



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

MASTER'S THESIS PRESENTATION

KENTARO NAKAMURA

Department of Statistics
The University of Chicago

Improving Text as Treatment Framework

MONDAY, April 17, 2023, at 10:00 AM
in Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Social scientists increasingly use machine learning to extract features of interest from high-dimensional data like texts. The extracted features, such as topics, are sometimes used as a treatment for causal inference, called "latent treatment" (Fong and Grimmer 2021). In the current literature, the required assumptions to use latent treatment are restrictive: even when texts are randomized, researchers must identify all the features correlated with latent treatment of interest and influence the outcome. The failure to enumerate such features can lead to significant confounding bias, but in a high-dimensional setting like texts, it is tough to list all these features. Moreover, the problem of prediction error makes the situation more complicated. In this paper, I develop a novel methodology to estimate the effect of latent treatment based on double machine learning and instrumental variable methods. I describe the identification strategies, assumptions, and estimations procedures. The performance of the method is confirmed through simulations with actual text data.