

Instructions

1. Please do not begin until given permission.
2. This exam is a closed-book exam. The only reference you can use during the exam is the help function inside of Stata. You may use an English dictionary (either electronic or paper-based one). You may also use a calculator as long as it is without a memory device in it. (You cannot use a calculator in a cell phone.)
3. **Please work on the desktop of the computer assigned to you. Below is how to change the working directory to the desktop:**
Double click My Computer ⇒ Local Disk (C:) ⇒ Documents and Settings ⇒ Folder with your user name ⇒ Desktop
Copy the directory path and paste it into Stata after the command `cd`. Please do not forget to use double quotes (" ") when you paste the directory path into Stata after the command `cd`.
Now, the working directory has been set to the desktop. Please do not change the working directory from the desktop during the exam.
4. There are 10 questions. The points allocated to each question are shown beside the question number. The full score is 24.
5. Type your name and student ID in the first line of your `do` file using `*` (Remember that Stata ignores the line starting with `*`) or whatever appropriate codes.
6. You need to turn in your `txt` file as well as `do` file at the end of the exam. **The name of your `do` file must be exam.do. The name of your `txt` file must be exam.txt.**
Your `txt` file must be created from your `do` file by using the commands `do` and `translate`. In creating your final `smcl` file (to be translated into your final `txt` file), you must literally use the command `do` to run your completed `do` file. Repeatedly highlighting a part of a `do` file and creating a final `smcl` file is NOT allowed.
7. You need to write Stata commands that implement the tasks given in the questions. You must write your commands in a `do` file in which you must clearly show question numbers using `*`.
8. Please reduce error commands as much as possible in your `do` file. If your `do` file contains error commands that can easily be removed, your credits will be lowered substantially.
9. If you need a few steps to obtain the requested answer, you must show ALL of your work.
10. Some questions are sequential, meaning that, for example, question 6 uses answer

to question 3. If you cannot answer question 3 but you can answer question 6, type your answer to question 6 using *. If your answer to question 6 is correct, I will give you the full credit to question 6 no matter whether you are correct in question 3 or not.

11. You must stop working when you are asked to stop. Literally, no typing any more. You must have finished converting your *smcl* file into a *txt* file by the end time of the exam.
12. **Please neither log out nor shut down your computer when you leave the room. The instructor will shut down your computer after confirming that your *do* and *txt* files have been collected.**
13. All students cannot take this exam simultaneously. To make the exam fair to everybody, you must not discuss the contents of this exam with other classmates until the second session of the exam is over.
14. You must follow the above 13 rules. Breaking any single rule automatically reduces your credits to zero as well as potential disciplinary actions taken by the Dean's office.

You must agree to the above requirements of the exam and sign below.

Print your student ID number:

Print your name:

Print today's date:

Sign:

Advice: Please save your *do* file as often as possible. Every year there is one or more students who lose their *do* files because they close their *do* files without saving them.

We discussed insurance tests in class. In particular, we learned the concepts of perfect and imperfect insurance in the context of rural areas in a developing country. In this exam, we are interested in examining the hypotheses of perfect insurance and imperfect insurance, using data from rural Indonesia. You are supposed to do some data-preparation work and run an OLS regression that is relevant to the research question. Empirical results (to be obtained) may not be necessarily consistent with theoretical expectations. Do not worry about the empirical results themselves. Rather, what's important is whether you understand basic Stata skills and econometric concepts learned in class.

We are going to use two data files (*hhinc.dta* and *hhexp.dta*). These data files were created from the rural sample of the Indonesia Family Life Survey (IFLS). Survey interviews were fielded in 1997 and 2000. These data are called panel data, in which the same sample households were surveyed in both 1997 and 2000. Both data files are household-level (each row represents a single household). There are no missing or irrelevant observations in both data files, so you do not need to worry about complications which arise from missing/irrelevant observations.

There are 3 variables in *hhinc.dta*. Below is a description of each variable.

- *id* is household id number.
- *rhinc1997* is total household income in 1997 in real terms (unit: Rupiah which is the currency in Indonesia).
- *rhinc2000* is total household income in 2000 in real terms (unit: Rupiah).

There are 3 variables in *hhexp.dta*. Below is a description of each variable.

- *id* is household id number.
- *rhexp1997* is total household expenditure in 1997 in real terms (unit: Rupiah).
- *rhexp2000* is total household expenditure in 2000 in real terms (unit: Rupiah).

Price differences across times and places have been adjusted in calculating *rhinc1997*, *rhinc2000*, *rhexp1997*, and *rhexp2000*. That is, all of these variables are in real terms.

1) 2 points

Check the number of observations, mean, minimum, and maximum of each of *rhinc1997* and *rhinc2000*. Just write a Stata command that shows you the number of observations, mean, minimum, and maximum of each variable.

2) 2 points

Merge the two data files, so that each sample household has *rhhinc1997*, *rhhinc2000*, *rhhexp1997*, and *rhhexp2000*.

3) 2 points

Report the numbers of sample households in the following three categories. (i) sample households which have both income and expenditure data, (ii) sample households which have income data but do not have expenditure data, and (iii) sample households which have expenditure data but do not have income data. Explain the command result using `*`.

4) 2 points

The household id variable *id* has 6 digits. The first 4 digits in *id* is community id. Create a community id variable which should be the first 4 digits in *id*. The name of the community id variable must be *commid*. Hint: You can use *substr* under the command *generate*. Use the help function. Search 'functions' and see help for 'string functions.'

5) 3 points

Create sample means of *rhhexp1997* by community. That is, you need to create community-specific means of *rhhexp1997*. The name of the new variable must be *rcomexp1997*. Repeat this for *rhhexp2000*. That is, you need to create community-specific means of *rhhexp2000*. The name of the new variable must be *rcomexp2000*.

6) 2 points

Calculate change in household income for each household. That is, calculate *rhhinc2000-rhhinc1997*. The name of the new variable must be *diff_rhhinc*. Similarly, create *diff_rhhexp=rhhexp2000-rhhexp1997* and *diff_rcomexp=rcomexp2000-rcomexp1997* for each household.

7) 3 points

We are going to run the following OLS regression.

$$\text{diff_rhhexp} = \beta_0 + \beta_1 \text{diff_rhhinc} + \beta_2 \text{diff_rcomexp} + \varepsilon$$

We assume away econometric identification problems (reverse causality and omitted variables). According to this model, household expenditure must be financed by what

and what in the context of rural areas of a developing country? Answer two sources from which household expenditure must be financed. Type your answer using *.

8) 3 points

The concept of perfect insurance predicts what values of β_1 and β_2 ? The concept of imperfect insurance predicts what values of β_1 and β_2 ? Type your answer using *.

9) 2 points

Run an OLS regression on the model shown in 7).

10) 3 points

Carefully interpret the OLS coefficient $\hat{\beta}_1$. You must discuss the magnitude of the coefficient estimate. If you cannot run a regression in 9) above, use $\hat{\beta}_1 = 0.001$ for this question. The unit of *diff_rhhexp*, *diff_rhhinc*, and *diff_rcomexp* is Rupiah which is the currency in Indonesia. Type your answer using *.

(End of Stata Exam 2011)