

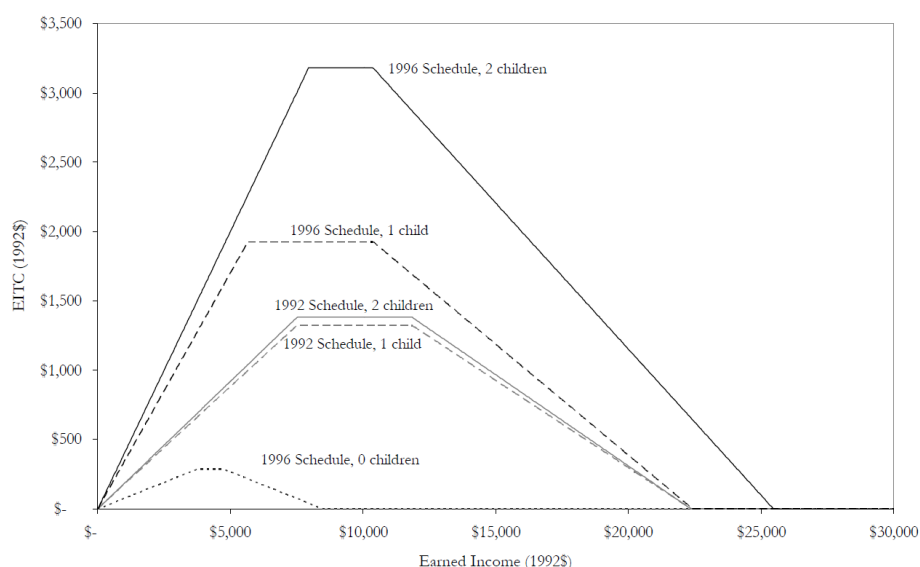
## Assignment

**Instructions** The assignment is due 12 noon on Monday 25 May 2020. Your work should consist of two computer files, uploaded to Wattle using the link provided:

- A Stata do-file which creates all of your results. This file must be annotated with explanatory comments, so that it is clear what results are sought, and it must run without syntax errors (assuming the data is in the current working directory).
- A typed report (Word or pdf format). Part of the assignment is to present results 'professionally'. This means that there should be no Stata commands or Stata output in the main text. Extract the information you need from the Stata output, and create nice tables and figures similar to those you see in textbooks and journal articles. Attach your Stata do- and log-files as appendices. There should be no mismatch between the do-file results and the reported results.

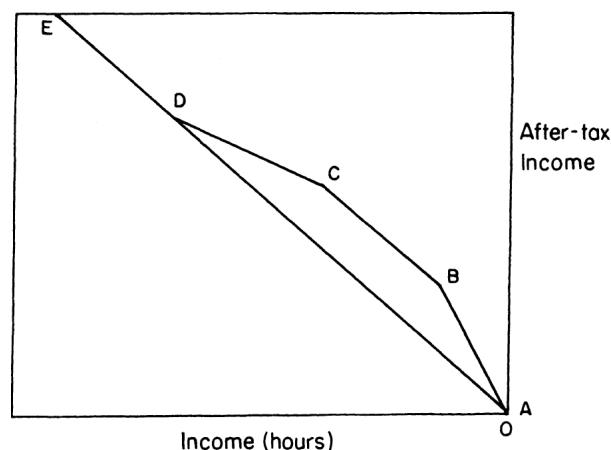
Your submission must be all your own original work. You must not collaborate with anyone. You may consult all of the EMET8005 course materials. If you have any questions about the assignment, please email [EMET8005@anu.edu.au](mailto:EMET8005@anu.edu.au). There is no penalty for clarification questions. If you are stuck or lost we may be able to provide a hint to unstick you, possibly with a small grade penalty.

**Introduction** The Earned Income Tax Credit (EITC) in the USA is a program to support low-income families with children. Most other welfare programs give money to poor people depending on some assessment of their needs. If recipients begin to earn more money, then their benefits are typically reduced, and this may discourage them from working. The EITC attempts to encourage work by giving money in proportion to people's own earnings. Only after their earnings reach a certain threshold are the benefits gradually reduced.



The figure above illustrates the program. The parameters are different in 1992 and 1996, but the principles are the same. There are four phases depending on people's earnings. For people with very low earnings, the tax office pays them a benefit amount proportional to their earnings until a maximum is reached (about 18% in 1992 and about 34% in 1996). For people in the next earnings range, the EITC amount is the same for everybody. In the third phase, the EITC is gradually reduced (at about the rate 13% in 1992 and 16% in 1996). Finally, the EITC is nil for people whose earnings are above a certain threshold. The rates and amounts are adjusted almost every year and vary with the number of children in the family, so the program is even more complicated than shown here.

The figure below illustrates how after-tax income is affected by the EITC. (The horizontal axis measures income or working hours increasing from right to left.) Without the EITC, after-tax income will increase proportionally with working hours, which corresponds to line ADE. The EITC changes the after-tax income to the line ABCDE. The line segment AB corresponds to very low earned income which is subsidized at a constant rate. The line BC corresponds to the income interval where the EITC amount is constant. The line CD corresponds to the earned income range where the EITC is phased out. The line DE corresponds to top earned income range which is unaffected by the EITC program.



Simple economic theory suggests that the EITC program encourages labour force participation for those who would not work otherwise, and doesn't discourage working for anyone. However, the effect on hours of work is ambiguous. Theorists think that people in the phase-in range could either increase or decrease their working hours, while those in the constant and the phase-out ranges would always reduce hours. (It may be helpful for you to think of after-tax income as 'consumption' and working hours as negative 'leisure'. Since consumption and leisure are both nice, 'utility' increases towards the upper right corner of the diagram.)

**Analysis and report** In this assignment you will use difference-in-differences methodology to study the impacts of changes in the EITC program on employment. As mentioned, the program parameters (the phase-in rate, the maximum credit, the income level where the credit begins to phase out, and the phase-out rate) vary across years. However, the changes were relatively minor during 1991–1993, while a big expansion occurred between 1993 and 1994 with further relatively minor expansions in the following years. The 1993/1994 expansion increased the generosity particularly for families with two or more children.

For the purposes of this assignment, assume that 1991–1993 are the pre-treatment years and that 1994–1996 are the post-treatment years. From 1994, there is also some support for very poor families without children, but the amount is small and we shall ignore this here. The vast

majority of eligible families consists of a single mother and her children. Hence, assume that the treatment group consists of single women with children and low income. A possible control group is single women without children. Since poverty is concentrated among people with not much education, we restrict the analysis to mothers with less than high school education.

Download the Stata data set `asEITC.dta` from Wattle. The original source of these data is various waves of the monthly Current Population Survey (CPS). The file has data for a sample of single women aged 20–54 with less than a high school education covering the years 1991–1996. All dollar amounts are in 1997 dollars.

There is a small confusing issue of timing that fortunately you don't really need to worry about. The variable `work` indicates whether the respondent was employed 'last year'. This is because the data are taken from the March CPS surveys where the respondents are interviewed about their employment and earnings in the last financial/tax year (January–December 'last year' when you look back from March). All the variables in this dataset refer to the last financial/tax year, so the timing should be consistent.

Begin by constructing dummy variables for the treatment group (call it *anykids*) and for the treatment period (call it *post93*).

- (a) Create a table of sample means and standard deviations for age, non-white race, years of education, whether working, family income, earnings, and unearned income over the years 1991–1993 for four groups (in separate columns): (1) single women without children; (2) single women with any children; (3) single women with one child; and (4) single women with two or more children. (Note another term for unearned income is non-labour income.)

Earnings are reported as zero for women who are not employed. Create a new variable with earnings conditional on working (ie missing for non-employed) and include summary statistics of this variable in the table as well.

Discuss the differences and similarities in the sample across groups.

- (b) A colleague suggests you can estimate the effect of the EITC expansion on employment by comparing single women with kids in 1994–1996 and single women without kids in 1994–1996. Describe how to do this in a regression framework (without additional control variables), carry out the regression, and present the results. Discuss the findings. Given the information in (a), how might this estimate be biased?
- (c) Create a graph which shows the average annual employment rates for each of the years 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.

*Hint:* There are various ways to go about creating a graph with averages by year. One way is to use the Stata command `collapse` which will replace the current data sets with another data set consisting of summary statistics. For example,

```
collapse (mean) work, by(year anykids)
```

will create a data set of the mean of `work` for each combination of `year` and `anykids`. The original data set will be erased from memory, and must be reloaded after the graphs is done. Check the Stata online help for details.

If you use `collapse`, remember to restore the original data when you are finished with the graph.

- (d) Given the level difference in average employment rates by group in (c) it may be hard to assess the results from the graphs. Instead create a graph showing the differences in the average employment rate between the treatment and control group across years. Comment on the graph.

*Hint:* Again, there are various ways to do this. One way begins with estimating a regression with a full set of time dummies as well as the time dummies interacted with *anykids*. The coefficients on the interaction terms capture the year-specific differences in average employment rates. Now find a way to plot these coefficients (with *year* on the horizontal axis).

- (e) Carry out a formal test of the hypothesis that the trends are parallel during 1991–1993. That is, test that the difference in employment rates between women with children and women without children is the same (not necessarily 0) across the years 1991, 1992, and 1993. Note that this amounts to two restrictions (like  $Dif_{1991} = Dif_{1992}$  and  $Dif_{1991} = Dif_{1993}$ ).
- (f) Calculate the unconditional (ie without any other controls) difference-in-difference estimates of the effect of the EITC expansion in 1993/1994 on employment of single women. Also calculate the standard errors. Take all women with children as the treatment group and all women with no children as the control group. Present the means and standard errors in a table as follows:

	1991–1993	1994–1996	Difference
Treatment group	?	?	?
	(?)	(?)	(?)
Control group	?	?	?
	(?)	(?)	(?)
Difference	?	?	?
	(?)	(?)	(?)

Discuss your results. What is the estimated treatment effect?

- (g) Recalculate the unconditional difference in difference estimates by allowing the treatment effects to vary for those with 1 and 2 or more children. Present your results in two tables, like the one in (f). 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? Discuss the practical implications of the results.
- (h) Now run a regression to calculate the difference-in-differences estimate of the effect of the EITC. Use all women with children as the treatment group. Do not include any variables other than the few needed to calculate that effect. How do these results compare with what you found in (f)? What is the interpretation of the coefficient on the variables you included?
- (i) Rerun the regression in (h) with dummies for each of the years 1991–1996. (Except some should be dropped to avoid the dummy variable trap. It doesn't matter which you drop.) Discuss why one might want to include these dummies. Does the estimate of the treatment effect change much?

- (j) Create a large table to present estimation results for five different models in the columns. Report the coefficient estimates and their standard errors for all variable in the model, except the time dummies, state dummies, and constant term. For the latter dummies, it suffices to indicate whether they are included in the model or not. Three digits after the decimal point should be sufficient.

For each model, discuss how the estimated treatment effect changes when additional controls are added or changed.

- Model 1: The first column of your table should show the results from the basic model in (i) with year dummies but no additional controls. You do not need to report the year dummies in the table, if space is limited.
- Model 2: Add controls to Model 1 for 'demographic' variables: unearned income, number of children, non-white race, age, age squared, years of education, and years of education squared. Scale unearned income, age squared, and education squared such that the standard error is larger than 0.001. For example, perhaps redefine *unearn* to be \$1,000s of 1997 dollars. In your comments, also interpret the parameter estimates on non-white and income variables.
- Model 3: Extend Model 2 by adding the state unemployment rate and allow its effects to vary by the presence of children. (You will need interact *urate* and *anykids*.) In your comments, also discuss what the estimates tell you about the importance of business cycles and how they affect the different groups?
- Model 4: Add state-specific dummies to Model 3. (Since we are usually not interested in the coefficients on the state dummies themselves, it is customary not to report them in the tables. Instead, we add a row saying that state dummies were included in the estimates presented in this column.) In your comments, also discuss why you think including state-specific intercepts would or would not change the estimated EITC effect.
- Model 5: Extend Model 4 by allowing the treatment effect to vary by those with one or with two or more children. In your comments, explain what you would expect to find given the nature of the EITC expansion and discuss whether your findings and your expectations agree.

*Hint:* You will need to create some of the control variables (eg age squared).

- (k) One way to probe whether the difference-in-differences methodology is delivering plausible estimates of the treatment effect is to do so-called 'placebo' experiments. That is, we apply the DD methodology to a period where we think there was no change in policy. If we end up finding a significant 'treatment' effect nevertheless, then we have a clear indication that something (the common trends assumption) is wrong.

For this, use the same treatment and control groups as before, but take data from only the pre-treatment period where we think the EITC parameters didn't change much. Now pretend that there was a policy change between 1991 and 1992, and define 1991 as the pre-treatment period and 1992–1993 as the post-treatment period. Estimate a (fake) treatment effect without additional controls, as in (h). Discuss the findings.