***Economics 142***
***Problem Set 6***

The data set welfare.csv contains 5,480 observations for people from the Self Sufficiency Program, a welfare experiment operated in Canada that created incentives for full time work. The experiment had a very strong incentive for people in the treatment group to find a full time job and leave welfare within a year of entering the experiment (when people were randomly assigned to a treatment or control group). The data set has the following variables:

treatment = 1 if in treatment group, 0 for the control group

imm = 1 if immigrant

hsgrad = 1 if high school graduate (or higher) education

agelt25 =1 if age is 25 or less

age35p = 1 if age is 35 or more

working_at_baseline = 1 if person was working when assigned to treatment/control group

anykidsu6 = 1 if any children under age 6

nevermarried = 1 if never married (all cases are single parents)

In addition there are measures of two outcomes at various times after assignment to the program:

- ft15, ft20, ft24, ft48 = set of dummies if working full time at months 15, 20, 24, and 48 after random assignment

- welfare15...welfare48 = set of dummies if receiving welfare at months 15, 20, 24, and 48 after random assignment

1. We are interested in a causal model of how working full time affects welfare participation. Our causal model for period $t$ is:

$$y_i = \beta_0 + \beta_1 FT_i + u_i, \tag{1}$$

where $y_i$ is an indicator equal to 1 if parent $i$ is on welfare in period $t$, and $FT_i$ is an indicator for full time status in the month. We are going to use "assigned to the treatment group" as our instrument, so $z_i = T_i \equiv Treatment_i$. A slight difference from the examples in class is that we are going to allow for a different set of coefficients $\beta_0$ and $\beta_1$ in each of months 15, 20, 24, and 48 after the baseline period when people were assigned their treatment status.

a) Estimate first stage models for the probability of working FT in months 15, 20, 24, 48, using treatment as the instrumental variable and no other control variables:

$$FT_i = \pi_0 + \pi_1 T_i + \eta_i$$

There will be separate estimates of $(\pi_0, \pi_1)$ for each of months 15, 20, 24, 48.

b) Estimate reduced form models for the probability of being on welfare in months 15, 20, 24, 48, using treatment as the instrumental variable and no other control variables:

$$y_i = \gamma_0 + \gamma_1 T_i + v_i$$

There will be separate estimates of $(\gamma_0, \gamma_1)$ for each of months 15, 20, 24, 48.

c) Estimate the "causal model (1) by OLS for each of months 15, 20, 24, 48.

d) Estimate the causal model by IV in each of months 15, 20, 24, 48.

e) Verify that in each month, $\widehat{\beta}_1^{IV} = \widehat{\gamma}_1/\widehat{\pi}_1$.

2. Now repeat steps (a)-(e) from question (1) ***for month 15 only***, but including as controls the variables {imm, hsgrad, agelt25, age35p, working_at_baseline, anykidsu6, nevermarried} in your first stage model, your reduced form model, and your OLS and IV versions of the causal model. Because SSP was a randomized experiment, $T_i$ is (approximately) uncorrelated with all these control variables. Does their addition affect your different estimates?