v3.dta"

drop \_merge

\*\*\* Use of most straighforward units to build missing unit conversion factors

\* Using purchases

rename h10aq4 qty\_conspurch

rename h10aq5 val\_conspurch

rename h10aq6 qty\_consaway

rename h10aq7 val\_consaway

rename h10aq8 qty\_consown

rename h10aq9 val\_consown

rename h10aq10 qty\_consfree

rename h10aq11 val\_consfree

rename h10aq12 price\_mkt

rename h10aq13 price\_farmg

\* Implicit price per unit

\* Price from quantities purchased

gen double pric\_purch=val\_conspurch/qty\_conspurch

gen double pric\_purch\_kg=pric\_purch/ucf\_kg

\* Market Price VS Price from quantities purchased

gen dif\_pric1=pric\_purch- price\_mkt

tab dif\_pric1

\* The difference is zero for 60,806 obs over 62,461 non missing obs(97.35%). For the rest the difference is less than 1 UGS, there

\* are 5 cases where the difference is between 198 and 544 UGS (knowing that 1 UGS~ 2500 $US). It seems that the value of consumption

\* was computed from market prices, but in that case there shouldn't be ANY difference in the market price and the price from

\* quantities purchased

\* ????????????????????????

\*???????????????? QUESTION TO ASK

\*--> I will use only the implied price from quantities purchased

\* Price from quantities consumed away from home

order descript untdes ucf qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg price\_mkt price\_farmg qty\_consaway val\_consaway qty\_consown val\_consown qty\_consfree val\_consfree

codebook qty\_consaway

\* 88501 missing over 89597 observations. There are too many

\* missing observations for this variable to yield useful implied prices.

\*--> I suggest we use prices from purchases for food consumed away from home

order descript untdes ucf qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg price\_mkt price\_farmg qty\_consaway val\_consaway qty\_consown val\_consown qty\_consfree val\_consfree

\* Price from quantities consumed from own production

codebook qty\_consown

\* 69018 missing over 89597. These implied prices need to be compared with farm gate prices.

gen double pric\_own=val\_consown/qty\_consown

gen double pric\_own\_kg=pric\_own/ucf\_kg

\* Farm gate prices

gen dif\_pric2= pric\_own- price\_farmg

tab dif\_pric2

\* The difference is zero for 20,021 obs out of 20,578 non missing obs (97.29%). The difference is between 900 and 13500 for 7 obs.

\* Let's take a closer look at these observations

sort dif\_pric2

order descript untdes ucf qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg price\_mkt price\_farmg qty\_consaway val\_consaway dif\_pric2 qty\_consown pric\_own val\_consown qty\_consfree val\_consfree

\* For 6 of the 7 observations the price from own consumption makes more sense. For the 7th observation it is the inverse, but since I am

\* sure about the price of sweet potatoes I will leave all prices from own consumption as is, and I have decided to use price from

\* own consumption rather than farm gate prices

sort product untcd

order descript untdes ucf qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg price\_mkt price\_farmg qty\_consaway val\_consaway qty\_consown val\_consown qty\_consfree val\_consfree

\* Prices from consumption of items received free or in kind

codebook qty\_consfree

\* 84640 missing observations out of 89597. There are too many

\* missing observations for this variable to yield useful implied prices.

\*--> I suggest we use market prices for food received free or in kind

\*\*\* Merge with file with information on households

sort hhid

merge m:1 hhid using "$path\_ug\out\hhdata\_6\_2009.dta"

drop \_merge

sort product untcd spdomain

order descript untdes ucf\_kg qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg price\_mkt price\_farmg spdomain

\* Create an index to specify the type of consumption

gen typ\_cons=1 if qty\_conspurch~=.

\*----------------------------------------

\* Average price per kg

\*----------------------------------------

\*--> This will be donne by region and rural/urban area

\*--> This will be done by type of purchase, as prices from consumption of own production are expected to be lower than market prices

\* !!!! Possible way to refine this: compute average for big units (e.g. 15l Tins) separately from small units (e.g. heap) as error

\* margins can be expected to be different in each case.

\* !!!!! For later.

\*\*\*----------- Average Price from purchases

\*--------------------------------------------------

\*\*\* Prices by region and rural/urban areas

egen mean\_pric\_purch\_kg=mean(pric\_purch\_kg), by (product spdomain)

order descript untdes ucf\_kg qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg mean\_pric\_purch\_kg mean\_pric\_purch\_kg spdomain

codebook mean\_pric\_purch\_kg

\* 7231 missing obs over 89597

tab descript if mean\_pric\_purch\_kg==.

\*--> Some of these are missing even though the unit conversion file provided ucf for them: this is because of the break down by region

\* and rural/urban area.

\*--> We'll proceed by first, second and third best:

\* First best: Average by food, region and rural/urban

\* Second best: Average by food and region

\* Third best: Average by food only

\*\*\* Prices by region

egen mean\_pric\_purch\_kg\_reg=mean(pric\_purch\_kg), by (product region)

order descript untdes ucf\_kg qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg mean\_pric\_purch\_kg mean\_pric\_purch\_kg spdomain ///

mean\_pric\_purch\_kg\_reg region

codebook mean\_pric\_purch\_kg\_reg

\* 6369 missing obs over 89597

\*\*\* Prices at the national level

egen mean\_pric\_purch\_kg\_nat=mean(pric\_purch\_kg), by (product)

order descript untdes ucf\_kg qty\_conspurch val\_conspurch pric\_purch pric\_purch\_kg mean\_pric\_purch\_kg mean\_pric\_purch\_kg spdomain ///

mean\_pric\_purch\_kg\_reg region mean\_pric\_purch\_kg\_nat

codebook mean\_pric\_purch\_kg\_nat

\* 5413 missing obs over 89597

\*\*\*------- Remaining missing average prices for price from purchase

\*\*\*--------------------------------------------------------------------

tab descript if mean\_pric\_purch\_kg\_nat==.

/\* These products are the ones with missing average prices

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

139 Other vegetables | 2,567 47.42 47.42

155 Cigarettes | 555 10.25 57.68

156 Other Tobacco | 833 15.39 73.06

157 Food | 894 16.52 89.58

160 Other juice | 139 2.57 92.15

161 Other foods | 425 7.85 100.00

------------------------------------+-----------------------------------

Total | 5,413 100.00

\*/

\*\*\* Other vegetables

\* I will use the average price of all vegetables

gen veggie=(product>=135 & product<=138)

tab product veggie

egen mean\_pric\_purch\_veg\_a=mean(pric\_purch\_kg) if veggie==1, by (spdomain)

egen mean\_pric\_purch\_veg=min(mean\_pric\_purch\_veg\_a), by (spdomain)

order product veggie mean\_pric\_purch\_veg\_a mean\_pric\_purch\_veg

codebook mean\_pric\_purch\_veg

\* No missing obs. No need to compute the average at the regional only or national level

\*\*\* Other Juice

\* The problem here arises because there are no units for "other juice". So I can not use the average price of soda and other soft drinks

gen juice=(product==151 | product==154)

tab product juice

egen mean\_pric\_purch\_jui\_a=mean(pric\_purch\_kg) if juice==1, by (spdomain)

egen mean\_pric\_purch\_jui=min(mean\_pric\_purch\_jui\_a), by (spdomain)

order product juice mean\_pric\_purch\_jui\_a mean\_pric\_purch\_jui

codebook mean\_pric\_purch\_jui

\* No missing obs. No need to compute the average at the regional only or national level

\*\*\* Food/Other Foods

\* The problem here arises because there are no units for "food" or "other foods". So I can not use the average price of some units to complete it

\* for those whose ucf is missing

\* I will compute the average price of all foods

gen fooditem=(product>=100 & product<=125) | (product>=127 & product<=147)

order product fooditem

egen mean\_pric\_purch\_food\_a=mean(pric\_purch\_kg) if fooditem==1, by (spdomain)

egen mean\_pric\_purch\_food=min(mean\_pric\_purch\_food\_a), by (spdomain)

order product fooditem mean\_pric\_purch\_food\_a mean\_pric\_purch\_food

codebook mean\_pric\_purch\_food

\* No missing obs. No need to compute the average at the regional only or national level

\*\*\*------- Other Fruits

\* I computed the average price for other fruits, but it is not reliable because it is based only (at most) on three observations

\* of "other fruits" given in "kg". I will instead use the average price of all the other fruits.

gen fruit=(product>=130 & product <=133)

tab product fruit

egen mean\_pric\_purch\_frui\_a=mean(pric\_purch\_kg) if fruit==1, by (spdomain)

egen mean\_pric\_purch\_frui=min(mean\_pric\_purch\_frui\_a), by (spdomain)

order product fruit mean\_pric\_purch\_frui\_a mean\_pric\_purch\_frui

codebook mean\_pric\_purch\_frui

\* No missing obs. No need to compute the average at the regional only or national level

\*\*\*----------- Average Price from own production

\*------------------------------------------------------

\*\*\* Prices by region and rural/urban areas

egen mean\_pric\_own\_kg=mean(pric\_own\_kg), by (product spdomain)

order descript untdes ucf\_kg qty\_consown val\_consown pric\_own pric\_own\_kg mean\_pric\_own\_kg mean\_pric\_own\_kg spdomain

codebook mean\_pric\_own\_kg

\* 27726 missing obs over 89597. It is normal that there are more missing than for prices from purchases since there are less

\* cases of own production than purchases

\*\*\* Prices by region

egen mean\_pric\_own\_kg\_reg=mean(pric\_own\_kg), by (product region)

order descript untdes ucf\_kg qty\_consown val\_consown pric\_own pric\_own\_kg mean\_pric\_own\_kg mean\_pric\_own\_kg\_reg region

codebook mean\_pric\_own\_kg\_reg

\* 25494 missing obs over 89597

\*\*\* Prices at the national level

egen mean\_pric\_own\_kg\_nat=mean(pric\_own\_kg), by (product)

order descript untdes ucf\_kg qty\_consown val\_consown pric\_own pric\_own\_kg mean\_pric\_own\_kg mean\_pric\_own\_kg\_nat region

codebook mean\_pric\_own\_kg\_nat

\* 23769 missing obs over 89597

\*\*\*------- Remaining missing average prices for price from own consumption

\*\*\*-------------------------------------------------------------------------

tab descript if mean\_pric\_own\_kg\_nat==.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

122 Fresh Fish | 1,167 4.91 4.91

123 Dry/ Smoked fish | 1,892 7.96 12.87

126 Infant Formula Foods | 38 0.16 13.03

139 Other vegetables | 2,567 10.80 23.83

147 Sugar | 4,261 17.93 41.76

148 Coffee | 512 2.15 43.91

149 Tea | 4,037 16.98 60.89

150 Salt | 6,157 25.90 86.80

155 Cigarettes | 555 2.33 89.13

156 Other Tobacco | 833 3.50 92.64

157 Food | 894 3.76 96.40

158 soda | 292 1.23 97.63

160 Other juice | 139 0.58 98.21

161 Other foods | 425 1.79 100.00

------------------------------------+-----------------------------------

Total | 23,769 100.00

\*/

\*\*\*------ Fresh Fish, Dry/Smoked Fish, Infant formula foods, Sugar, Coffee, Tea, Salt, Soda

\* The average price is missing for these items because there are no observations for them in consumption of own produce with unit

\* conversion factors.

\*!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

\*!!!!!!!!!!!!!--> I will use the price from purchases. It is overestimating the cost, but it is the best option for now

\*\*\*------ Other vegetables

\* I will use the average price of all vegetables

egen mean\_pric\_own\_veg\_a=mean(pric\_own\_kg) if veggie==1, by (spdomain)

egen mean\_pric\_own\_veg=min(mean\_pric\_own\_veg\_a), by (spdomain)

order product veggie mean\_pric\_own\_veg\_a mean\_pric\_own\_veg

codebook mean\_pric\_own\_veg

\* No missing obs. No need to compute the average at the regional only or national level

\*\*\* Food/Other Foods

\* The problem here arises because there are no units for "food" or "other foods". So I can not use the average price of some units to complete it

\* for those whose ucf is missing

\* I will compute the average price of all foods

egen mean\_pric\_own\_food\_a=mean(pric\_own\_kg) if fooditem==1, by (spdomain)

egen mean\_pric\_own\_food=min(mean\_pric\_own\_food\_a), by (spdomain)

order product fooditem mean\_pric\_own\_food\_a mean\_pric\_own\_food

codebook mean\_pric\_own\_food

\* No missing obs. No need to compute the average at the regional only or national level

\*\*\*------ Other juice

\* The problem here arises because there are no units for "other juice". So I can not use the average price of soda and other soft drinks

egen mean\_pric\_own\_jui\_a=mean(pric\_own\_kg) if juice==1, by (spdomain)

egen mean\_pric\_own\_jui=min(mean\_pric\_own\_jui\_a), by (spdomain)

order product juice mean\_pric\_own\_jui\_a mean\_pric\_own\_jui

codebook mean\_pric\_own\_jui

\* 2138 missing obs over 89597

egen mean\_pric\_own\_jui\_a\_reg=mean(pric\_own\_kg) if juice==1, by (region)

egen mean\_pric\_own\_jui\_reg=min(mean\_pric\_own\_jui\_a\_reg), by (region)

order product juice mean\_pric\_own\_jui\_a\_reg mean\_pric\_own\_jui\_reg

codebook mean\_pric\_own\_jui\_reg

\* No missing obs. No need to compute the average at the national level

\*\*\*------- Other Fruits

\* I computed the average price for other fruits, but it is not reliable because it is based only (at most) on three observations

\* of "other fruits" given in "kg". I will instead use the average price of all the other fruits.

egen mean\_pric\_own\_frui\_a=mean(pric\_own\_kg) if fruit==1, by (spdomain)

egen mean\_pric\_own\_frui=min(mean\_pric\_own\_frui\_a), by (spdomain)

order product fruit mean\_pric\_own\_frui\_a mean\_pric\_own\_frui

codebook mean\_pric\_own\_frui

\* No missing obs. No need to compute the average at the regional only or national level

\*----------------------------------------

\* Final Price per kg

\*----------------------------------------

\*\*\*------------ Price of purchased items

gen double finpric\_purch\_kg=pric\_purch\_kg

replace finpric\_purch\_kg=mean\_pric\_purch\_kg if finpric\_purch\_kg==. /\* Average price by region and rural/urban areas \*/

replace finpric\_purch\_kg=mean\_pric\_purch\_kg\_reg if finpric\_purch\_kg==. /\* Average price by region \*/

replace finpric\_purch\_kg=mean\_pric\_purch\_kg\_nat if finpric\_purch\_kg==. /\* Average price at the national level \*/

replace finpric\_purch\_kg=mean\_pric\_purch\_veg if finpric\_purch\_kg==. & product==139 /\* Other vegetables \*/

replace finpric\_purch\_kg=mean\_pric\_purch\_food if finpric\_purch\_kg==. & (product==157 | product==161) /\* Food/Other Food \*/

replace finpric\_purch\_kg=mean\_pric\_purch\_jui if finpric\_purch\_kg==. & product==160 /\* Other juice \*/

replace finpric\_purch\_kg=mean\_pric\_purch\_frui if product==134 & untcd~=1 /\* Other fruits \*/

codebook finpric\_purch\_kg

tab descript if finpric\_purch\_kg==.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

155 Cigarettes | 555 39.99 39.99

156 Other Tobacco | 833 60.01 100.00

------------------------------------+-----------------------------------

Total | 1,388 100.00

\*--> The reason for computing these prices per kg being to obtain quantities in kg in order to compute calorie values, I will

not try to find prices per kg for cigarettes and tobacco

\*/

\*\*\*------------ Price of consumption from own production

gen double finpric\_own\_kg=pric\_own\_kg

replace finpric\_own\_kg=mean\_pric\_own\_kg if finpric\_own\_kg==. /\* Average price by region and rural/urban areas \*/

replace finpric\_own\_kg=mean\_pric\_own\_kg\_reg if finpric\_own\_kg==. /\* Average price by region \*/

replace finpric\_own\_kg=mean\_pric\_own\_kg\_nat if finpric\_own\_kg==. /\* Average price at the national level \*/

\* Fresh Fish, Dry/Smoked Fish, Infant formula foods, Sugar, Coffee, Tea, Salt, Soda

replace finpric\_own\_kg=finpric\_purch\_kg if finpric\_own\_kg==. & (product==122 | product==123 | product==126 | product==147 | ///

product==148 | product==149 | product==150 | product==158)

replace finpric\_own\_kg=mean\_pric\_own\_veg if finpric\_own\_kg==. & product==139 /\* Other vegetables \*/

replace finpric\_own\_kg=mean\_pric\_own\_food if finpric\_own\_kg==. & (product==157 | product==161) /\* Food/Other Food \*/

replace finpric\_own\_kg=mean\_pric\_own\_jui if finpric\_own\_kg==. & product==160 /\* Other juice \*/

replace finpric\_own\_kg=mean\_pric\_own\_jui\_reg if finpric\_own\_kg==. & product==160 /\* Other juice (avg by region) \*/

replace finpric\_own\_kg=mean\_pric\_own\_frui if product==134 & untcd~=1 /\* Other fruits \*/

codebook finpric\_own\_kg

tab descript if finpric\_own\_kg==.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

155 Cigarettes | 555 39.99 39.99

156 Other Tobacco | 833 60.01 100.00

------------------------------------+-----------------------------------

Total | 1,388 100.00

\*--> The reason for computing these prices per kg being to obtain quantities in kg in order to compute calorie values, I will

not try to find prices per kg for cigarettes and tobacco

\*/

\*----------------------------------------

\* Quantities in kg

\*----------------------------------------

/\* The strategy for getting quantities in grams is as follows:

- use the consumption quantities given when unit conversion factors are available

- when unit conversion factors are not available: use the value of the consumption

and the price per kg (imputed from the observations where unit conversion factors are available)

to compute quantities per kg

\*/

\*\*\*------------- Consumption from purchases

gen double qty\_purch\_kg = qty\_conspurch\*ucf\_kg if qty\_conspurch~=. & ucf\_kg~=.

replace qty\_purch\_kg = val\_conspurch/finpric\_purch\_kg if qty\_conspurch~=. & qty\_purch\_kg==.

tab descript if qty\_purch\_kg==. & qty\_conspurch~=.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

155 Cigarettes | 515 42.08 42.08

156 Other Tobacco | 709 57.92 100.00

------------------------------------+-----------------------------------

Total | 1,224 100.00

\*--> OK. All food items purchased have been converted into quantities in kg

\*/

\*\*\*------------- Consumption from own production

gen double qty\_own\_kg = qty\_consown\*ucf\_kg if qty\_consown~=. & ucf\_kg~=.

replace qty\_own\_kg = val\_consown/finpric\_own\_kg if qty\_consown~=. & qty\_own\_kg==.

tab descript if qty\_own\_kg==. & qty\_consown~=.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

155 Cigarettes | 7 7.95 7.95

156 Other Tobacco | 81 92.05 100.00

------------------------------------+-----------------------------------

Total | 88 100.00

\*--> OK. All food items have been converted into quantities in kg

\*/

\*\*-------------- Consumption of items received in kind/free

\* I will use prices from purchases

gen double qty\_free\_kg = qty\_consfree\*ucf\_kg if qty\_consfree~=. & ucf\_kg~=.

replace qty\_free\_kg = val\_consfree/finpric\_own\_kg if qty\_consfree~=. & qty\_free\_kg==.

tab descript if qty\_free\_kg==. & qty\_consfree~=.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

155 Cigarettes | 35 21.34 21.34

156 Other Tobacco | 129 78.66 100.00

------------------------------------+-----------------------------------

Total | 164 100.00

\*--> OK. All food items have been converted into quantities in grams

\*/

\*\*-------------- Consumption of items away from home

/\*

\* Items consumed outside the home are usually more expensive. I will use the difference in prices per kg from the data set

to evaluate the % difference in prices of items purchased or consumed away from home

\*/

gen test=1 if qty\_consaway~=. & qty\_conspurch~=. & ucf\_kg~=.

gen double price\_away\_kg= val\_consaway/ (qty\_consaway\* ucf\_kg)

order test qty\_conspurch val\_conspurch pric\_purch\_kg qty\_consaway val\_consaway price\_away\_kg product ucf\_kg

\*--> The prices per kg are the same. It seems that when this file was constructed the same prices were used for food

\* purchased and for food consumed away from home. I will therefore use prices of items purchased for the items consumed

\* away from home

drop test price\_away\_kg

gen double qty\_away\_kg = qty\_consaway\*ucf\_kg if qty\_consaway~=. & ucf\_kg~=.

replace qty\_away\_kg = val\_consaway/finpric\_purch\_kg if qty\_consaway~=. & qty\_away\_kg==.

tab descript if qty\_away\_kg==. & qty\_consaway~=.

/\*

descript | Freq. Percent Cum.

------------------------------------+-----------------------------------

155 Cigarettes | 53 65.43 65.43

156 Other Tobacco | 28 34.57 100.00

------------------------------------+-----------------------------------

Total | 81 100.00

\*--> OK. All food items have been converted into quantities in kg

\*/

\*-------------------------------------------------

\* Preparation of the file for the GAPP project

\*-------------------------------------------------

/\* Note: expenditure on food, beverages and tobbaco was recorded during the last seven days

so I will divide the amounts and quantities by 7 in order to have daily values \*/

\*--------- Food VS Nonfood item

gen food\_cat=(product~=153 & product~=154)

tab food\_cat,m

\*--------- Creation of a variable to aggregate the total value of FOOD consumption

\* As per the instructions of the GAPP project, items received in kind are not to be included

gen double valuez= (val\_conspurch + val\_consaway + val\_consown)/7 if food\_cat==1

label variable valuez "Amount paid in food buying transaction: national currency"

\*--------- Creation of a variable to aggregate the total quantity of FOOD consumption

\* As per the instructions of the GAPP project, items received in kind are not to be included

gen double quantityz= (qty\_conspurch + qty\_consaway + qty\_consown)/7 if food\_cat==1

label variable quantityz "Quantity of food in the transaction: Kilogrammes"

keep hhid product food\_cat valuez quantityz

label variable food\_cat "0,1: Food product or not"

\*\*\*\*\*\*\* Clean up

erase "$path\_ug\work\2009\temp\_A4\_1.dta"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Table A5: Amount and Quantity of products (food as well as non food) - Household and Product level

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\* Non-Durable Goods and Frequently Purchased Services during the last 30 Days

\* Note: I copied the labels into excel so as to create a variable with the description of the codes, starting with the numerical code

\* as per GAPP's specifications. This file's name is "descript\_30days"

clear

use "$path\_ug\in\2009\HSEC10B\_CLN.dta"

rename hh hhid

label variable hhid "Household ID"

rename h10bq2 product

label variable product "Product code: numerical"

sort product

merge m:1 product using "$path\_ug\in\2009\descript\_30days.dta"

drop \_merge

sort hhid product

\*\*\* Rent

\* I want to see if there are households with missing rent

save "$path\_ug\work\2009\temp\_A5\_1.dta",replace

gen produ\_rent=(product==301 | product ==302 | product ==303)

egen double exp\_rent= rowtotal(h10bq5 h10bq7 h10bq9)

collapse (sum) exp\_rent, by(hhid produ\_rent)

keep if produ\_rent==1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Arrêt ici le 4 Janvier