

Research Project

This note provides detailed information about the class project. The goal of the project is to give you an opportunity to apply methods discussed in this class to a problem of interest to you, and to share your findings with the rest of the class.

Objectives

The purpose of the final project is to deploy some of the methods introduced in class on an empirical application of your choice. The question should involve causal inference (including prediction policy questions) and make use of machine learning methods or other methods introduced in the course.

You will be assessed on the following aspects of your project. Of these 3 metrics, the most important is #2, i.e., the quality of the empirical component of the project.

1. **Quality of research question.** Did you formulate a question whose answer is interesting? Did you identify a dataset that can credibly be used to answer the question?
2. **Quality of empirical strategy and execution.** A focus of this class is in surveying methods that can leverage flexible data sources and rigorously reflect causal identifying assumptions. A good empirical approach will involve carefully justifying all modeling assumptions made.
3. **Quality of final presentation.** Did you communicate and document your findings in a clear, polished presentation?

For your research question, we recommend being as explicit as possible about what the causal or policy question is, and about what your identification strategy is. Is your data from a randomized trial, or are you relying on observational study assumptions? If you have panel or longitudinal data, what are you assuming about dynamics?

For your empirical strategy, we recommend comparing and benchmarking at least 2–3 alternative methods whenever possible (e.g., when choosing a predictive method to use). We also recommend using simulation-based or semi-synthetic evaluations to assess your empirical strategy.

Forming Groups

The assignment is to be completed in groups of **at most 4 students**. On Wednesday, April 6, we will have an in-class session to facilitate group formation. Please bring an introductory slide about yourself, ideally printed out. We will break into groups and you will go around and introduce yourself and describe your background and research interests. We will remix groups several times so that you can meet other students.

Choosing a Topic

Once you have formed a group, you should decide on a topic. There are two main approaches for doing so:

- **Bring your own problem and dataset.** You choose your own research question, and work based on it. For PhD students, questions related to your thesis work or other research interests are particularly encouraged.
- **Use a provided dataset.** We maintain a list of dataset of publicly available experiments at <https://github.com/gsbDBI/ExperimentData>. You may use one of these datasets as a basis for your project. However, as a warning many of those datasets are noisy and do not have a lot of treatment effect heterogeneity.

Regardless of how you choose your topic, we recommend running your question by a member of the teaching team (e.g., in office hours) to make sure we don't identify any obvious issues.

Logistics

There are two **deliverables** for this assignment. The written part must be submitted on the Canvas website by Wednesday, June 1, at 11:59pm.

The first part is a **3–5 page write-up** of your findings (12 pt., 1.5 spacing, 1 inch margins, page count not including any figures or references). The write-up should explain your question, your choice of methods, and what conclusions you may draw. You should also submit your code and output, preferably produced as a knit file that shows the code and the results together.

The second part is a **class presentation** of the project during the last week of class.