Assignment 2

**221.644.01 Econometric Methods for Impact Evaluation of Health Programs**

**Prof. Antonio J. Trujillo, Ph.D.**

Assistant Professor

Health Systems Programs

Department of International Health

Johns Hopkins Bloomberg School of Public Health

## **PART I (20 points):** Briefly answer the following four questions. Each answer should be limited to one paragraph shorter than 10 lines.

1. Please describe the main elements behind regression discontinuity design. How does it measure local average treatment effect?
2. Briefly discuss the relevance of the main assumptions behind RDD.
3. What are the main consequences in terms of estimating the program impact using RDD if one variable that is influenced by program participation is omitted?
4. Briefly explain how one may enhance the validity of RDD.

**Part II (80 points): Two numerical exercises**

**Part II.A.** [40 points]Please answer the following numerical exercises related to the link between insurance (Medicare) and use of medical services (mammograms). In order to complete this assignment, students will need to use the following database: Medicarecleandata.dta. The Word files codebook\_assignment2 and varlayout\_assignment2 contain the name and description of the variables to be used in the analysis. Each question will be worth 5 points.

Using RDD, you will estimate the impact of having Medicare for those with less than a high school degree. You will restrict your analysis for women in the 50 to 80 age range.

These exercises are based on Decker, S.L. 2005. It is important to keep in mind that you will not be able to replicate her results because I only provide three years of data instead of the 11 years of data used in her analysis. Furthermore, we will not control for any survey design characteristic of the database. Therefore, one could use the normal regression commands in STATA.

The variables from the database to be used in the analysis are: STATE, ORACE; HISPANIC; GENHLTH; MEDICAR2; TYPCOVR1; TYPCOVR2; NOCOV12, SMOKE100, AGE, MARITAL, EDUCA, INCOME2, WEIGHT, HEIGHT, HADMAM, HOWLONG, SEX, YEAR. You should construct your variables as Decker did in her paper. Lastly, in the analysis you should create a dummy for missing value in each covariate of interest. Observation with missing values for the dependent variable should be deleted.

**Part II.A.1: Impact of Medicare on mammograms use for those with less than a high school degree**:

1. Please show graphically if the use of mammograms for those with less than a high school degree has a discontinuity at 65. You should use regular plot that shows the mean value of the outcome of interest at each age. You should show in your graph a straight line at the cut-off point. Briefly discuss your results
2. Please create a table that shows for the following 6 age groups: (65-10); (65-5); (65-2); (65+2);(65+5); and (65+10), the mean and standard error of the following variables: GENHLTH; MEDICAR2; SMOKE100, AGE, MARITAL, EDUCA, INCOME2, WEIGHT, HEIGHT, HADMAM, HOWLONG, SEX . Briefly discuss the implication of your results.
3. Now, first, please run the simple model (case 1 in your lecture notes), of use of service as a function of program participation and age. Second, add the variable dummy education to this model. In both models you should control for the dummy variable, race, state, year effect. Please briefly interpret your results.
4. Now, please run an interaction model (case 2 in your lecture notes), of use of service as a function of program participation and age. Second, add the variable dummy education to this model (be sure to add the interaction program dummy only to the age variable). In both models you should control for the dummy variable race, state, year effect. Please briefly interpret your results.
5. Now, please run a non-linear model quadratic in age (case 3 in your note), of use of service as a function of program participation and age. Second, add the variable dummy education to this model (be sure to add the interaction program dummy only to the age variable). In both models you should control for the dummy variable race, state, year. Please briefly interpret your results.
6. Now, please run a non-linear model cubic in age (case 3 in your note), of use of service as a function of program participation and age. Second, add the variable dummy education to this model (be sure to add the interaction program dummy only to the variable age). In both models you should control for the dummy variable race, state, year effect. Please briefly interpret your results.
7. Based on your results 1-7 (please be sure to justify which model specification is more adequate given the data), briefly discuss the impact of Medicare on mammograms use among those with less than a high school degree.
8. Briefly discuss some limitations of your analysis.

**Part II.B. [40 points] Impact of Public Policy on the Health Insurance Coverage of Young Adults**

Please answer the following numerical exercises related to the link between age and insurance coverage. In order to complete this assignment, students will need to use the following database: levinedata2.dta. The Word file varlayout\_assignment2 contains a description of the data.

1. [7 pts] Please show graphically if the percent insured has a discontinuity at age 19 when comparing the time periods 1990-1995, 1999-2003, and 2004-2008. Also show graphically if the percent covered by public health insurance has a discontinuity at age 19 when comparing the same time periods. Finally, show if the percent covered by public health insurance has a discontinuity at age 19 for these time periods. What do your results tell you? Discuss the possibility of crowd out as well as take up.
2. [7 pts] Run the following regressions:

* Dependent variable- any insurance coverage; Independent variable- individual is eligible for SCHIP, state, year, age, unemployment rate of the individual’s state, household income as a share of the poverty line, and the square of the household income as a share of the poverty line. Limit your sample to individuals between the ages of 16 & 22. (Hint- you will need to create a dummy variable for each state fips code, year and age. The following webpage might be helpful: http://www.ats.ucla.edu/stat/stata/faq/dummy.htm)
* Dependent variable- public insurance coverage; Independent variable- individual is eligible for SCHIP, state, year, age, unemployment rate of the individual’s state, household income as a share of the poverty line, and the square of the household income as a share of the poverty line. Limit your sample to individuals between the ages of 16 & 22. (Hint- you will need to create a dummy variable for each state fips code, year and age. The following webpage might be helpful: http://www.ats.ucla.edu/stat/stata/faq/dummy.htm)
* Dependent variable- private insurance coverage; Independent variable- individual is eligible for SCHIP, state, year, age, unemployment rate of the individual’s state, household income as a share of the poverty line, and the square of the household income as a share of the poverty line. Limit your sample to individuals between the ages of 16 & 22. (Hint- you will need to create a dummy variable for each state fips code, year and age. The following webpage might be helpful: http://www.ats.ucla.edu/stat/stata/faq/dummy.htm)
* NOTE: Your answers will be slightly different than what Levine et al. report, because we will not cluster standard errors around age.

You only need to report the coefficient on being eligible for SCHIP and any others that you find interesting. Discuss your results. Make sure to note if the coefficient is statistically significant or not.

1. [7 pts] Repeat question 2 using a sample that is restricted to income being below 150% of the poverty line. (The sample should still be restricted to individuals between ages 16 and 22.)
2. [7 pts] Repeat question 2 using a sample that is restricted to income being above 150% of the poverty line but below 300%. (The sample should still be restricted to individuals between ages 16 and 22.)
3. [7 pts] Repeat question 2 using a sample that is restricted to income above 300% of the poverty line. (The sample should still be restricted to individuals between ages 16 and 22.)
4. [5 pts] Compare your results from questions 2-5. What do they tell you about crowding out and take up?