On the Importance of Heteroskedasticity in Causal Inference

Jieyu Gao[1] and Ivan Jeliazkov[2]

Abstract

There is a large and rapidly growing causal inference literature, yet little is known about how the presence of heteroskedasticity impacts these estimation approaches. In this paper, we attempt to fill this gap by recognizing, understanding, and carefully documenting the impact of heteroskedasticity in several popular causal models. We build upon and extend existing Bayesian methods for several well-known settings such as regression discontinuity designs (both sharp and fuzzy), models based on the potential outcome frameworks, and propensity score matching. Key features of the Bayesian approach in these settings include flexible modeling and context-specific computationally efficient estimation algorithms, the ability to recover functions of the treatment parameters instead of only focusing on averages, as well as the fact that estimation employs all available data, rather than subsets based on proximity. Simulation studies are used to gauge the modeling and algorithmic methodology based on approaches in discrete data and time-varying parameter analysis. In addition, we also contribute on the applied side by examining applications involving the effect of academic probation on students' subsequent academic performance and the effect of Medigap on healthcare expenditures. The findings in both applications provide strong evidence for the necessity of accounting for the presence of heteroskedasticity and illustrate the practical consequences of misspecification.

Jieyu Gao: University of California, Irvine; jieyug1@uci.edu
Ivan Jeliazkov: University of California, Irvine; ivan@uci.edu