**HW 6**

**First Review Video Tutorial 10, once you are finished complete the following assignment. Please work with a partner, but submit your own code and your own answer sheet.**

Load the data set in the homework folder titled: **DATA\_for\_HW6.dta**. Answer the following questions. Your answers should be typed using the answer sheet provided and handed in along with a .do file attached to the very end of the answer sheet. Similar to our class tutorials, this data set includes repeat transactions for residential single family homes. Each home, which is uniquely identified by the variable **parcel\_no**, may appear multiple times if said home transacted…multiple times. The price of each home sale is given by the variable **price**. The transaction date of each home is given by the string variable **tdate.** The string variable **floodplain\_sfha** is equal to “T” if the home is located in flood risk area and “F” or missing otherwise. In this first part of this project, we are going to estimate the impact the Boulder County flood had on the bid price of homes in flood risk areas, before and after the flood which commenced on 9/11/2013. Note that the variable **dist\_flood\_ft** measures the distance between each home and the boulder county flood.

This data consists of **158,198** observations. As a first step, normalize the data by (a) keeping home sales that transacted between [-365 days, 365 days] of 9/11/2013 (you need to convert the variable tdate to do this); (b) keeping homes that have a floodplain\_sfha value of JUST “T” or “F” (e.g. drop the missing entries). Now, verify that you now have EXACTLY **7,876** observations in the data. If you don’t have 7,876 observations, **you made a coding error**. Finally generate a variable named **SFHA** that is equal to 1 if floodplain\_sfha == "T" and zero otherwise. (SFHA stands for special flood hazard risk area).

1. What is the average price of homes located in the SFHA (i.e. those for which SFHA==1) restricting attention to homes that sold before the flood (i.e. number of days elapsed <0)? Please note that the variable indicating the sale price of a home is called “**price**”. (Note, your answer should = **$440,640.5)**
2. What is the average price of homes located in the SFHA (i.e. those for which SFHA==1) restricting attention to homes that sold after the flood (i.e. number of days elapsed >=0)?
3. Based on your answers to (1) and (2), did home prices in the SFHA increase following the flood, or decrease?
4. What is the average price of homes located OUT of the SFHA (i.e. those for which SFHA==0) restricting attention to homes that sold before the flood (i.e. number of days elapsed <0)?
5. What is the average price of homes located OUT of the SFHA (i.e. those for which SFHA==0) restricting attention to homes that sold after the flood (i.e. number of days elapsed >=0)?
6. Based on your answers in (4) and (5), did home prices OUT of the SFHA following the flood, or decrease?
7. Based on all of your answers above, does the evidence thus far suggest the flood caused an increase or a decrease in the price of housing in the SFHA. Justify your answer.
8. How would your answer to 7 change if market interest rates for houses in Boulder County were overall higher after the flood relative to before the flood.
9. Go back to question 3, by exactly what amount did home prices increase or decrease in the SFHA following the flood?
10. Go back to question 6, by exactly what amount did home prices increase or decrease out of the SFHA following the flood?
11. Subtract the number you reported in question 10 from the number you reported in question 9 and report this difference-in-differences on the answer sheet.
12. Next, create two new variables. First, create the binary variable “Post” that is equal to 1 if number of days elapsed >=0 and zero otherwise. Create the binary variable SFHA\_Post which is equal to the variable SFHA multiplied by the variable Post. Now, estimate the following regression: , where refers to the variable in STATA called SFHA, and so on. Report the results in a table and comment specifically on the magnitude of your estimate for relative to the number you report in question 11 (hint: they should be equal!).

*Next, we will explore the relationship between home prices, home development, and the elevation of homes. Please clear your memory (in STATA that is!) and reload the original dataset* **DATA\_for\_HW6.dta.** *In the late 1800s, the Federal Government allocated 1,800 acres of land near Boulder County, Colorado. With concerns of urban populations expanding into wild and rural lands, voters passed what is now referred to as the “Blue Line”. The Blue line refers to an elevation contour. More specifically, city voters voted to restrict city water services to any homeowner located in a home with an elevation of 1750 meters or more with the goal of limiting expansion of development in the mountainous regions of the city. (Note that the variable* ***elevation*** *in our data measures the elevation of each home in the data,* ***in meters)****. In other words, homes with an elevation of 1751 meters were restricted access while homes with an elevation of 1749 were given access to city water services.*

**Question 13:**

First, estimate a series of multiple regression models to estimate the impact of elevation on home values. Specifically, you will estimate **three** variations of the following model and then report your results in a regression table as shown in class using the outreg2 command. The model of interest takes the form:

where represents a vector of additional control variables that will vary depend on which column of the table we consider. In **Column (1)** of your regression table, estimate EQ 1 **without any control** **variables other than elevation and elevation squared.** (Note, this regression should include **158,198** observations). In **column (2)**, add quadratic controls for the year of each home sale (e.g. include the variable year\_of\_sale and year\_of\_sale\_squared which you will have to compute yourself). Also include quadratic controls for the age of each home at time of sale (e.g. age of home and age of home squared where age equals the sale year minus the built year of each home). (Hint, you will need the variable **builtyear** to answer this question as well as the variable **tdate**). Lastly, In **column (3)**, in addition to the controls that you included in column (2), further add quadratic controls for main floor square footage using the variable **mainfloorsf**. Once finished, insert your completed table in the word document write up and title your table: “Table (Q13)”

**Question 14:**

We are going to provide some visual evidence of Boulder County’s “Blue Line” policy on property values by fitting non-parametric plots in STATA using the lpolyci command (if unfamiliar, please reference our class tutorials). Specifically, on the **same graph**, provide two lpoly trend lines. In the first trend line, plot price against elevation restricting attention to homes located at an elevation within the interval: [1250m, 1750m). In the second trend line, plot price against elevation restricting attention to homes located between [1750m, 2250m]. Your command in STATA should look something like

twoway (lpoly price elevation if … ) (lpoly price elevation if …)

Label the y-axis “Price”. Label the x-axis “Elevation”. Note, you must write STATA code within the twoway plot type to produce these titles. Likewise, also write code to add a vertical line at 1750m (hint: xline(1750). Finally, write code that will add a legend to your STATA plot which separately indicates the lpoly trend line for homes located in each respective interval. In the space provided in the answer sheet, please copy and paste your STATA graph.

**Question 15)**: Does your graphical evidence suggest that the blue line lead to price decreases or increases for homes located above 1750m? Explain.

**Question 16)**: What is the average difference in the price of homes in the interval [1750m, 2250m], and [1250m, 1750m). Is this difference statistically meaningful? (Note, you will have to conduct a test for the difference of means to answer this part (hint, take a look at question 17 before conducting this test).

**Question 17)**: Next, restrict attention to homes in the interval [1250m, 1750m]. Then, generate two variables. First, generate the variable called **treat\_1750\_2250** which is equal to 1 if **elevation**>=1750 & **elevation**<=2250. Likewise **treat\_1750\_2250** = 0 if **elevation**>=1250 & **elevation**<1750. Once finished, run the regression

Report your estimate of and also comment on its magnitude relative to the difference you reported in question 16 (hint, it should be the same). Please report your results in an excel table.

**Question 18)**: Generate a new variable called elev\_m\_1750 which is equal to elevation – 1750 so that if elev\_m\_1750>0 then elevation > 1750 and if elev\_m\_1750 <0 elevation < 1750 and if elev\_m\_1750 = 0 elevation = 1750. Building off of everything you did in question 17 (including the observations dropped from the sample), also generate another new variable called **treat\_x\_elev\_m\_1750** which is equal to the product of the variables **elev\_m\_1750** and **treat\_1750\_2250**. Once finished, run the regression

After reporting these results in an excel table, comment how the inclusion of these additional covariates changed the estimate for . (Also note, you just estimated a regression discontinuity model. Along these lines, think very carefully about interpreting . represents the effect of turning off and on (e.g. going from 0 to 1) holding and which is equivalent to fixing elevation at 1750 which conceptually represents the estimated difference in the prices of homes immediately above and immediately below the 1750 discontinuity).

**Question 19**): Explain very clearly how the formal estimates you provide in 17 and 18 are reflected/reconciled/explained by the visual evidence you provide in question 14. Also, suppose that a researcher argued that model estimates of from question 17 are biased because (a) temperature influences home values, is omitted from the model, and is correlated with since homes at higher elevations face lower temperatures. Do you agree with this critique? Why or why not. If you do agree with this researchers critique, is this critique relevant for model estimates of obtained from question 18? Why or why not.