# When Causal Forests Mislead
## Evaluating the precision of Confidence Intervals

**Author**
Rares Iordan
r.iordan@student.tudelft.nl
**Supervisor**
Rickard Karlsson
**Professor**
Jesse Krijthe

This study tackles an important issue in evaluating the reliability of confidence intervals in causal forests by examining how data characteristics and hyperparameters influence actual coverage rates compared to theoretical benchmarks. A primary discovery is the identification of a practical limit for reliable confidence interval coverage: When the sum of confounders and effect modifiers exceeds 4, coverage rates drop considerably below 80%, even for simple treatment effect functions.

## 01 Introduction

- Machine learning algorithms excel at prediction but struggle with causal relationships
- In fields like healthcare, understanding how interventions affect outcomes is often more valuable than just predicting the outcomes
- Treatment effects are heterogeneous - different subpopulations respond differently to the same treatment
- Understanding this heterogeneity enables personalized decision-making and targeted interventions
- Causal forests can estimate these effects, but their statistical properties remain uncertain

## 03 Research Questions

1. How sensitive are confidence interval coverage rates to data characteristics (confounders, effect modifiers, polynomial complexity, confounding strength)?
2. Which hyperparameters most influence coverage rates and how?
3. Do tree count and sample size interact to affect coverage rates?

## 04 Methodology

- Data Generating Process: Polynomial data-generating process that generates synthetic observational datasets with known causal structure
- Sobol Sampling: 1536 samples over five parameters: Polynomial degree {1,2,3,4,5}, Confounding strength (0,10], Number of confounders/instruments/effect modifiers {0,1,2,3,4,5,6,7,8}
- HDMR Analysis: High-Dimensional Model Representation with maximum interaction order of 2 to evaluate sensitivity of coverage rates
- Grid Searches: Exhaustive grid search methodology for hyperparameter analysis

## 05 Findings

- **Critical Threshold Identified**
  - When the combined number of confounders and effect modifiers exceeds 4, coverage rates decline dramatically below 80% even for the simplest treatment effect function. This threshold appears robust, as increasing computational resources provided only marginal improvements.
- **Most Influential Parameters**
  - The number of confounders, their interaction with effect modifiers, and effect modifiers are the dominant factors (sensitivity indices ≈ 0.28 ± 0.04, ≈ 0.14 ± 0.03, and ≈ 0.10 ± 0.02)
- **Hyperparameter Recommendations**
  - max_depth: Leave unset for best results
  - max_samples: Increase to 0.5 for best coverage rate performance
  - min_balancedness_tol: Set to 0.5 for optimal coverage
  - n_estimators: Use ≥ 2400 trees for best performance
  - min_impurity_decrease: Keep at default (0.0) to avoid degrading inference quality

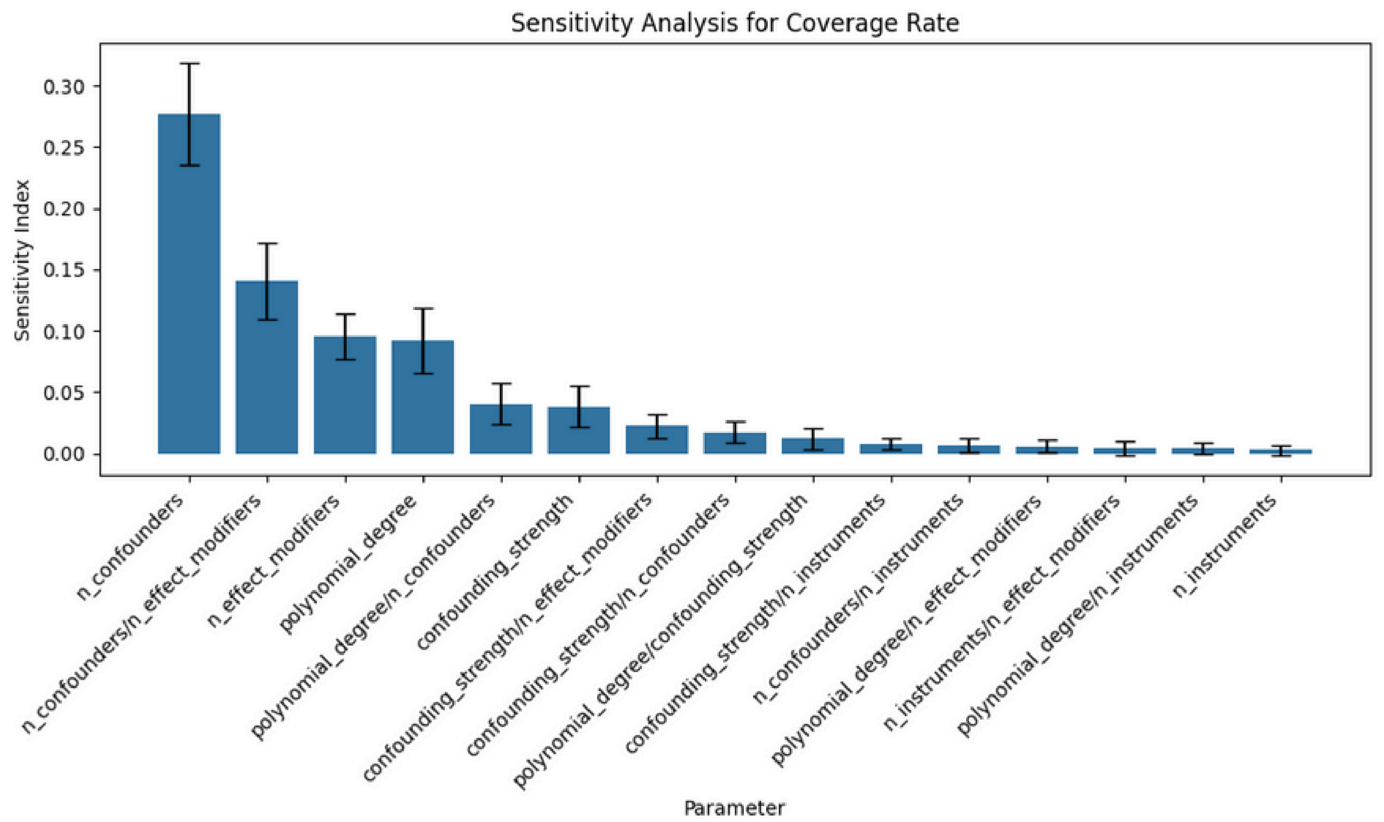
## 02 Background - Causal Forests

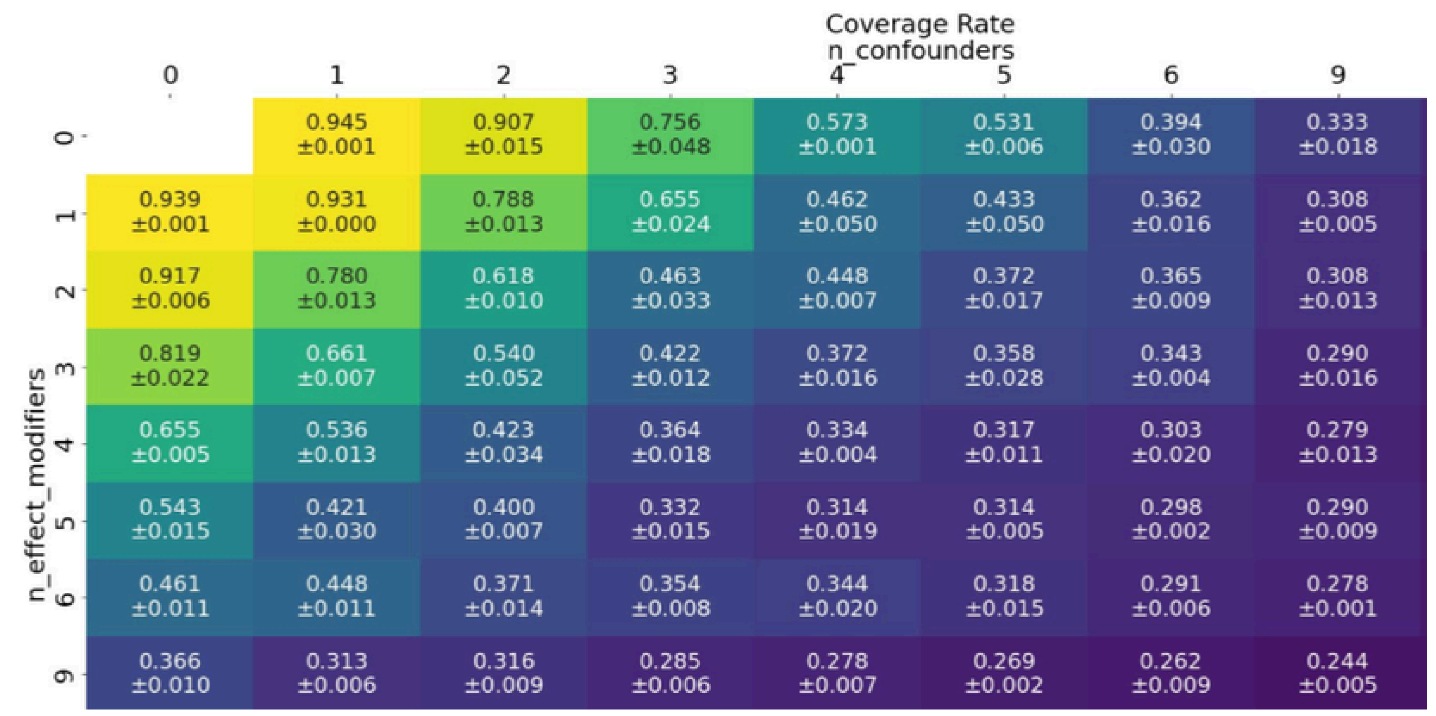
- Evolution from Random Forests
- Adapted for heterogeneous treatment effect estimation by Wager and Athey (2018)
- Later enhanced by Athey et al. (2019)
- "Honest" Trees for valid statistical inference and confidence intervals
  - Use outcome variable Yi either for splitting or leaf estimation, but not both.
- Uses bootstrap of little bags for estimating variance

$$\hat{V}_{BLB}(x) = \underbrace{\frac{1}{G}\sum_{g=1}^{G}(\hat{\tau}^g(x) - \hat{\tau}(x))^2}_{\text{Between-group variance}} - \underbrace{\frac{1}{\ell - 1} \cdot \frac{1}{G}\sum_{g=1}^{G}\frac{1}{\ell}\sum_{b\in g}(\hat{\tau}_b(x) - \hat{\tau}^g(x))^2}_{\text{Within-group correction}}$$

- Treatment effect estimates are asymptotically Gaussian

$$\hat{\tau}(x) \pm 1.96\sqrt{\hat{V}_{BLB}(x)}$$




**Related Literature**
Leo Breiman. “Random forests”. In: Machine learning 45.1 (2001), pp. 5–32.
Stefan Wager and Susan Athey. “Estimation and inference of heterogeneous treatment effects using random forests”. In: Journal of the American Statistical Association 113.523 (2018), pp. 1228–1242.
Susan Athey, Julie Tibshirani, and Stefan Wager. "Generalized random forests". In: Annals of Statistics 47.2 (2019), pp. 1148–1178.