

MASTER'S THESIS PRESENTATION

Development of a Python Package for Matching Observational Data

WHEN

May 5, 2022

12:20 PM

WHERE

Zoom Meeting

For ZOOM presentations, details will be provided in an email announcement for this seminar.

**Jack Potrykus, MS candidate**

Observational studies are widely used throughout econometrics, psychology, and medical research. Matching is a field in causal statistics concerned with algorithms to minimize the effect of selection bias in the observational data on analyses of treatment effects. In particular, in binary treatment/control studies, these algorithms work by matching each treatment observation to one or more “nearby” control observations. This paper breaks the matching procedure down into two key components: how distance is measured, and how matches are assigned. In doing so, it explores several distance metrics, in particular the propensity score (Rosenbaum and Rubin 1983), the prognostic score (Ben B. Hansen 2008), and exact matching (Iacus, King, and Porro 2012) and its machine learning extensions (Liu et al. 2019; Wang et al. 2021); it then explores several matching algorithms that can be used to produce the matched subset once the distance is calculated, in particular the Hungarian algorithm (Munkres 1957) and greedy algorithms (D. Ho et al. 2011). I then showcase the functionality of matching, an open-source Python package I have developed for matching observational data which takes a graph-centric approach, something which no other package offers. I finally explore some practical considerations of parameter tuning when matching via experiments on simulated data.

stat.uchicago.eduTHE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS