EC230A
Winter 2013

Hilary Hoynes
hwhoynes@ucdavis.edu

Empirical Problem Set 1
Due: Thursday Feb 14

You are welcome to collaborate on the problem set. You can share do/log files but please write up your own results.

This problem set is replication/extension of the Eissa and Liebman paper:

> Eissa and Liebman, "Labor Supply Response to the Earned Income Tax Credit," *QJE*, 1996.

In EL (Eissa&Liebman) they use difference in difference techniques to examine the 1986 expansion of the Earned Income Tax Credit (EITC). In this exercise we apply their technique to analyze the 1993 expansion of the EITC.

We will estimate models like in EL for the 1993 expansion. In EL, the treatment group is single women with children, and the control is single women without children. We will estimate the same model. As an extension we will also estimate models where the treatment effect is allowed to differ between those with 1 versus 2+ children.

We have data for 1991-1996. We will assume that the before expansion years are 1991-1993 and the after expansion sample is 1994-1996.

We will use STATA for the problem set. You should have downloaded the STATA data set **ps1.dta**.

Make sure you attach your Stata code with your completed problem set. That way I can figure out where you went wrong (if you do) …

<u>Data set overview</u>

The data set consists of a sample of single women ages 20-54 with less than a high school education covering the years 1991-1996 from the Current Population Survey.
(Note EL estimates models separately for different education groups. We only have the lowest education group. Ignore their results by education group.)

Here is output from the STATA command DESCRIBE which lists the variables in the data set along with their definitions.

```
. describe

Contains data from ps1.dta
  obs:        13,746
 vars:            11
 size:       618,570 (40.8% of memory free)
```

```
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
state           float  %9.0g                  State of Residence
year            float  %9.0g                  Year [taxyear]
urate           float  %9.0g                  State Unemp Rate
children        byte   %8.0g                  Number of Children
nonwhite        byte   %8.0g                  Dummy=1 if Hispanic/Black
finc            double %10.0g                 Annual Family Income (97$)
earn            double %10.0g                 Annual earnings (97$)
age             byte   %8.0g                  Age of woman
ed              byte   %8.0g                  Years of education
work            byte   %8.0g                  Dummy =1 if Employed last year
unearn          double %10.0g                 Unearned Income (97$)
-------------------------------------------------------------------------------
Sorted by:
```

All dollar amounts are in 1997 dollars.

1) Describe the 1993 expansion of the EITC. What does theory predict about the impact of this EITC expansion on the labor supply of single women with children? Discuss the predictions for the intensive and extensive margins. Would you expect a difference by family size? Why or why not. Explain.

2) Construct variables for difference-in-difference analysis as in EL equation 1. You will need a variable for the "treatment" (call it ANYKIDS) and for after the expansion (call it POST93). Calculate the sample means by presence of children as in Table 1 of EL. Discuss the differences in the sample between single women with and without children. Calculate sample means separately for single women with 1 versus 2 or more children. How do the samples differ?

[Don't worry about the variables that are in EL that are not in the input data set. The variables in their Table I that you do not have are: preschool children, weekly participation, and hours of work. You can create filing unit size = number of children + 1 (for the woman). You can create earnings conditional on work from earnings.]

3) Estimate the impacts of the EITC on employment using a "single difference" model. In particular, take one after year (1996) and compare the employment of single women with kids to single women without kids. Estimate the unconditional "single difference" in a regression framework. What do you find? Given the information in (2) above, how might this estimate be biased?

4) Create a graph which plots mean annual employment rates by year (1991-1996) for single women with children (treatment) and single women without children (control). Discuss the differences that show up in the graph. Use this information to critique the validity of using single women without children as the control group. In particular examine the "pre-treatment" trends and how they differ by group.

5) Given the difference in mean employment rates by group in (4) it may be hard to analyze the results. Instead, estimate a regression with a full set of time dummies

and another set of time dummies interacted with ANYKIDS. Plot the set of dummies interacted with ANYKIDS (with year on the X axis). Interpret the dummies. What does this show and how does it differ from (4)?

6) **Calculate the unconditional difference-in-difference estimates of the effect of 1993 EITC expansion on employment of single women as in EL Table II. Also calculate the standard errors. Organize your table as in EL Table II. Explain your results. What is the estimated treatment effect? How does it compare to the magnitude of the treatment effect in EL? Do you expect the 1993 expansion to generate larger or smaller effects compared to the 1986 expansion? Why?**

[Use the specification in EL Table II, Panel B, Control group 1. Treatment group is less than high school single women with children; the control group is less than high school single women without children. Before is 91-93, after is 94-96.]

Note: Make sure that you test properly—it is the standard error of the mean not the standard error of the variable.

7) **Recalculate the unconditional difference in difference estimates by allowing the treatment effects to vary for those with 1 and 2 or more children. Again organize your results as in EL Table II. This amounts to considering one treatment group at a time, the control group in both cases is single women without children. Which treatment effect is larger? Is this consistent or inconsistent with the theoretical predictions?**

8) **Estimate conditional difference in difference estimates by running PROBIT regressions as in EL Table III. Estimate model labeled WITHOUT COVARIATES in column (1). What is the treatment effect? How does it compare to the unconditional treatment effect that you found in part (6)? Is this what you expected? Why or why not? What is the interpretation of the coefficient on KIDS? What about POST93?**

Please report dp/dx values for the probit. Alternatively, estimate as a linear probability model (OLS). How do the results differ?

9) **As stated in the notes to EL Table III, all specifications include dummies for 1984, 1985, 1989, 1990. What time dummies do you identify in your model in (8)? Explain. Now, drop the time dummies (but still keep POST and POST*ANYKIDS). How do the results change? Explain.**

10) **Estimate model labeled DEMOGRAPHIC CHARACTERISTICS in column (2). How does the treatment effect change? What do the parameter estimates for NONWHITE and OTHER INCOME tell you?**

The only variable we do not have in the data set is number of preschool children.

11) Add unemployment rate and allow its impact to vary by presence of children. (Like in EL column 3.) How do your results change? What does it tell you about the importance of cycles and how they affect the different groups?

12) Add state fixed effects to model. Given the research design in the paper, would you expect adding state fixed effects to change the EITC effect?

13) Allow the treatment effect to vary by those with 1 or 2+ children. What do you find? Looking back at the EITC expansion, what do you expect to find? Does this make sense?

14) Suppose you are concerned about the validity of women without children as a control group. Instead you use families with 2+ children as the treatment group and families with 1 child as the control group. Estimate the treatment effect controlling for all variables you have in (12). What do you find and how does it differ from the results in (13)? How sensitive are these results to whether you allow the unemployment rate effect to vary with those with 1, vs 2+ children? Discuss.

15) As you did in (5), examine the trends in the main T and C groups. But now do it in the conditional model (conditioning on the Xs you have in 13). How do the trends change?

Hint: Drop POST and replace with a full set of year effects. Drop ANYKIDS and replace with full set of year effects interacted with ANYKIDS.

16) Estimate a "placebo" treatment model. In particular take data from the pre period only. The treatment and control groups are unchanged. But the "placebo post" period begins in 1992 (or you can try 1993). What do you find?