QTM Honors Thesis Proposal

Research Problem:

Can we observe treatment effect heterogeneity for a randomized control trial using images as a source of data?

Background:

In 2023, Connor T. Jerzak, Fredrik Johansson, and Adel Daoud were interested in examining treatment effect heterogeneity for a randomized control trial (RCT) studying the effect of a youth development program in certain regions of Uganda. Traditional RCT analyses, which rely on tabular variables like age and ethnicity taken much closer to the time of the experiment, often miss broader historical and geographical factors that can affect the conditional average treatment effect (CATE), or the outcome. Jerzak et al. were interested in filling this gap using satellite imagery, which provides vital data on the environmental and locational context of experimental units.

Motivation:

I have been working as a research assistant for Dr. Alejandro Sanchez-Becerra through the QTM Fellows program since August of 2023, on a project to collect NGO information. I have experience working on projects involved with poverty-alleviation programs and understand how important these findings can be for expanding and improving initiatives for the cause. My research back then largely consisted of collecting data, but as I have more experience dabbling with complex statistical methods now, I would like to expand on the methods that I have learned in class, such as Causal Designs and Inference, and apply these concepts to real world applications. By applying these models, I seek to uncover insights from previously conducted RCTs in global poverty and explore the broader applicability and validity of these methods.

Objectives & Methodology:

RCTs have long been the ideal choice for calculating the ATE of interventions. By replicating the experiment that Jerzak et al. have conducted, I can gain a deeper understanding of this new approach utilizing images for treatment effect heterogeneity. Jerzak created a package in R called "causalimages," a tool that can be used for causal inference with earth observation, bio-medical, and social science images. This package leverages TensorFlow networks that are commonly used in Python and transitions these models to R for statistical purposes. The first step of the thesis would be to replicate his findings by using satellite images provided with the tutorial of the package. Once I am accustomed to the package and can interpret the use cases, I aim to combine these findings and methodology with Dr. Sanchez Becerra's research (developing statistical tools to improve the evaluation of poverty-alleviation programs) and further investigate the validity of the results.