Notes on the Cleaning Procedures of Consumption Expenditure, Individual and Household, as well as the Labor Force Data

Reno Dewina

Philippines 1997

39,298 households

File Name	Comments
00_basic info	-There are two different type of raw datasets: household data (FIES 1991_v1) and individual data (LFS Oct 1997.dta.). They are combined together to get a complete picture named as 1997 LFS-FIES_clean.dta. Hid was generated by combining several components regn_s+lprov+lmun+lbgy+lea+lhcn. -Hhsize_s has been provided which includes boarders, domestic helpers and non- relatives. -Final weights and raising factors yield the closest population. Household weight (wta_hh)= rfadj/10000. - Household weights yield a close approximation to the total population in 1997 which is 70.3 million compared to 73 million based on data from WDI.
01_expfood	**-First and foremost, we treat missing values in expenditure data as zero, because they are missing likely because the household did not consume any. This rule applies to all expenditure data aggregation programs thereafter.** -Use cleaned version of consumption data FIES 1997v1_clean. -Consumption data are mostly aggregated by categories and calculated annually regardless of the recall periods in the questionnaire. -Food categories: cereals, roots, fruit, meat, dairy, fish, cofct, and food not elsewhere classified (fdnec). -Food eaten outside home should be included in restaurant and catering expenses (11_exphotl) -Non-alcoholic beverages: aggregated as nonal. -There is no separation between food that are purchased or self-produced.
02_expalch	OK
03_expclth	-The expenses on clothing and footwear are combined together under clothes category. ---Not clear what exactly included under the broaden category HSCLOTH.
04_exphous	-Rent is not specific mentioned whether actual or imputed, but we assume it is an actual rent. Thus rent=rent_act and rent_ -TOTHOUS does not include rent, but the country aggregate does. -Components included in the TOTHOUS are expenses on house repair and fuel.
05_expfurn	-TOTFURN is generated by combining non-durable furniture (ndfur), durable furniture (dufur) and (hoper). -No detail information on what type of furniture included under each category above
06_exphlth	-TOTHLTH represents aggregate of health expenses. In the dataset, it is aggregated under medic variable.
07_exptrsp	-Transportation and communication are combined together as trcom. -No more detailed information available
08_expcmnq	-Since communication variable has been put together under transportation, there's nothing much to do to generate this file. -This file basically contains of all missing values.
09_exprcre	-TOTRECTN is generated based on rcrtn -Rcrtn represents the aggregate of recreation expenses.
10_expeduc	-Education expenses (TOTEDUC) is equal with educ. -Educ is the aggregate of education expenses.
11_exphotl	- fdout is categorized under Hotcat, and hotacc is missing; - No data available on hotel expenses

	-So TOTHOTL should be equal to hotcat
12_expmisc	OK
13a_Pindex1	-Food prices and quantities consumed are in the questionnaire, but not in the data.
13_fdpindex	
14_hhexp	OK
PHL_1997_E	-No regional food price deflators.
PHL_1997_I	-The main problem in handling the datasets because there is no questionnaire available for the PHL 1997 data. -Thus, we only rely on some kind of coding book in analyzing them. - Limit the max age to 98 -In the marital status, add 7 "Unknown". -There's no information on language, ethnicity and religion -For education there is no variable in the dataset that distinguishes who is currently attending school or not (variable "atschool"). Several attempts were made to create atschool variable, for example using responses to a question "why are you not looking for a job?" and those who reported being students can be counted as atschool=1. However, this question was not applied for people who are aged below 10. The other variable that we tried to use to figure out who is currently in school, was the grade variable (in combination with age). The code book that we have only indicates that the grade represents "highest completed grade". Using the grade variable also did not work completely because we have to make too many assumptions. -There's no direct information for whether ever attended school. Consider everybody who was currently in school, or who reported a degree as ever attended. Caveat: People without a degree may have attended school. -Education levels are determined based on highest degrees. People with no degrees are not necessarily uneducated. Change the value code "0" to "No degree". -In Philippines education system, high school follows elementary school, and lasts for four years. - High school is not distinguished as lower and upper secondary schools. Count the first two years as lower secondary and the last two years as upper secondary. -Not in all levels in our data we know which year the individual is in. For currently enrolled students, we can use the age to separate them -We used the age variable to divide the high school graduates into lower and upper levels. Specifically, elementary school graduates and high school undergrads of age <=14 are counted as "Completed primary, but less than completed lower secondary". -High school undergrads of age >=15 and currently in school are counted as "Completed lower secondary (or post-primary vocational), but less than completed senior secondary". The condition of "currently in school" has to be added because there may be older adults who have only finished two years of high school, and thus cannot be counted as upper secondary education even though they are older than 15. This step excludes those whose highest education degrees are high school undergraduates. Thus, the education variables, EDLEVEL_EA and EDLEVEL, are incomplete. -Another variable, EDLEVEL_PHL, is created (_PHL indicates this variable is Philippines specific), which does not separate secondary education into lower and upper levels. -EDYEARS is constructed for those who have completed their education, based on the theoretical length of each level of education -Based on discussion with Juan (3/30), grade=1 (in the coding book) which represents grade 1-3, we take the average of 2 for non-current student, and 1 (2-1) if they are currently at school. Similarly, for grade=5 (in the coding book) which represents 1st to 3rd year of high school, we will count 8 years for non-current student and 7 (8-1) for current students. -Elementary undergraduates could be grade 1 to 5. We don't have further information, so consider them 5 years of education. Elementary grad is 6 years. Likewise, high school undergraduates could be in year 1 to 3 in high school. Consider them 9 years of education. High school grad is 10 years. College undergrads could

	be in year 1 to 3. Consider them 13 years of education. College grad is 14 years. Masters degrees and PhD are bundled together. Consider them 17 years of education. -Wage data in the survey is daily. Annualized wage is calculated as (pay per day)*(days worked last week)*(365/7), of which, days worked last week = (hours worked last week)/(normal hours worked per day during last week).
PHL_1997_G	-Because the individual level data are collected from the LFS, there are no additional variables, such as fertility, that are good for gender analysis. Only the raw values of sectors and occupations are extracted in addition to the _I file.
PHL_1997_E	No regional food price deflator
PHL_1997_H	-Light materials (cogon, nipa, anahaw) for walls and roofs are a type of grass. Put them in bamboo in our category. -Our categories for construction materials of roofs do not differentiate salvaged materials from grass/leaves. So have to put light materials and salvaged materials from the raw data together in the standardized file. -For source of water, tubed/piped well either owned or shared is considered as a protected well -For the actual distance to water source, need to clarify the unit. No questionnaire is found. -Household availability of electricity is a Yes/No indicator. Take Yes as Central/local system. -Water-sealed toilets are considered as flush toilets. Closed pits are considered as improved latrine pits. -Employment of household heads is from individual employment of the past 7 days.