Heterogeneous Treatment Effect Analysis Based on Machine Learning Methodology

Xiajing (Jean) Gong, Meng Hu, Liang Zhao

Division of Quantitative Methods and Modeling, Office of Research and Standards, Office of Generic Drugs,
Center for Drug Evaluation and Research, U.S. Food and Drug Administration (FDA), Silver Spring, Maryland, USA

INTRODUCTION

- Heterogenous treatment effect (HTE) analysis focuses on examining the variation of treatment effects over the population. A good understanding of HTE information can provide a more accurate advice on the treatment plan to achieve optimal outcome (e.g., for personalized medicine), as compared to average effect analysis.
- One unique challenge of conducting HTE analysis is that the quantity to be estimated (i.e., treatment effect) is often unknown on a given dataset, as in many cases each subject can only be exposed to one condition of treatments. Regression-based models, such as the conventional two-step method, were developed to conduct the HTE analysis, but their model performance would be compromised when dealing with nonlinear and/or high-dimensional data.
- Recently, the machine learning (ML) methodologies have been employed in the HTE analysis, especially with the tree-based approaches. As a latest advances in tree-based HTE method, causal forest [1] – a method based on random forest - has been developed to conduct HTE analysis, exhibiting flexibility to handle complex data while sidestepping potential issues by the single-tree method.

OBJECTIVES

- To conduct simulations to mimic scenarios with different levels of complexity in term of HTE, by which the causal forest and two-step method can be systematically examined with respect to the ability to identify the treatment effect heterogeneity.

METHODS

❑ Simulation of HTE data

- Interactions between treatment and covariates of subjects can lead to the HTE among the study population.
- By changing the relations of treatment indicator and heterogeneity covariates in the outcome model, various scenarios (models) were created:

I. no heterogeneity covariates
II. linear relationship between heterogeneity covariates
III. nonlinear relationship between heterogeneity covariates
IV. high dimensional data

❑ ML-based and conventional HTE analysis methods

- We adopted causal forest and the two-step method as the proxies for ML-based and conventional method respectively, for HTE analysis and performance evaluation.

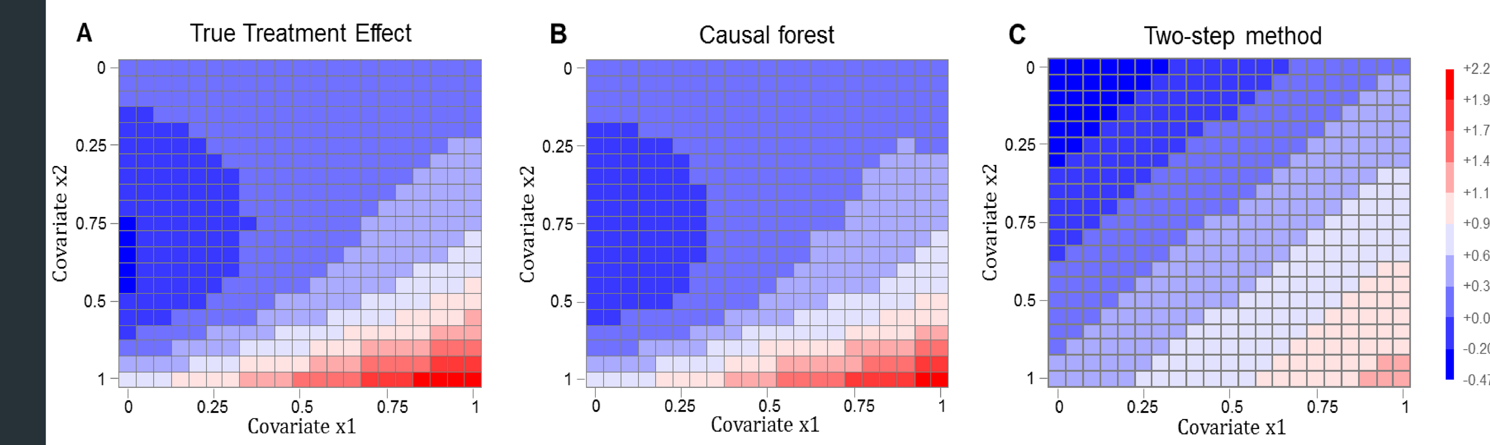
Casual forest, a machine-learning-based method, is a promising tool in real-world applications for heterogenous treatment effect (HTE) analysis.

It can potentially be used to improve business intelligence in the Agency.

Take a picture for more **details**

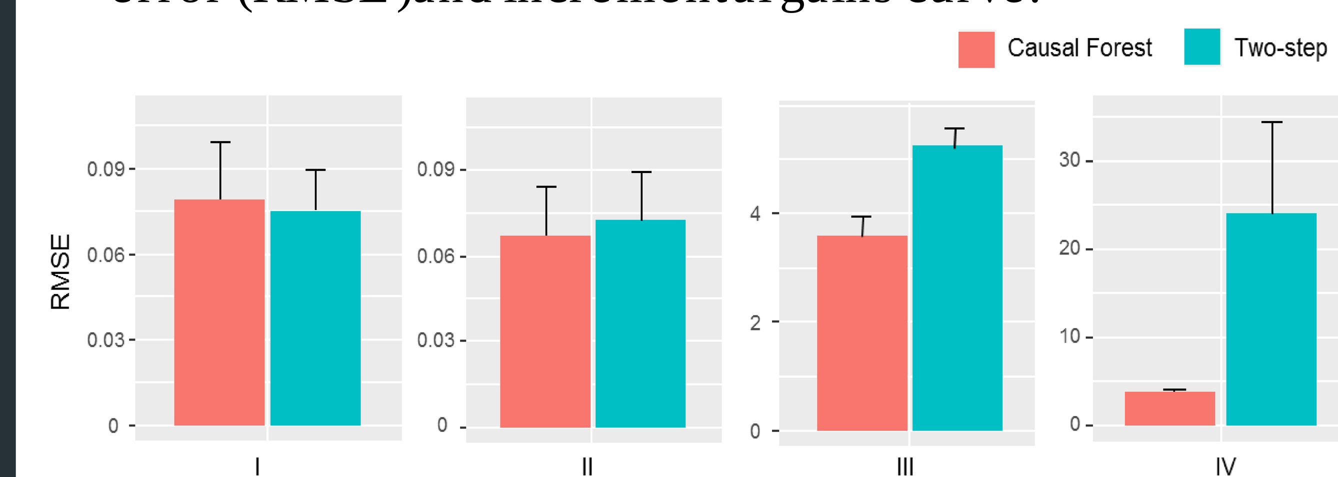
RESULTS

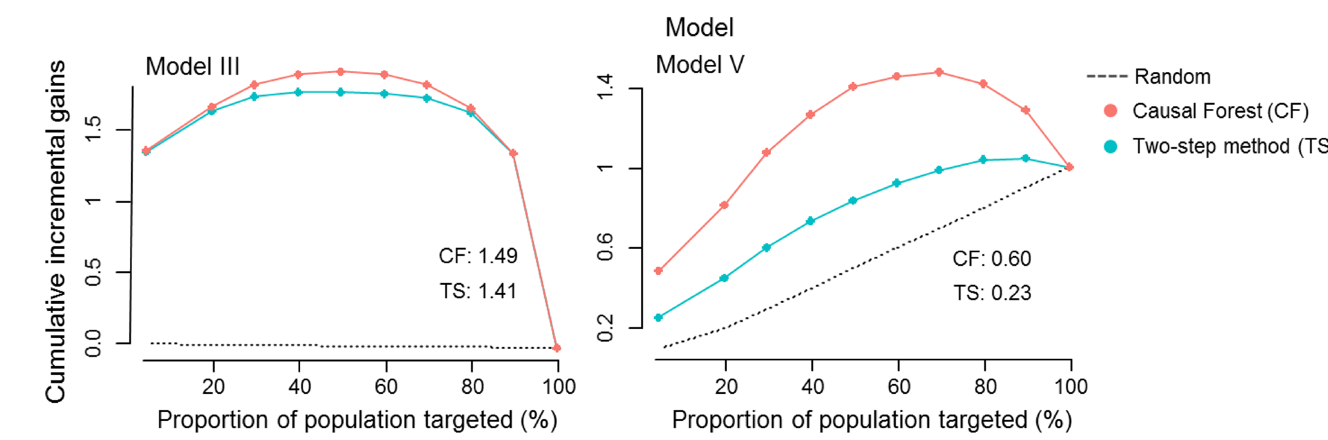
Exemplary Case: treatment effect of a dataset with nonlinear relationship between the two covariates x1, x2 and treatment effect.



Simulations based on Model I-IV

Causal forest and two-step method were applied on simulation data to perform HTE analysis. Prediction performance were evaluated in terms of root mean square error (RMSE)and incremental gains curve.





Causal forest can outperform the linear regression-based two-step method for HTE analysis with nonlinearity and high-dimensionality present in the data.

REFERENCES

1) Wager, S. and S. Athey, *Estimation and inference of heterogeneous treatment effects using random forests.* J Am Stat Assoc, 2018. 113(523).

U.S. FOOD & DRUG ADMINISTRATION