ECON 104 **Truth in Numbers**

Homework 3

Please estimate the effects of the MLDA on alcohol consumption. The figures and tables you create will form the core of the second paper. The dataset **"NHIS.csv"** is a sample drawn from the National Health Interview Sample Adult Files 1997-2007. It includes the following variables: *HS_diploma, Hispanic, white, black, uninsured, employed, married, workin_lw, going_school, male,days_21* (Days to 21st birthday), *drinks_alcohol* (Reports they drink alcohol),*AGE_yrs*, *perc_days_drink* (Percent of days on which they report drinking). Each row on the dataset is a record for an individual.

1. Determine what range of age to include in your age profile of whether or not people drink alcohol (variable name: *drinks_alcohol*). Please attach a panel with various choices of range. Why did you pick the one you did? Please include your code. (Hint: You will need to bin the data into one week bins for this to work as the outcome is binary).

2. Determine what bin width to use for your age profile of whether or not people drink alcohol (variable name: *drinks_alcohol*). Please attach a panel with various choices of bin width. Why did you pick the one you did? Please include your code.

3. Determine what range of alcohol consumption levels to use in your age profile of whether or not people drink alcohol (variable name: *drinks_alcohol*). Please attach a panel with various choices of the range of *drinks_alcohol*. Why did you pick the one you did? Please include your code.

4. Create a table of regression estimates of the increase in the proportion of people who drink with different order polynomials going from least to most flexible. You should also create a figure with the regressions superimposed over the scatter plot to help you assess which model best fits the age profile. Which regression specification do you prefer and why?

5. Create a carefully labeled figure of the first stage relationship between the MLDA and whether or not people drink alcohol (variable name: *drinks_alcohol*) with the range and bin width choices you made in steps 1-3. Also please superimpose the fitted values of your preferred regression specification from step 4. Please include your code.

6. Does the MLDA reduce the proportion of people who drink? If so by how much.