BETTER POLICIES FOR BETTER LIVES

Instrument Validity Tests with Causal Trees

Helmut Farbmacher, Raphael Guber

Discussion by Nicolas Woloszko (OECD)

Modelling with Big Data and Machine Learning: Interpretability and Model Uncertainty

Bank of England and King's College

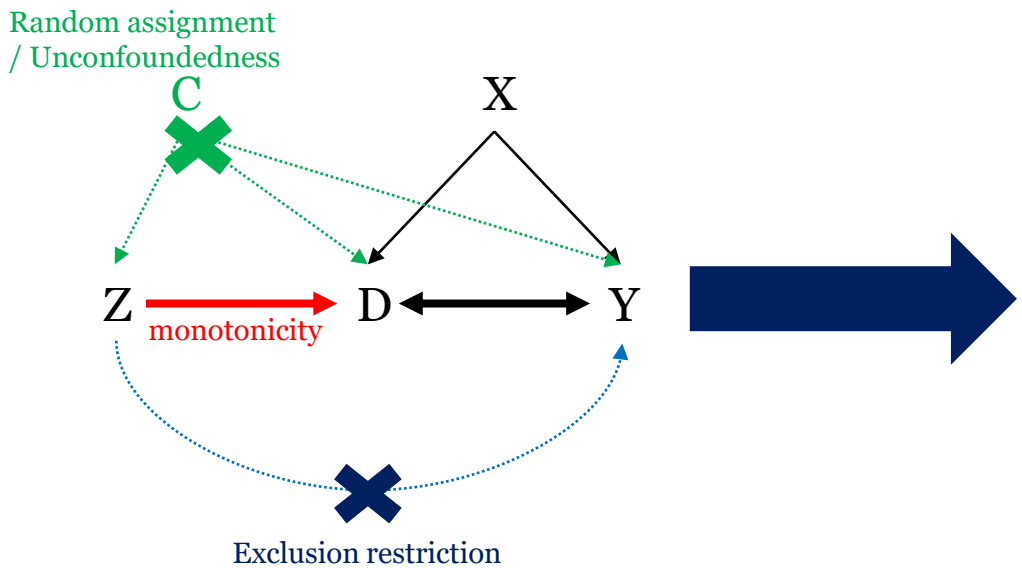
November 4 2019

Main contributions

- Provides a methode to search the feature space for violation of LATE assumptions (Imbens, Angrist, 1994).
- Expands from existing work (including Huber 2015) that searches for feature subspaces where LATE assumptions are violated by relying on:
  - Causal Trees (Athey, Imbens, 2015) and Causal Forests (Athey, Wager, 2018)
  - a multiple testing procedure based on Chernozhukov *et al.* 2016.
- Provides an application to the analysis of the effect of #children on labour supply using sex composition as an instrument

OECD

A clever use case of causal trees



$$\forall B, d, \tau(d, x) = E[Q_d(B)|Z = 1, X = x] - E[Q_d(B)|Z = 0, X = x] \leq 0$$

- The LATE assumptions are derived into a necessary condition that has a causal form
- This allows to use of causal trees/forests in order to test for LATE assumptions violations

Practical difficulties of implementation

- Choice of B in the case of a continuous outcome
- Dealing with « *honesty* »: sample splitting schemes lead to juggling with random trees/forests or using *cross-fitting* (Chernozhukov et al. 2018)
- It could be useful to have the procedure in pseudo-code, and/or make the code publicly available.

OECD

Applications

- All papers that pass this new test will implement it ☺ Unless going local becomes a new standard.
- Brings valuable insights into the relationship between instrument and treatment. In the China application, first-born being a male is a key driving factor.
- Other *derived* applications of causal forests and causal ML methods