Fall 2021
SS 340
Loujaina Abdelwahed

Submit your work at the beginning of class on Monday (October 18). You may work with other students, but the write-ups must be unique. Please save your files as "HW2.doc" and upload them to MS Teams.

Part I: Answer the following questions

1. You are hired to conduct a study to determine whether smaller class sizes lead to improved student performance on fourth graders. Suppose you can collect data on several thousand fourth graders in a given state.
   a) To run a simple regression model, what are the variables you need to know for each student? **(5 points)**

   b) Write down the simple regression model. What do you expect the relationship between class size and test score to be? What is the interpretation of $\hat{\beta}_1$? **(5 points)**

   c) Do you think this simple regression will give you the causal relationship between class size and test scores? Why/ why not? **(5 points)**

   d) What other variables would you control for in this regression? Think carefully about whether these are good, useless or bad controls. **(5 points)**

2.

a) First, consider a regression where the independent variable is the neighborhood income around a school attendance zone and the dependent variable is student test scores. What is the likely sign of the coefficient on neighborhood income? **(5 points)**

b) Now consider a regression where the independent variable is a measure of violent crime incidents around the school and the dependent variable is student test scores. What is the likely sign of the coefficient on violent crime? **(5 points)**

c) Finally, consider a regression of violent crime incidents on area income levels. What is the likely sign of the coefficient on area income levels? **(5 points)**

d) Now consider the sign of omitted variable bias in the first regression, neighborhood income levels on student test scores. What is the sign of omitted variable bias if we omit a measure of violent crime around a school? **(5 points)**

Fall 2021
SS 340
Loujaina Abdelwahed

Part II: Data Analysis

Attached are the public use micro datafiles from the latest Consumer Expenditure Survey (CE). The CE survey is one of the most important government surveys that provides data on expenditures, income, and demographic characteristics of consumers in the United States, and is used to calculate weights for the Consumer Price Index.

CE data are collected by the Census Bureau for BLS in two surveys, the Interview Survey for major and/or recurring items and the Diary Survey for more minor or frequently purchased items. The data attached to the HW are from the Interview Survey.

Data are collected on a quarterly basis. In each quarter of the year, the BLS asks a sample of HH about their consumption of different items and their income in the PREVIOUS quarter. Note that this is very important information. So, files for year 2019Q1 asks households about their consumption in the last quarter of year 2018, 2019Q2 asks households about their consumption in first quarter of 2019, … etc. This means that to get data on consumption in the last quarter of 2019, you have to use interviews from the first quarter of year 2020.

I am providing for you the data from the five quarters.

Note that, each quarter, a different set of households are being interviewed.

The CE survey has many types of files. I am providing to you only the fmli files, which are the summary files. I did not clean the data at all… the files have ALL the variables. What I have done for you is I have code to the codebook, and I wrote for you the definitions of the variables that you would need for the analysis. So stick to them:

- NEWID: HH identification number
- FAM_SIZE: family size
- NUM_AUTO: number of vehicles owned
- FINCBTXM: final income before taxes (imputed in case of missing information) in the past 12 months
- TOTEXPPQ: total expenditure in the previous quarter
- TOTEXPCQ: total expenditure in the current quarter
- ALCBEVCQ/ ALCBEVPQ: spending on alcoholic beverages in current and previous quarter
- FOODCQ/ FOODPQ: spending on food in current quarter and in previous quarter
- FDHOMECQ/ FDHOMEPQ: food at home this quarter and previous quarter
- FDMAPCQ/ FDMAPPQ: meals as pay this quarter and previous quarter
- FDAWAYCQ/ FDAWAYPQ: food away from home excluding meals as pay this quarter and previous quarter
- MAJAPPCQ/ MAJAPPPQ: spending on major appliances in current quarter and previous quarter
- TENTRMNC/ TENTRMNP: spending on entertainment (sporting events, movies, and recreational vehicles) in current and previous quarters

Fall 2021
SS 340
Loujaina Abdelwahed

- EDUCACQ/ EDUCAPQ: spending on education in this quarter and previous quarter
- ELCTRCCQ/ ELCTRCPQ: spending on electricity this quarter and in previous quarter

The purpose of this question is to use real data on consumer expenditure and income to compute the "marginal propensity to consume (mpc)" (those of you who took macro should know what is this. If you have not taken macro, you can google this term and learn it).

For this question, we will focus on data from the 2019 Interview Year (not the 2019 calendar year). I want to focus on interview year because it is more straight forward. If you want to think about calendar years in CE, it gets really tricky because the month of interview will also matter for whether the household is counted or not.

So let's focus on Interview Year 2019.

a. Which files you need to use to construct your sample? (there have to be 4 files)
   Append the files together to construct the dataset (**10 points)**
b. Present some summary statistics for the main income and expenditure variables. Do not forget to write a paragraph or two that tell story about the data from your summary statistics. **(5 points)**
   P.S. For each expenditure variable (e.g. total expenditure), you have 2 variables, one ends with CQ and one ends with PQ (stand for current quarter and previous quarter). If you are being interviewed in the first month of the quarter, CQ variables will be 0. So you gotta add the CQ and PQ variables. This gives you quarterly expenditure. To annualize it, you multiply it by 4.
   You do not have to do this for income, because income is already annual

c. Make sure to "clean" the data. You can drop any observations that does not seem right (let us know what you dropped and justify why) **(10 points)**
d. Write out the regression equation you plan to estimate to figure out the MPC **(5 points)**
e. Estimate the regression and interpret the coefficients (interpret both $\hat{\beta}_1$ *and* $\hat{\beta}_0$) . **(10 points)**
f. Do you think that MPC can differ for high and low income households? **(10 points)**
   P.S. you can define high-income as households with income above the median and low-income as households with income below the median.
g. In the data, you have spending on some categories: food at home, food away from home, alcoholic beverages, major appliances, and entertainment. Estimate the MPC for an essential category and for a non-essential category. Are the MPCs for each of these two categories equal? **(10 points)**

For questions e, f, and g, always interpret your coefficients. Do not formally test if the coefficients are equal because in this case, you cannot use the t-test anymore.

Fall 2021
SS 340
Loujaina Abdelwahed