

Econ 758 Homework 1

Ege Can, John Appert

February 22, 2019

1 Question 1

- a Give a short description of the relevant aspects of the EITC expansion in 1993.(Hint: Have a look at Eissa and Hoynes, 2004.) Briefly discuss the theoretical predictions for the impact of the reform on the labor market participation of single women with children. You do not need to present a formal model!

The earned income tax credit began as a reform to traditional welfare programs in that it requires the recipient to earn income in order to receive benefits. This feature of the EITC was designed to counteract disincentives to work in traditional welfare programs.

There was an unintended effect from the program however. While the program provides increasing benefits based on the number of children in a family, it was focused on household income. This means that in a household where the mother is a secondary earner the choice to work is made after the husband's earnings are taken into account. This feature results in the decision for a women to join the labor market to occur only if the husband has not maximized the benefit of EITC.

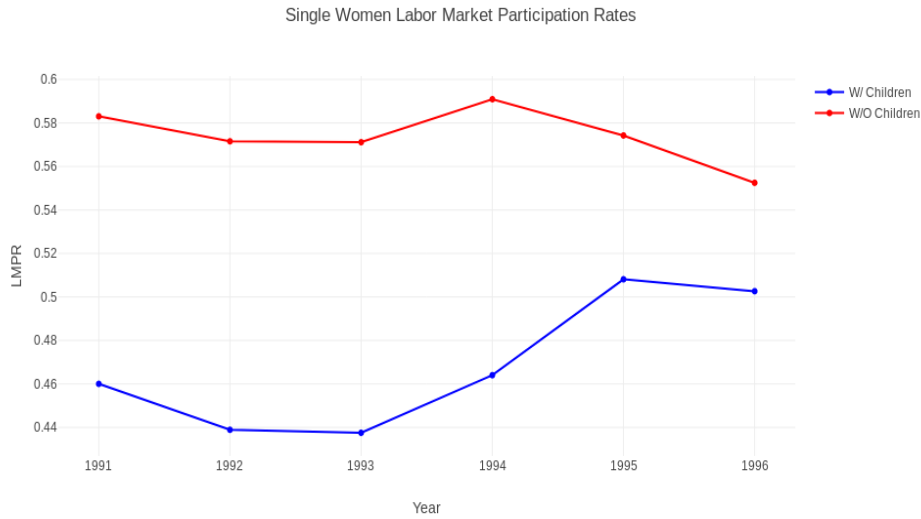
In 1993 the EITC was modified to provide increased benefits to single parents. In this case we would expect the impact single mothers to be significantly positive as we are increasing the benefit for them to work.

- b Would you expect the number of children to influence the size of the effect Why or why not? Explain.
Yes, we expect the number of children to impact the size of the effect. The more children someone has the greater the benefit for entering the workforce. Therefore, as the number of children increases we would expect a greater rise in the labor market participation rate.
- c Generate a table with descriptive statistics (Table 1, structured as in Table I in Eissa and Liebman, 1996), which contains the sample means of the variables nonwhite age ed work earn for two groups: single women with and without children. You do not need to display the standard deviations. Briefly discuss the differences.
- d Now calculate the sample means separately for single women with one child and women with two or more children (add the information to Table 1). How do they differ from each other

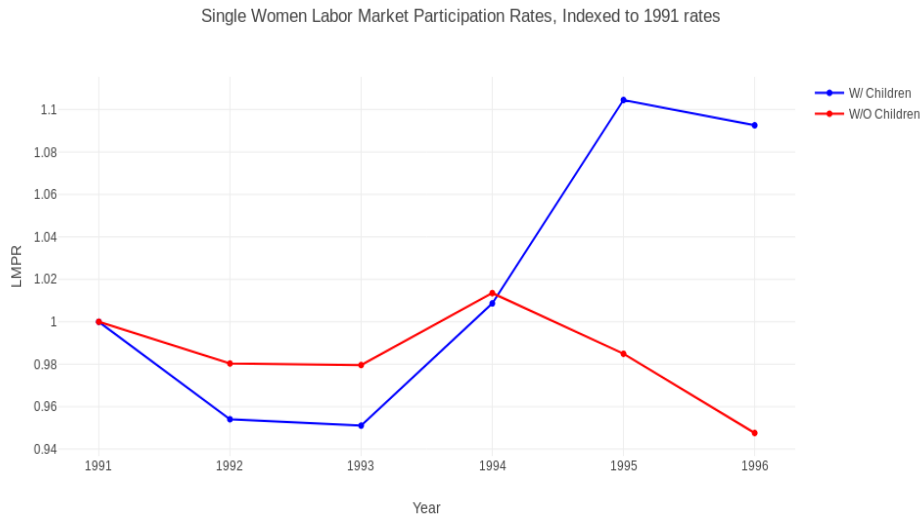
2 Question 2

For the following analysis you need to generate two dummy variables to identify the treatment group (single women with children) [call it child] and the post-treatment period (1994-1996) [call it post1993].

- a Create a figure (Figure 1) that illustrates the annual mean labor market participation rates by year (1991-1996) for single women with children (treatment group) and single women without children (control group).Label the axes and include a title and a legend into the graph.



- b Now normalize the value of the labor force participation rate for each of the two groups to group-specific 1991 values. That is, the mean of the labor market participation rates in 1991 become equal to 1. Plot a graph (as the one before, including labeling, title, and legend) in Figure 2.



- c Based on Figures 1 and 2, discuss the validity of using single women without children as control group. When looking at figure one it is difficult to determine whether or not the idea of using single women without children as a control group is valid. The levels of labor market participation are significantly different and both trends seem similar. However, once we index the labor market participation rate to 1991 and look at changes in the level with respect to 1991 we see that both groups track closely until 1994 when there is a divergence. This implies that we can use single woment without children as a control group.
- d Calculate the sample means of labor force participation rates (work) of women with and without children for the pre-(average over 1991-1993) and post-reform (average over 1994-1996) period. Organize your table (Table 2) as in Table II in Eissa and Liebman (1996).

TABLE II Labor Force Participation Rates of Single Women				
Group	Pre-1993 (1)	Post-1993 (2)	Diff (3)	Diff-in-Diff (4)
With Children	0.446 (0.0076)	0.491 (0.0084)	0.0448	
Without Children	0.575 (0.0088)	0.573 (0.0094)	-0.002	0.047

- e Calculate the within-and between-group differences as well as the unconditional difference-in-differences estimate and add them to Table 2. Briefly comment on your results.

TABLE II Labor Force Participation Rates of Single Women				
Group	Pre-1993 (1)	Post-1993 (2)	Diff (3)	Diff-in-Diff (4)
With Children	0.446 (0.0076)	0.491 (0.0084)	0.0448	0.047
Without Children	0.575 (0.0088)	0.573 (0.0094)	-0.002	
Between Group Differences	-0.129	-0.082		

We find that there is an increase in the labor supply of single women with children from the pre-1993 period to the post-1993 period. This is inline with what we would expect from the theory.

- f Repeat the comparison separately for women with one child and for women with at least two children for the years before and after the EITC expansion. Again compute the within-and between-group differences and the difference-in-differences estimates. Compare each of the two groups separately to single women without children (the control group). Display the results in Table 3 and discuss your findings. For which of the two groups do you find larger treatment effects? Is this consistent with the theoretical predictions?

We find a treatment effect of approximately 3 percent for single women with one child versus a treatment effect of almost 6 percent for women with two children. This is in line with theoretical predictions as the incentive to work increases with the number of children.

TABLE III Labor Force Participation Rates of Single Women				
	Pre-1993 (1)	Post-1993 (2)	Within Group Dif- ference (3)	Difference- in- Differences (4)
A. Treatment and Treat- ment				
Two or more Children	0.396	0.449	0.031	
One Child	0.523	0.554	0.053	0.022
Between Group Differ- ences	-0.127	-0.105		0.022
B. Treatment and Control				
One Child	0.523	0.524	0.031	
No Children	0.575	0.573	-0.002	0.033
Between Group Differ- ences	-0.052	-0.019		0.033
C. Treatment and Control				
Two or More Children	0.396	0.449	0.053	
No Children	0.575	0.573	-0.002	0.055
Between Group Differ- ences	-0.179	-0.124		0.055

g Return to the comparison of women with and without children. Estimate the difference-in-differences effect from the EITC expansion by running OLS regressions. As dependent variable, use the dummy indicating labor market participation(work). First run a regression without controls (“unconditional diff-in-diff estimate”). Then add control variables (urate nonwhite age ed) to obtain the “conditional diff-in-diff estimate”. Present your results (including standard errors) in Table 4 and interpret them. Compare the estimates and their statistical significance for the conditional and unconditional difference-in-differences estimates. Also comment on the estimated coefficients of child and post1993.

TABLE IV		
OLS Results for post 1993 Treatment		
Dependent variable: work		
Variable	Unconditional (1)	Conditional (2)
constant	0.575***(0.009)	0.496
child	-0.129***(0.012)	-0.118***(0.012)
post1993	-0.002(0.013)	-0.0234*(0.014)
interactionterm	0.046***(0.017)	0.049***(0.017)
urate		-0.016***(0.003)
nonwhite		-0.044***(0.009)
education		0.017***(0.002)
age		0.002***(0.001)
Observations	13476	13476
R-Squared	0.0126	0.0273

Standard error in parentheses

***p<0.01, **p<0.05, *p<0.1

- h Estimate a conditional (i.e., including $urate$ $nonwhite$ age ed), “placebo” treatment model on the pre-treatment period. For this purpose, take data from the years 1991-1993 only and leave the treatment and control groups unchanged. Assume for the analysis that the placebo reform would have taken place on January 1st, 1992 (generate a dummy variable $postplacebo$ that is one for year 1992 and after and an interaction with $child$) and present your results (including standard errors) in Table 5. What do you find?

TABLE V		
OLS Results for placebo Treatment		
Dependent variable: work		
Variable	Unconditional (1)	Conditional (2)
constant	0.583***(0.015)	0.540***(0.048)
child	-0.123***(0.020)	-0.111***(0.020)
postPlacebo	-0.012(0.018)	-0.000(0.018)
interactionterm	-0.010(0.024)	0.049(0.024)
urate		-0.021***(0.004)
nonwhite		-0.039***(0.012)
education		0.016***(0.002)
age		0.002***(0.001)
Observations	7,401	7,401
R-Squared	0.0167	0.0312

Standard error in parentheses

***p<0.01, **p<0.05, *p<0.1

We find that when building a "placebo" model there are no significant effects on labor supply of single women with children in the period preceding the change in the EITC. This supports the argument that the change we see in the later period is due to the changes in the EITC and not some other unobserved variable. In this case we find the coefficient of the interaction variable is -0.0127 with a standard error of 0.024.

3 Code

3.1 Regression Outputs from Python

Dep. Variable:	work	R-squared:	0.013
Model:	OLS	Adj. R-squared:	0.012
Method:	Least Squares	F-statistic:	58.45
Date:	Wed, 20 Feb 2019	Prob (F-statistic):	1.54e-37
Time:	08:01:06	Log-Likelihood:	-9884.9
No. Observations:	13746	AIC:	1.978e+04
Df Residuals:	13742	BIC:	1.981e+04
Df Model:	3		

	coef	std err	t	P> t	[0.025	0.975]
const	0.5755	0.009	65.060	0.000	0.558	0.593
parent	-0.1295	0.012	-11.091	0.000	-0.152	-0.107
Post1993	-0.0021	0.013	-0.160	0.873	-0.027	0.023
interact	0.0469	0.017	2.732	0.006	0.013	0.081

Omnibus:	5.965	Durbin-Watson:	1.934
Prob(Omnibus):	0.051	Jarque-Bera (JB):	2175.929
Skew:	-0.051	Prob(JB):	0.00
Kurtosis:	1.054	Cond. No.	7.14

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Dep. Variable:	work	R-squared:	0.027
Model:	OLS	Adj. R-squared:	0.027
Method:	Least Squares	F-statistic:	55.09
Date:	Wed, 20 Feb 2019	Prob (F-statistic):	3.84e-78
Time:	08:04:28	Log-Likelihood:	-9781.8
No. Observations:	13746	AIC:	1.958e+04
Df Residuals:	13738	BIC:	1.964e+04
Df Model:	7		

	coef	std err	t	P> t	[0.025	0.975]
const	0.4959	0.036	13.960	0.000	0.426	0.565
parent	-0.1179	0.012	-9.891	0.000	-0.141	-0.095
Post1993	-0.0234	0.014	-1.730	0.084	-0.050	0.003
urate	-0.0164	0.003	-4.962	0.000	-0.023	-0.010
nonwhite	-0.0445	0.009	-4.945	0.000	-0.062	-0.027
age	0.0020	0.000	4.466	0.000	0.001	0.003
ed	0.0171	0.002	10.477	0.000	0.014	0.020
interact	0.0495	0.017	2.905	0.004	0.016	0.083

Omnibus:	4.872	Durbin-Watson:	1.939
Prob(Omnibus):	0.088	Jarque-Bera (JB):	2046.360
Skew:	-0.046	Prob(JB):	0.00
Kurtosis:	1.112	Cond. No.	330.

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Dep. Variable:	work	R-squared:	0.031
Model:	OLS	Adj. R-squared:	0.030
Method:	Least Squares	F-statistic:	34.06
Date:	Wed, 20 Feb 2019	Prob (F-statistic):	4.84e-47
Time:	08:05:32	Log-Likelihood:	-5254.1
No. Observations:	7401	AIC:	1.052e+04
Df Residuals:	7393	BIC:	1.058e+04
Df Model:	7		

	coef	std err	t	P> t	[0.025	0.975]
const	0.5403	0.048	11.281	0.000	0.446	0.634
parent	-0.1092	0.020	-5.490	0.000	-0.148	-0.070
Post1992	-0.0002	0.018	-0.009	0.993	-0.036	0.036
urate	-0.0210	0.004	-4.750	0.000	-0.030	-0.012
nonwhite	-0.0394	0.012	-3.265	0.001	-0.063	-0.016
age	0.0019	0.001	3.237	0.001	0.001	0.003
ed	0.0157	0.002	7.103	0.000	0.011	0.020
interact	-0.0127	0.024	-0.525	0.599	-0.060	0.035

Omnibus:	0.010	Durbin-Watson:	1.968
Prob(Omnibus):	0.995	Jarque-Bera (JB):	1083.431
Skew:	0.003	Prob(JB):	5.44e-236
Kurtosis:	1.126	Cond. No.	328.

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.2 Python Code from Jupyter Notebook

```
1 import pandas as pd
2 import numpy as np
3
4 #import data visualization library
5 import plotly.plotly as py
6 import plotly.graph_objs as go
7 from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
8
9 #suppress warnings
10 import warnings
11 warnings.simplefilter("ignore")
12
13 #regression
14 import statsmodels.api as sm
15
16 #Read in the data
17 df=pd.read_stata('ps1.dta')
18
19 #Initial data munging
20
21 df['employed']=np.where(df['work']==1,1,0)
22 df['unemployed']=np.where(df['work']==0,1,0)
23 df['parent']=np.where(df['children']!=0,1,0)
24
25 #pivot by year and parent and then reset the index
26 df1=df.groupby(['year', 'parent']).sum()
27 df1=df1.reset_index()
28
29 #calculate the lfpr for both parents and no parents
30 df1['urate']=(df1['employed'])/(df1['employed']+df1['unemployed'])
31 parent=df1[df1['parent']==1]
32 nparent=df1[df1['parent']==0]
33
34 #Generate figure 1
35 # Add data
36 year = parent['year']
37 parentLMPR= parent['urate']
38 nparentLMPR = nparent['urate']
39
40
41 # Create and style traces
42 trace0 = go.Scatter(
43     x = year,
44     y = parentLMPR,
45     name = 'W/ Children',
46     line = dict(
47         color = ('blue'),
48         width = 2)
49 )
50 trace1 = go.Scatter(
51     x = year,
52     y = nparentLMPR,
53     name = 'W/O Children',
54     line = dict(
55         color = ('red'),
56         width = 2,)
57 )
58
59
60 data = [trace0, trace1]
61
62 # Edit the layout
63 layout = dict(title = 'Single Women Labor Market Participation Rates',
64               xaxis = dict(title = 'Year'),
65               yaxis = dict(title = 'LMPR'),
66               )
```



```

67
68 fig = dict(data=data, layout=layout)
69 py.iplot(fig, filename='raw-plot')
70
71 #Reindexing to 1991
72 pBaseLevel=parent.iloc[0,3]
73 nBaseLevel=nparent.iloc[0,3]
74 parent['index']=parent['urate']/pBaseLevel
75 nparent['index']=nparent['urate']/nBaseLevel
76
77 #Generate figure 2
78 # Add data
79 year = parent['year']
80 piLMPR= parent['index']
81 niLMPR = nparent['index']
82
83
84 # Create and style traces
85 trace0 = go.Scatter(
86     x = year,
87     y = piLMPR,
88     name = 'W/ Children',
89     line = dict(
90         color = ('blue'),
91         width = 2)
92 )
93 trace1 = go.Scatter(
94     x = year,
95     y = niLMPR,
96     name = 'W/O Children',
97     line = dict(
98         color = ('red'),
99         width = 2,)
100 )
101
102
103 data = [trace0, trace1]
104
105 # Edit the layout
106 layout = dict(title = 'Single Women Labor Market
107     Participation Rates, Indexed to 1991 rates',
108     xaxis = dict(title = 'Year'),
109     yaxis = dict(title = 'LMPR'),
110 )
111
112 fig = dict(data=data, layout=layout)
113 py.iplot(fig, filename='index-plot')
114
115 #Calculating diff-in-diff
116 parent=df[df['parent']==1]
117 nparent=df[df['parent']!=1]
118
119 #calculate the average of the treatment group pre-1994
120 tc1=parent[parent['year']<1994]
121 tc1_empl=tc1['work'].sum()
122 tc1_mean=tc1_empl/len(tc1)
123
124 #calculate the average of the treatment group post-1994
125 tc2=parent[parent['year']>1993]
126 tc2_empl=tc2['work'].sum()
127 tc2_mean=tc2_empl/len(tc2)
128
129 #calculate the average of the control group pre-1994
130 cg1=nparent[nparent['year']<1994]
131 cg1_empl=cg1['work'].sum()
132 cg1_mean=cg1_empl/len(cg1)
133
134 #calculate the average of the control group post-1994

```

```

135 cg2=nparent[nparent['year']>1993]
136 cg2_empl=cg2['work'].sum()
137 cg2_mean=cg2_empl/len(cg2)
138
139 #calculate diffs
140 dif1=tc2_mean-tc1_mean
141 dif2=cg2_mean-cg1_mean
142 dif_dif=dif1-dif2
143
144 #print (tc1_mean, tc2_mean, cg1_mean, cg2_mean)
145
146 l1=["Treatment Group", len(parent), tc1_mean, tc2_mean, dif1, '']
147 l2=["Control Group", len(nparent), cg1_mean, cg2_mean, dif2, dif_dif]
148
149 table=[l1, l2]
150
151 headers=['Group', 'Sample Size', 'Pre-1993',
152          'Post-1993', 'Difference', 'Difference-in-differences']
153
154 table2=pd.DataFrame(table, columns=headers)
155
156 #table2
157
158 #diff-in-diff w/ one child and two children
159
160 one_child=df[df['children']==1]
161 two_child=df[df['children']>1]
162
163 #calculate the average of the treatment group with one child pre-1994
164 tglc1=one_child[one_child['year']<1994]
165 tglc1_empl=tglc1['work'].sum()
166 tglc1_mean=tglc1_empl/len(tglc1)
167
168 #calculate the average of the treatment group with one child post-1994
169 tg2c1=one_child[one_child['year']>1993]
170 tg2c1_empl=tg2c1['work'].sum()
171 tg2c1_mean=tg2c1_empl/len(tg2c1)
172
173 #calculate the average of the treatment group with two children pre-1994
174 tglc2=two_child[two_child['year']<1994]
175 tglc2_empl=tglc2['work'].sum()
176 tglc2_mean=tglc2_empl/len(tglc2)
177
178 #calculate the average of the treatment group with two child post-1994
179 tg2c2=two_child[two_child['year']>1993]
180 tg2c2_empl=tg2c2['work'].sum()
181 tg2c2_mean=tg2c2_empl/len(tg2c2)
182
183 #calculate diffs
184 dif3=tglc2_mean-tglc1_mean
185 dif4=tg2c2_mean-tglc2_mean
186 dif_dif3=dif3-dif2
187 dif_dif4=dif4-dif2
188
189 l3=["One Child", len(one_child), tglc1_mean, tg2c1_mean, dif3, '']
190 l4=["Control Group", len(nparent), cg1_mean, cg2_mean, dif2, dif_dif3]
191 l5=["Two Child", len(two_child), tglc2_mean, tg2c2_mean, dif4, '']
192 l6=["Control Group", len(nparent), cg1_mean, cg2_mean, dif2, dif_dif4]
193
194 table=[l1, l2, l3, l4, l5, l6]
195
196 headers=['Group', 'Sample Size', 'Pre-1993', 'Post-1993',
197          'Difference', 'Difference-in-differences']
198
199 table2=pd.DataFrame(table, columns=headers)
200
201 #table2
202

```

```

203 #1st regression
204 df['Post1993']=np.where(df['year']<1994,0,1)
205 df['interact']=df['Post1993']*df['parent']
206 X=df[['parent', 'Post1993', 'interact']]
207 y=df['work']
208 mod=sm.OLS(y, sm.add_constant(X))
209 res=mod.fit()
210 print (res.summary())
211
212 #second regression
213 X=df[['parent', 'Post1993','urate',
214       'nonwhite', 'age', 'ed', 'interact']]
215 y=df['work']
216 mod=sm.OLS(y, sm.add_constant(X))
217 res=mod.fit()
218 print (res.summary())
219
220 #placebo regression
221
222 df2=df[df['year']<1994]
223 df2['Post1992']=np.where(df2['year']<1992,0,1)
224 df2['interact']=df2['Post1992']*df2['parent']
225 X=df2[['parent', 'Post1992','urate',
226         'nonwhite', 'age', 'ed', 'interact']]
227 y=df2['work']
228 mod=sm.OLS(y, sm.add_constant(X))
229 res=mod.fit()
230 print (res.summary())
231 \end{lstlisting}

```

3.3 Stata Code

```

1 *Question 1 c
2 use "\\files\users\ecan\Desktop\ps1.dta", clear
3
4 *mean for education – years of education
5 sum ed if child>0
6 sum ed if child==0
7 *mean for nonwhite – if hispanic or black it is 1
8 sum nonwhite if child>0
9 sum nonwhite if child==0
10 *mean for age – age of the women
11 sum age if child>0
12 sum age if child==0
13 *mean for work – if worked last year
14 sum work if child>0
15 sum work if child==0
16 *mean for earn – annual earning
17 sum earn if child>0
18 sum earn if child==0
19
20 *Question 1 d
21 *mean for education – years of education
22 sum ed if child==1
23 sum ed if child>1
24 *mean for nonwhite – if hispanic or black it is 1
25 sum nonwhite if child==1
26 sum nonwhite if child>1
27 *mean for age – age of the women
28 sum age if child==1
29 sum age if child>1
30 *mean for work – if worked last year
31 sum work if child==1
32 sum work if child>1
33 *mean for earn – annual earnings
34 sum earn if child==1
35 sum earn if child>1
36

```

```

37 *Question 2
38 generate child=1 if children>0
39 replace child=0 if children==0
40 generate post1993=1 if year>1993
41 replace post1993=0 if year<1994
42
43 *Question 2 part a
44 **Obtaining the means for annual labor participation for women with children
45 sum work if child==1 & year==1991
46 sum work if child==1 & year==1992
47 sum work if child==1 & year==1993
48 sum work if child==1 & year==1994
49 sum work if child==1 & year==1995
50 sum work if child==1 & year==1996
51
52 collapse (mean) work, by(post1993 child)
53
54 **Obtaining the means for annual labor participation for women without children
55 sum work if child==0 & year==1991
56 sum work if child==0 & year==1992
57 sum work if child==0 & year==1993
58 sum work if child==0 & year==1994
59 sum work if child==0 & year==1995
60 sum work if child==0 & year==1996
61
62 *Question 2 d & e
63 **Obtaining the means for post 1993 and pre 1993 for control and treatment groups
64 sum work if child==1 & post1993==1
65 sum work if child==1 & post1993==0
66 sum work if child==0 & post1993==1
67 sum work if child==0 & post1993==0
68
69 *Question 2 f
70 sum work if children==1 & post1993==1
71 sum work if children==1 & post1993==0
72 sum work if children>1 & post1993==1
73 sum work if children>1 & post1993==0
74
75 *Question 2 g
76 generate interactionterm=post1993*child
77 reg work child post1993 interactionterm
78 reg work child post1993 interactionterm ed urate nonwhite age
79
80 *Question 2 h
81 use "\\files\users\ecan\Desktop\ps1.dta", clear
82 drop if year>1993
83 generate child=1 if children>0
84 replace child=0 if children==0
85 generate postplacebo=1 if year>1991
86 replace postplacebo=0 if year==1991
87 generate interactionterm2=postplacebo*child
88 reg work postplacebo child interactionterm2
89 reg work postplacebo child interactionterm2 urate ed nonwhite age

```