MGRECON 548Q: Empirical Economic Analysis
Fuqua School of Business, Duke University

Team Project

Instructions:

- This is a team assignment.

- You and your pre-assigned team will make one submission.

- All team members will have the opportunity to earn the same grade, however, each student's individual efforts will be subject to a peer review.

The goal of this assignment is for you to apply all that we have learned throughout the semester. Your task is to conduct a causal analysis on a topic and question of your choosing.

Deliverable & Submission

- The deliverable consists of two components: i) a slide deck, submitted either in powerpoint or PDF format (75%), and ii) code supporting your analysis (25%).

- The slide deck should be fully self contained, implying that the reader need not be forced to look at the code or outside resources to understand your analysis or conclusions. You may use "notes" feature in powerpoint to provide details that are not easily conveyed on a slide. Design the slide deck to so that the grader can view the slides and notes within approximately 20 minutes.

- The code should be well organized and documented heavily. All accompanying files and files should be indicated in an accompanying README.txt file.

- All documents (presentation, code, data) should be zipped and submitted via Canvas. Only a single submission per team, wherein the zip file is titled with your Team number.

Content & Requirements

- Pick a topic of your choosing. There are no constraints here, but I recommend choosing something that is within your domain of professional interest: finance, health, etc...Then, identify a causal question of interest.

- Your slide deck should include the following sections. The percentages next to each item are the point allocations for the presentation portion of yoru assignment grade.

- Motivation (10%) — Why is this question of interest and important?

- Literature (10%) — Conduct a brief review of prior academic and industry reports that have addressed this question. How (if at all) is your attempt to answer the question different than the prior literature (e.g. different technique, different data, etc..)?

- Methodology (50%) — Detail your methodology. Explain each step, ensuring that you justify how your proposed method addresses the causal question of interest. Be sure to discuss the potential presence of omitted variable bias. Note: No single technique can answer a question completely. Often, multiple techniques are required. To the extent possible, be sure to implement at least one of the advanced techniques we've developed in class (See below for list) to elicit causality. The more advanced techniques you can employ, the stronger will be your deliverable. If your particular problem does not support multiple advanced techniques, be sure to detail those reasons.

- Data (5%) — Detail your data source(s) and provide exploratory analyses (EDA).

- Findings (20%) — Discuss your findings. Be sure to contrast with the prior literature and your expectation. Also be sure to connect back to the causal question of interest.

- Conclusion (5%) — Summarize. Provide any caveats or limitations to your analysis.

Tips for Success

- Milestones

  - Pick an area for research (e.g. finance, health care, etc..)
  - Crystallize a specific question
  - Find data
  - Clean data
  - Define methodology
  - Implement methodology
  - Refine
  - Draft
  - Final deliverable

- Workflow

  - Ensure all team members know their roles and responsibilities.
  - A project manager / lead is invaluable.
  - Assign very clear tasks. These should be actionable, have a "RP" (responsible person), and have a due date. Post these tasks somewhere for all team members to monitor.
  - Foster an open line of communication within your team. Use technology channels for chatting and document sharing for all team members to access; e.g. nothing should be stored on someone's laptop locked away from the rest of the team.
  - Assign dates for completion of each milestone (be flexible/agile to alter and revise as needed). Pro Tip: work backwards from the project completion date when deciding milestone deadlines.

- Picking a topic:

  - The best projects i) are aligned with your career interests, ii) challenge you to exercise the skills we are developing, and iii) are tractable given the time/labor/computing resources available.
  - Avoid the common trap of starting your ideation on Kaggle. Data availability is indeed important, but this approach often wastes time (e.g. we've spent 8 days cleaning and doing EDA on a random dataset we found on Kaggle and just now realized that we don't have a treatment variable).
  - A common approach is for the group to pick a general field of interest (health care, finance, etc..). Next, the group must find a specific causal question within that field. Beware of jumping onto the first idea. An honest intellectual exercise should be well vetted. Think about the risk/reward of the entire exercise (e.g. topic A might have easily accessible data, but if we can find data for topic B it will allow us to use more sophisticated techniques.)
  - Ensure that there is a causality question.

  Methodology:

  - Don't fall into the trap of simply running a bunch of regressions and identifying the statistically significant variables. Such an approach does permit us to ask well structured causal questions.

- Leverage academic literature and professional research, which can act like a "subject matter expert", providing guidance on expected signs of any biases and possible omitted variables.
- The research cycle does NOT start with data/EDA. The order should be Whiteboard → Keyboard, then iterate.
- Although forecasting might be useful, this is NOT our focus. Although data exploration and visualization is powerful, that is NOT our focus.

- Summary of Causality methods we cover in this course:
  <u>Basic</u>

  - Random Sampling
  - OLS

  <u>Advanced</u>

  - Regression Discontinuity: "Does the MLDA (minimum legal drinking age) impact driving deaths in young people? Compare drivers at 20yrs old and 22yrs old."
  - Instrument Variables: Solution to OVB by finding another variable that impacts the treatment variable but not the omitted variable. "What is the impact of participating in a special private educational program on student grades? The problem is that those parents seeking this program might be systematically different than those parents who prefer to remain in public school. This leads to OVB. We can use a dummy variable to indicate whether the student won a random lottery to get into the program as an instrument to elicit the treatment effect."
  - Difference in Difference: "Does the MLDA impact driving deaths in young people? Compare drivers in state with 21yrs and similar state with 19yr old."
  - Panel Regression: "Does the MLDA impact driving deaths in young people? Compare multiple states over time, with control variates."
  - Simultaneous Equations: "What is the impact of savings on housing expenditures? Problem is that housing expenditures =f(savings), and savings=f(housing exp)."