HAD 5744: Health Econometrics 1
Assignment 3
Dr. Hoagland

Assignment #3
DUE:

Late assignments will be penalized at the rate of 10% per day
Note there is a **hard cutoff for submission** of this assignment **on December 8, 2024**

Total points: 100 possible

Please complete two problems for this assignment: problem 1 (Difference-in-Differences) and any one of the remaining two problems. I recommend choosing to explore a method that aligns with your research interests, particularly for future planned projects.

HAD 5744: Health Econometrics 1
Assignment 3
Dr. Hoagland

1. **Difference-in-Differences: COVID-19 and Sourdough Consumption.** (Adapted from Nick Huntington-Klein and Peter Nencka.) During the early days of COVID-19, there was a brief craze for homemade sourdough bread, as stores were out of yeast (sourdough can be made at home using yeast from the air and does not require store-bought yeast). We will be estimating whether COVID lockdowns actually increased interest in sourdough bread.

   We will be measuring interest in sourdough bread using Google Trends data in the USA. Google Trends tracks the popularity of different search terms over time. We will be comparing the popularity of the search term "sourdough" against the control groups: the search terms "cereal," "soup," and "sandwich," the popularity of which we suspect might not have been meaningfully affected by COVID lockdowns.

   a. Load the data set "a3_p1_sourdough_trends.csv" and look through the data. Make a line graph with `date` on the x-axis and `hits` on the y-axis, with a separate line for each `keyword`. Also add a vertical line for the "start of the pandemic" which we'll decide for our purposes is March 15, 2020. Describe what you find – does it lend support for a particular hypothesis? That is, what is (a) the shape any potential effect takes (i.e. is it a permanent increase in popularity? Temporary?), and (b) whether you might be concerned about any of the control groups we've chosen.
   b. Create a "Treated" indicator for sourdough. Perform a test of whether the prior trends differ between the treated and control groups, using a linear trend. Then, if you were concerned about any of the control groups in part (b), drop them (and keep them dropped for the rest of the assignment) and rerun the test. Interpret your test results—does anything concern you here?
   c. Estimate a difference-in-differences model (where the treatment occurred on March 15) of the effect of a lockdown on sourdough popularity. What are your two fixed effects? Report and interpret your results, with your standard errors clustered at the `keyword` level.
   d. What do you conclude about the effect of the COVID-19 pandemic on sourdough bread consumption?

2. **Synthetic Control.** Many countries and provinces have enacted programs to incentivize organ donation, but what are the effects of these programs? Bilgel and Galle (*Journal of Health Economics, 2015*) found strong support for the implementation of tax deductions on the volume of living (nonrelated) organ donations. More recently, many have argued that transitioning from a tax *deduction* to a tax *credit* is a more equitable way to handle incentives for living organ donation.

   In this problem, we will assess the effects of one such transition, when Louisiana started a tax *credit* for donations in 2015. We will use US data on kidney donations from the Organ Procurement and Transplantation Network (OPTN).

   a. Load the "a3_p2_organ_donations.csv" dataset. Present a figure showing the rates of living organ donations for each state over time, with a vertical line at 2015, the year of Louisiana's adoption – make sure that Louisiana is clearly visible. Discuss your results. Note that all variables should be **population-adjusted.**
   b. Would a typical difference-in-differences strategy here suffice to evaluate Louisiana's policy? Why/why not? Why would/wouldn't a synthetic control approach be better?
   c. Based on (a) and your poking around in the data, are there any states that should be dropped from your donor pool? Justify your decision and drop any states you select for the rest of the assignment.
   d. Construct a synthetic control using your donor pool and the following variables: all pre-treatment levels of the outcome, pre-treatment levels of total donations, state population, and state GDP. Report the relative weights of both variables and donor units. Discuss.
   e. Show a balance table between synthetic and real Louisiana. Also include the main figure showing the trends for real and synthetic Louisiana. Was your synthetic control construction successful? Justify, and speculate as to what might have contributed to this.
   f. What is the estimated effect? What do you conclude about Louisiana's decision to move from a tax deduction to a tax credit? What are the policy implications of your findings?

   g. **Extra credit:** Conduct two placebo tests:
      i. First, rerun the synthetic control with each state in your donor pool as "treated" at 2015. Show the estimated effect for Louisiana in the full distribution of these estimated placebo effects. What is the percentile of Lousiana's estimated MSPE compared to those in the full distribution?
      ii. Second, rerun the synthetic control with the actual date of implementation in 2018 (the date the law went into effect). How does this change your estimated effect?

3. **Quantile and Nonparametric Regression.** In the U.S., traditional Medicare is a universal, publicly-funded health insurance that is available only to individuals aged 65 and older. This problem assesses how qualifying for Medicare (by turning 65) gives individuals new from medical expenditure risk, and how that differs across the population. This problem is based on Barcellos and Jacobson, 2015.

   At age 65, out-of-pocket expenditures drop by 33 percent at the mean and 53 percent at the ninety-fifth percentile. Medical-related financial strain, such as difficulty paying bills and collections agency contact, is dramatically reduced. This problem asks you to dig into this empirically.

   a. Load the "a3_p3_medicare.RData" dataset. What is the distribution of out-of-pocket medical spending ("totexp") for individuals before they are on Medicare versus after? Does this tell you anything about the financial protection afforded to Medicare patients – why or why not?
   b. What is a naïve OLS estimate of the effect of having Medicare (i.e., age >= 65) on expected total expenditures? Control for any variables you feel are relevant, and be sure to weight your regressions using the "perwt" weighting variable (throughout this problem).
   c. Now report quantile regression results at 20 evenly spaced quantiles of the income distribution ("faminc"). Report the raw regression coefficients in a figure (feel free to drop outliers, as long as that is noted). How do you interpret these results?
   d. Scale these coefficients by the quantile value of the income distribution and recreate the figure. How does this scaling change the interpretation of the results?
   e. Instead of quantile regression coefficients, construct a binscatter that illustrates the relationship between the income distribution and overall health expenses. Report two relationships: one for individuals on Medicare and one for individuals who are too young to qualify. Adjust for the relevant covariates and weights you used in parts (b) and (c). How does the difference between these two curves change across the distribution? What is the interpretation of that change?
   f. What do you conclude about the distributional effects of Medicare as a form of risk protection? Can you think of any practical mechanisms (open back-doors) that explain your findings? What are the policy implications of your findings?

   g. **Extra credit:** What do your results from (e) imply about the use of nonparametric regression here? Specify and implement a nonparametric conditional density estimator (of your choice) estimating the effect of Medicare coverage over the income distribution. Discuss the assumptions you would need for this estimator to be valid and interpret your results.