Econ 665

Assignment #3:

Due Date: December 7, 2017

General Instructions

Please answer the following questions:. You should submit your answers in a well-written text, using Stata output as well as the do file that you used to get the answers. In your report, you can use another document for your analysis but make sure it answers the questions in each section below. While your answers should be submitted individually and **written independently**, you are allowed to work in cooperation with your classmates. It is expected, however, that you give proper credit to your classmates who facilitated your understanding.

Female Participation in Microcredit program: Another Look

Recall the data set in problem 1 and 2. It is a data set from the Bangladesh Household Survey. The information was collected at the household and community levels in 1998 (the data set we used in the previous assignments) but baseline information was also collected in 1991. There is a panel data set that we will use for this question.

We want to use this data set to start thinking of what one can do when one wants to measure the impact of a program on its participants (potential and actual participants). The program in this case is microcredit and we want to assess whether it leads to an increase in expenditures.

If you recall, some of these households have members who are participants in a micro-credit project while some households do not have participants. The variable "dfmd" is the indicator variable if the household have a female member who is a micro-credit participant.

One way of measuring the impact of a program is by using panel data, collected from a baseline survey before the program was implemented and after the program has been operating for some time. These two surveys should be comparable in the questions and survey methods used and must be administered to both participants and nonparticipants.

For this question, we do not assume that the counterfactual was obtained through randomization. We will use a different approach that we saw in class in order to deal with the unobserved bias that may affect the estimation of the impact of the

programme. Using the panel data allows elimination of unobserved variable bias, provided that it does not change over time This approach, the double-difference (DD, also commonly known as difference-in-difference, or dif-in-dif) method has been popular in nonexperimental evaluations.

The DD method estimates the difference in the outcome during the postintervention period between a treatment group and comparison group relative to the outcomes observed during a preintervention baseline survey. The allocation into the treatment group vs the control group was not randomized.

Question 1. Simple DD (20%)

The simplest way of calculating the DD estimator is to manually take the difference in outcomes between treatment and control between the surveys. The panel data hh_9198.dta are used for this purpose.

Please find attached the panel data set hh_9198.dta .

use hh_9198

Question 1a. Please find some summary statistics: How many observations? How many households? What is the mean of total expenditures (exptot) overall and by year ?

Question 1.b We want to take the difference between the two years. The problem is that since the data are at the level of household, year, one cannot simply make the difference between 1998 and 1991.

Here's how to do it using stata. The following commands open the data file and create a new 1991-level outcome variable (per capita expenditure) to make it available in observations of both years. Then, only 1998 observations are kept, and a log of per capita expenditure variable is created; the difference between 1998 and 1991 per capita expenditures (log form) is created.

```
use hh_9198
gen exptot0=exptot if year==0
egen exptot91=max(exptot0), by(nh)
keep if year==1
gen lexptot91=ln(1+exptot91)
gen lexptot98=ln(1+exptot)
gen lexptot9891=lexptot98-lexptot91
```

Please explain in simple language how the above commands worked.

Question 1.c : Use the "ttest") and take the difference variable of outcomes created earlier ("lexptot9891") and compare it for microcredit participants and nonparticipants. In essence, test if "lexptot9891" for those with dfmfd=1 is different from those with dfmfd==0.

What do you find?

Question 2. Regression Implementation (25%)

Instead of manually taking the difference of the outcomes, DD can be implemented with a regression using the panel data set.

On the basis of the discussion in Ravallion (2008), the DD estimate can be calculated from the regression

$$Y_{it} = a + DD * T_i t + \beta * T_i + \theta t + \varepsilon_{it}$$

where T_i is the treatment variable, t is the time dummy, and the coefficient of the interaction of t and T_i (i.e. DD) gives the estimate of the impact of treatment on outcome Y.

The following commands open the panel data file, create the log of outcome variable, and create a 1998-level participation variable available to both years—that is, those who participate in microcredit programs in 1998 are the assumed treatment group.

```
use hh_9198,clear
gen lexptot=ln(1+exptot)
gen dfmfd1=dfmfd==1 & year==1
egen dfmfd98=max(dfmfd1), by(nh)
```

The next command creates the interaction variable of treatment and time dummy (year in this case, which is 0 for 1991 and 1 for 1998).

```
gen dfmfdyr=dfmfd98*year
```

The next command runs the actual regression that implements the DD method:

```
reg lexptot dfmfdyr dfmfd98 year
```

Question 2.a : Please discuss the results .

Question 2.b: A basic assumption behind the simple implementation of DD is that other covariates do not change across the years. But if those variables do vary, they should be controlled for in the regression to get the net effect of program participation on the outcome.

So please augment the regression model by including other covariates that may affect the outcomes of interest:

reg lexptot year dfmfd98 dfmfdyr sexhead agehead educhead lnland vaccess pcirr rice wheat milk oil egg [pw=weight]

Please discuss the results.

Question 3: Checking Robustness of DD with Fixed-Effects Regression (30%)

But the results above are just plain OLS. Another way to measure the DD estimate is to use a fixed-effects regression instead of ordinary least squares (OLS). Fixed-effects regression controls for household's unobserved and time-invariant characteristics that may influence the outcome variable.

The Stata "xtreg" command is used to run fixed-effects regression. In particular, with the "fe" option, it fits the fixed-effect models.

xtreg lexptot year dfmfd98 dfmfdyr, fe i(nh)

Question 3.a : Please discuss the results.

Question 3.b By including other covariates in the regression, the fixed-effects model can be extended in the following way:

xtreg lexptot year dfmfd98 dfmfdyr sexhead agehead educhead
lnland vaccess pcirr rice wheat milk oil egg, fe i(nh)

Please discuss the results.

Question 3.c Could you do other estimations in order to estimate the impact of the microcredit program? Which ones? What test(s), if any, would you run? Would you modify the commands above?

Question 4. Assessment of results (25%)

Reflecting on the results you obtained in questions 1, 2 and 3 above, what do you think one can say about the impact of the microcredit program on household expenditures? Should one trust the results using the DD methodology? Why? Or why not? What type of information would you need in order to assess further the validity of the DD methodology to assess the impact of the program in this particular case?