**Impact Evaluation in Practice**

**Applied Exercise**

Instructions: This problem set reviews the empirical application of a core set of impact evaluation methods. While the data are fictional, they are modeled after real-world examples of impact evaluation.

Context: This exercise analyzes the impact of a health insurance subsidy program for the poor, the HISP. One of the primary objectives of the program is to reduce the burden of health-related expenses for low-income households. You have been tasked with conducting an impact evaluation of the effect of the HISP on out-of-pocket health expenditures. You will use the dataset named “evaluation.dta” to conduct your analysis. The evaluation design will change with each case. A description of the data is provided in the appendix.

To begin:

1. Set up a “do file” to record your analysis and call it “solution.do”.

**Impact Evaluation Case 1: Before and After**

Under this scenario, you will estimate the effect of the program by comparing the change in outcomes over time for a group of households that enrolled in the program. Assume full compliance, meaning that all of the households eligible for the program enrolled in it.

1. Compare the average health expenditures before (round = 0) and after the program (round = 1) for the group eligible households (eligible =1) in treatment communities (treatment\_locality = 1)

**Impact Evaluation Case 2: Self Selected Treatments**

We will call this the random offering scenario. Under this scenario, enrollment in the health insurance program is voluntary and all households in the pilot communities are eligible to enroll. You have at your disposal a baseline survey (round = 0) which was collected before the health insurance program rolled out, and one follow up survey (round = 1) collected 12 months following the start of the program. For this case, analyze only communities where the program was offered (keep if treatment\_locality = 1). When running regressions make sure to cluster standardized errors by locality.

1. Describe the number of observations, the average, the standard deviation and the confidence intervals of the variables in the baseline database.
2. For the baseline, compare the average monthly per capita out of pocket health expenditures (health\_expenditures) and other covariates between households that sign up for the program (enrolled\_ro = 1) and those that didn’t sign up. Are the two groups comparable?
3. Your friend (whose doesn’t know a lot about impact evaluation) suggests that you use regression analysis with a method called Ordinary Least Squares (OLS) to estimate the effect of the program on the beneficiaries. Estimate the correlation between program participation and out of pocket health expenditures in the treatment period (round = 1). Run two regressions:
   * 1. Without controls
     2. Including characteristics of the household head and spouse
4. What is the “effect” of the program on the participants out of pocket expenditures?
5. Do you think the estimated coefficient form regression i or ii is the true program impact? What type of biases might you be worried about?

**Impact Evaluation Case 3: Randomization**

It turns out that when communities were being selected for inclusion in the health insurance pilot, there were many more eligible communities than could be covered with the available budget. The provincial authorities decided to run a lottery to select the communities that would participate in the insurance scheme in year 1, thus giving all communities a fair chance to start in the program first. Your data contains information on communities selected at random for participation in year 1, as well as on communities that would only enter the program in subsequent years. The variable “treatment\_locality” indicates treatment communities (treatment\_locality = 1) and non-treatment or control communities (treatment\_locality = 0). For this case, use both treatment and control communities in your analysis. The sample is structured as follows:

|  |  |  |
| --- | --- | --- |
|  | Treatment Communities | Control Communities |
| Eligible | 5,929 | 5,328 |
| Ineligible | 3,990 | 4,580 |

1. Compare baseline out of pocket health expenditures and other covariates between eligible households in treatment and control communities. Is the sample balanced on observables? Is this what you would expect and why or why not?
2. In the treatment period (round =1), compare the average out of pocket health expenditures for the eligible population in treatment and control communities. Is this the impact of the health insurance program on out-of-pocket health expenditures?
3. Now use an OLS regression to estimate of the effect of the program on out-of-pocket health expenditures:
   * 1. Without controls
     2. Including characteristics of the household head and spouse
     3. Including baseline covariates
4. What is the impact of the program on out-of-pocket health expenditures? What is the percent decrease that can be attributed to the program?
5. Which are some of the potential biases that you may still be concerned with under this method?
6. Replicate the analysis (step 3) on the ineligible population as a validity test. Do these results help confirm your confidence in the validity of the program impacts?

**Impact Evaluation Case 4: Instrumental Variables**

In case 3 you estimated the impact of the program by comparing eligible populations in treatment and control populations. However, you know that not ALL eligible households in treatment communities decided to enroll in the program. You would like to know what the effect of the program was for those eligible households that actually signed up for the program and were covered with the health insurance. To do this, you will need to correct for endogenous selection into treatment using a valid instrument.

For this case, we will alter the experimental design. Assume that the ministry of health makes an executive decision that the HISP should be made available to any household that wants to enroll. However, you know that realistically, this national scale-up will be incremental over time, and so you reach an agreement to accelerate enrollment in a random subset of villages through a promotion campaign. You undertake an intensive promotion effort in a random subsample of villages, including communication and social marketing campaigns aimed at increasing awareness of the HISP. The variable “promotion\_locality” indicates promotion communities (promotion\_locality = 1) and non-promotion localities (promotion\_locality = 0). The variable enrolled\_rp indicates whether or not a household enrolled in the program in this random promotion scenario. Since the promotion treatment was randomized, you hypothesize that it is a valid instrument to correct the selection into treatment when comparing those who enrolled in the program with the control communities.

1. Estimate a “two stage least squares” (ivreg) regression of the effect of the program on out-of-pocket health expenditures for treated households, using promotion as your instrumental variable.
   * 1. Without controls
     2. Including characteristics of the household head and spouse
     3. Including baseline covariates
2. How would you interpret this new estimated coefficient?

**Impact Evaluation Case 5: Regression Discontinuity Design (RDD)**

After doing some more investigation on the design of the health insurance program, you find out that authorities decided to target the program to low-income households, so only households below the existing poverty line would be eligible to enroll. The poverty index was constructed through a proxy means test that assigned each household in the community a score between 0 and 100. All households with a score less than or equal to 58 were classified as poor and were thus eligible to enroll in the health insurance program. All households with a score above 58 were considered non-poor and remained ineligible to enroll. The variable “eligible” identifies households below the poverty line (eligible = 1) and ineligible households (eligible = 0).

1. First, normalize the poverty score so that the threshold is 0. Estimate an OLS regression of the poverty index on out of pocket health expenditures in the treatment period. Include both the left and right sides of the poverty threshold in the regression.
2. Plot the fitted value of out-of-pocket health expenditures on the poverty index score (note: you can generate the fitted value with the “predict” command and plot it just using “graph7”).
3. Estimate the effect of the program on out-of-pocket health expenditures using an RDD model. In addition to your treatment/control variable, make sure to control for program eligibility by including both the left and right sides of the poverty threshold in your regression.
4. Introduce in your regression model squared, cubic, and quartic poverty score terms.
5. Plot the fitted value of out-of-pocket health expenditures estimated in step 4 against the poverty score. Is there any evidence of non-linearity?
6. Perform a “false experiment” using the RD model in the baseline period.
7. Plot the fitted value of out-of-pocket health expenditures estimated in step 6 against the poverty score.

**Impact Evaluation Case 6: Differences in Differences (dif in dif)**

You are not very satisfied with the results of your analysis. Thinking more in-depth you realize that since you have data for two periods for each household in the sample, you can use these data to solve some of the challenges encountered in the previous cases. You will now compare the change in out-of-pocket health expenditures between the households that enrolled in the program and those that did not enroll.

1. Generate a new variable with the difference in out-of-pocket health expenditures between baseline (round = 0) and follow-up (round = 1). Call the new variable “delta\_he”.
2. Compare the average change in out-of-pocket health expenditures between the households that signed up for the program and those that didn’t in treatment localities. Assume we are in the random scenario, so make sure to use the variable enrolled\_ro.
3. Run an OLS regression to estimate the effect of the program on out-of-pocket health expenditures using difference in differences:
   * 1. Without controls
     2. Including characteristics of the household head and spouse
     3. Including baseline covariates
4. What is the “effect” of the program on out-of-pocket health expenditures?
5. How does this method improve upon the estimates obtained in Case 1? Are there any potential biases you are still concerned about?

**Impact Evaluation Case 7: Matching**

Wanting to improve on your previous estimates, you decide to use some matching techniques to select a group of participant and non-participant households that look similar based on observable characteristics.

1. Estimate the probability of participating in the program (pscore) given the observable covariates of participant and non-participant households (use a logit regression followed by “predict”). Once again, assume we are in the random offering scenario.
2. Generate quintiles (5 groups) of the predicted probability of participating in the program (xtile).
3. Compare the average of the covariates between the households that took up the program and that didn’t within each quintile. Which quintiles are best balanced?
4. Use a linear regression to estimate the effect of the HISP on the health expenditures of households that decided to enroll, within the most balanced quintiles.
   * 1. Without controls
     2. Including pscore as a control plus squared and cubic terms for age, education and household size. Does this improve your result?
     3. Including baseline covariates
5. What is the “effect” of the program on out-of-pocket health expenditures?
6. How does this method improve upon the estimates obtained in Case 2? Are there any potential biases you are still concerned about?
7. Stata has a command, psmatch2, that more formally runs a regression using matching (you can install psmatch 2 with the command “ssc install psmatch2, replace”). Use psmatch2 to run the regression and see if you get the same results.

**Impact Evaluation Case 8: Power Calculations**

Let us say that the minister of health was pleased with the quality and results of the evaluation of the Health Insurance Subsidy Program (HISP). However, before scaling up the HISP, the government decides to pilot an expanded version of the program (which they call HISP+). HISP pays for part of the cost of health insurance for poor rural households, covering costs of primary care and drugs, but it does not cover hospitalization. The minister of health wonders whether an expanded HISP+ that also covers hospitalization would further lower out-of-pocket health expenditures. They ask you to design an impact evaluation to assess whether HISP+ further lowers health expenditures for poor rural households. In this case, choosing an impact evaluation design is not a challenge for you: HISP+ has limited resources and cannot be implemented universally immediately. As a result, you have concluded that randomized assignment would be the most viable and robust impact evaluation method. The minister of health understands how well the randomized assignment method works and is supportive. To ﬁnalize the design of the impact evaluation, you need to determine how big a sample is needed. Note that since we are using a random assignment scenario, you can drop the ineligible (eligible = 0) households from the dataset to simplify your code.

1. Use data from the follow-up HISP survey to obtain the benchmark mean and standard deviation for the two outcome indicators of interest to the minister of health—health expenditures and hospitalization.
2. Determine the sample size needed for a minimum detectable effect of $1, $2, and $3 decrease in household out-of pocket health expenditures. Compare the sample sizes required depending on the power level you use, 0.8 or 0.9.
3. Determine the sample size needed for a minimum detectable effect of 1%, 2%, and 3% increases in the hospitalization rate. Compare the sample sizes required depending on the power level you use, 0.8 or 0.9.
4. Which sample size would you recommend estimating the impact of HISP+ on household health expenditures? Would that sample size be sufficient to detect changes in the hospitalization rate?
5. Now assume that the HISP+ program will be assigned at the village level, thus generating clusters. First, find the intra-cluster correlation for health expenditures.
6. Determine the sample size required if the maximum number of clusters possible is 100 (50 treatment villages and 50 comparison villages), using a power level of 0.8. How do the number of clusters and units per cluster change by the minimum detectable effects of $1, $2, and $3?
7. Assume a minimum detectable effect of $2. How does the total number of observations required vary with the total number of clusters available? Determine the sample sizes for scenarios where the total number of clusters is 30, 58, 81, 90, and 120. Use a power level of 0.8.
8. In how many villages would you advise the minister of health to roll out HISP+?
9. Power calculations can also be done in the program Optimal Design, which is available for free download online. Plot the following graphs in Optimal Design:
   1. The relationship between total number of subjects and power, with minimum detectable effects of $1, $2, and $3. Note that you will need to convert the effects into standardized effect form.
   2. The relationship between total number of clusters and power, assuming 102 units per cluster.
   3. The relationship between number of units per cluster and power, assuming there are 100 total clusters available.

**Data Appendix**

1. The analysis database (in STATA format) is “evaluation.dta”
2. Outcome variable:
   1. health\_expenditures = out of pocket health expenditure (per capita per year)
3. Control variables (covariates):
   1. Household head and spouse characteristics:
      1. age\_hh = age of the head (years)
      2. age\_sp = age of the spouse (years)
      3. educ\_hh = education of the head (completed years of schooling)
      4. educ\_sp = education of the spouse (completed years of schooling)
      5. indigenous = 1 if the head speaks an indigenous language
      6. female\_hh = 1 if the head of the household is a woman
   2. Baseline characteristics
      1. hhsize = household size at baseline
      2. dirtfloor = dirt floor household at baseline = 1
      3. bathroom = household with private bathroom at baseline = 1
      4. land = number of land hectares at baseline
      5. hospital\_dist = distance to closest hospital
4. Other variables:
   1. locality\_identifier = locality identifier
   2. household\_identifier = unique household identifier
   3. round = survey round (0 = baseline; 1 = follow-up)
   4. enrolled\_ro = household that enrolled in the program = 1 under the random offering scenario
   5. enrolled\_rp = household that enrolled in the program = 1 under the random promotion scenario
   6. eligible = household eligible to enroll in the program = 1
   7. treatment\_locality = treatment community = 1
   8. poverty\_index = poverty index 1-100 (eligible ≤ 58)
   9. hospital = household member visited hospital in the past year = 1