An application of Causal Forests to explore the heterogeneous effects of social health insurance

Noemi Kreif, PhD
Research Fellow
Centre for Health Economics
University of York, UK

17th Armitage Workshop, Cambridge
07/11/2019
Motivation

- Large scale “system level” health policies implemented to achieve Universal Health Coverage
  - E.g. national health insurance programme, primary health care reforms
- Motivated to improve the health of those most in need, but rolled out universally

For whom?

- For the treated
- For the “compliers”

Around the discontinuity

Difficult to communicate, hard to generalise for future policies

Techniques:

- Differences-in-differences
- Instrumental variables
- Propensity scores / double robust methods
- Regression discontinuity design

Average treatment effects
Motivation

- Large scale “system level” health policies implemented to achieve Universal Health Coverage
  - E.g. national health insurance programme, primary health care reforms

- Motivated to improve the health of those most in need, but rolled out universally

- Almost never randomised: estimating the causal effect of policies challenging
  - Differences-in-differences
  - Instrumental variables
  - Propensity scores / double robust methods
  - Regression discontinuity design
Motivation

- Large scale “system level” health policies implemented to achieve Universal Health Coverage
  - E.g. national health insurance programme, primary health care reforms
- Motivated to improve the health of those most in need, but rolled out universally
- Almost never randomised: estimating the causal effect of policies challenging

Average treatment effects

- Differences-in-differences
- Instrumental variables
- Propensity scores / double robust methods
- Regression discontinuity design
Motivation

- Large scale “system level” health policies implemented to achieve Universal Health Coverage
  - E.g. national health insurance programme, primary health care reforms
- Motivated to improve the health of those most in need, but rolled out universally
- Almost never randomised: estimating the causal effect of policies challenging
  - Differences-in-differences
  - Instrumental variables
  - Propensity scores / double robust methods
  - Regression discontinuity design

Average treatment effects

For whom?

For the treated: Differences-in-differences, Instrumental variables, Propensity scores / double robust methods
For the “compliers”: Regression discontinuity design
For the population: Differences-in-differences, Instrumental variables, Propensity scores / double robust methods
Around the discontinuity: Regression discontinuity design
Motivation

- Large scale “system level” health policies implemented to achieve Universal Health Coverage
  - E.g. national health insurance programme, primary health care reforms
- Motivated to improve the health of those most in need, but rolled out universally

- Almost never randomised: estimating the causal effect of policies challenging

  Average treatment effects
  - Differences-in-differences
  - Instrumental variables
  - Propensity scores / double robust methods
  - Regression discontinuity design

For whom?

Difficult to communicate, hard to generalise for future policies
Motivation: heterogeneous treatment effects

- Policy maker needs information on
  - whether a policy worked for a given group
    - E.g. Did subsidised health insurance work for those with the greatest health need (Lancet 2019)

- Pre-specified subgroup analysis restrictive...

- Non-randomised evaluations rarely pre-specified -> "cherry picking"

- Can use the data to learn about important subgroups (vanDerWeele et al. 2019)
Motivation: heterogeneous treatment effects

- Policy maker needs information on
  - whether a policy worked for a given group
    - E.g. Did subsidised health insurance work for those with the greatest health need (Lancet 2019)
  - who were the (relative) winners and losers
    - To learn whether programme improved equity
    - To improve generalisability for the design of future programmes
Motivation: heterogeneous treatment effects

- Policy maker needs information on
  - whether a policy worked for a given group
    - E.g. Did subsidised health insurance work for those with the greatest health need (Lancet 2019)
  - who were the (relative) winners and losers
    - To learn whether programme improved equity
    - To improve generalisability for the design of future programmes

- Pre-specified subgroup analysis restrictive...
- Non-randomised evaluations rarely pre-specified -> “cherry picking”
- Can use the data to learn about important subgroups (vanDerWeele et al. 2019)
Estimating heterogeneous treatment effects

- Ideally want to learn individual treatment effect:
  \[ Y_i(1) - Y_i(0) \] [not feasible]

Still difficult: dimensionality of \( X \) potentially large, continuous & multi-valued

No unmeasured confounders \( Y^1, Y^0 \perp A \mid X \)

Positivity \( 0 < P(A = 1 \mid X) < 1 \)
Estimating heterogeneous treatment effects

- Ideally want to learn individual treatment effect:
  \[ Y_i(1) - Y_i(0) \] [not feasible]

- Instead: conditional average treatment effect (CATE) function:
  \[ \tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] \]

No unmeasured confounders \( Y^1, Y^0 \perp A | X \)
Positivity \( 0 < P(A = 1| X) < 1 \)

Still difficult: dimensionality of \( X \) potentially large, continuous & multi-valued \( X \)
Estimating the *CATE*

1. Through flexible outcome regression $\mu(A, X) = E[Y|A, X]$
   - predict $\hat{\tau}(x) = \hat{\mu}(1, X) - \hat{\mu}(0, X)$ (e.g. Künzel et al. 2019)
   - can use machine learning for $\mu(A, X)$
   - problem: asymptotics unclear
Estimating the \textit{CATE}

1. Through flexible outcome regression $\mu(A, X) = E[Y|A, X]$
   - predict $\hat{\tau}(x) = \hat{\mu}(1, X) - \hat{\mu}(0, X)$ (e.g. Künzel et al. 2019)
   - can use machine learning for $\mu(A, X)$
   - problem: asymptotics unclear

2. Directly estimate $\tau(x)$
   - double-robust transformation of the outcome
   - machine learning methods to “predict” $\tau(x)$ (Wager and Athey, JASA 2019)
Estimating the CATE

1. Through flexible outcome regression $\mu(A, X) = E[Y|A, X]$
   - predict $\hat{\tau}(x) = \hat{\mu}(1, X) - \hat{\mu}(0, X)$ (e.g. Künzel et al. 2019)
   - can use machine learning for $\mu(A, X)$
   - problem: asymptotics unclear

2. Directly estimate $\tau(x)$
   - double-robust transformation of the outcome,
   - machine learning methods to “predict” $\tau(x)$ (Wager and Athey, JASA 2019)

3. Only try to learn predictors of $\tau(x)$ (Chernozokov and Demirer et al. 2018)
4. “Only” try to learn the value of the counterfactual mean outcome under the optimal rule (e.g. Luedtke and van der Laan 2016)
Tree/forest-based prediction algorithms

- **Regression tree** predicts outcome based on average outcomes in a “leaf”
  - Splits tree to minimise in sample MSE, predicts out-of-sample
  
  \[
  \hat{\mu}(x) = \frac{1}{|i: X_i \in L(x)|} \sum_{|i: X_i \in L(x)|} Y_i
  \]

- Individual trees low bias but high variance
- **Random forests**: (weighted) average of predictions from many trees
- Tuning parameters: tree depth, number of trees
  - Can be chosen using cross validation
Tree/forest-based prediction algorithms

- **Regression tree** predicts outcome based on average outcomes in a “leaf”
  - Splits tree to minimise in sample MSE, predicts out-of-sample
    \[ \hat{\mu}(x) = \frac{1}{|i: X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i \]
  - Individual trees low bias but high variance

- **Random forests**: (weighted) average of predictions from many trees
  - **Tuning parameters**: tree depth, number of trees
    - Can be chosen using cross validation
Causal Forests for CATE  
(Wager and Athey 2018, Athey et al. 2019)

Aims to estimate:

$$\hat{\tau}(x) = \frac{1}{|i: W_i = 1, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i - \frac{1}{|i: W_i = 0, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i$$
Causal Forests for CATE
(Wager and Athey 2018, Athey et al. 2019)

Aims to estimate:

\[ \hat{\tau}(x) = \frac{1}{|i: W_i = 1, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i - \frac{1}{|i: W_i = 0, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i \]

1. Builds “causal trees”: maximise heterogeneity in estimated treatment effect as opposed to root mean squared error in prediction of the outcome
Causal Forests for CATE  
(Wager and Athey 2018, Athey et al. 2019)

Aims to estimate:

\[ \hat{\tau}(x) = \frac{1}{|i: W_i = 1, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i - \frac{1}{|i: W_i = 0, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i \]

1. Builds “causal trees”: maximise heterogeneity in estimated treatment effect as opposed to root mean squared error in prediction of the outcome

2. “Honesty”: in one tree, \( i \) is either used to select splits or estimate \( \tau(x) \)
Causal Forests for CATE  
(Wager and Athey 2018, Athey et al. 2019)

Aims to estimate:

\[ \hat{\tau}(x) = \frac{1}{|i: W_i = 1, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i - \frac{1}{|i: W_i = 0, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i \]

1. Builds “causal trees”: maximise heterogeneity in estimated treatment effect  
as opposed to root mean squared error in prediction of the outcome

2. “Honesty”: in one tree, \( i \) is either used to select splits or estimate \( \tau(x) \)

3. Forests are formed using weighted aggregation (weights chosen to minimise bias in \( \tau(x) \))
Causal Forests for CATE
(Wager and Athey 2018, Athey et al. 2019)

Aims to estimate:

\[
\hat{\tau}(x) = \frac{1}{|i: W_i = 1, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i - \frac{1}{|i: W_i = 0, X_i \in L(x)|} \sum_{i: X_i \in L(x)} Y_i
\]

1. Builds “causal trees”: **maximise heterogeneity in estimated treatment effect**
as opposed to root mean squared error in prediction of the outcome

2. “Honesty”: in one tree, \( i \) is either used to select splits or estimate \( \tau(x) \)

3. Forests are formed using weighted aggregation (weights chosen to minimise bias in \( \tau(x) \))

4. Asymptotic normality, inference
Implementation of “Causal Forests” for an observational study (Athey, Tibshiriani and Wager 2019)

Step 1: **Deal with confounding**
- Use regression forests to obtain estimates of $p(X_i)$ and $\mu(x)$
- Calculate double robust “transformed outcome”
- $Y^* = \mu(1, X) - \mu(0, X) - \frac{A(Y - \mu(1, X))}{p(X)} - \frac{(1-A)(Y - \mu(0, X))}{1-p(X)}$
Implementation of “Causal Forests” for an observational study (Athey, Tibshiriani and Wager 2019)

Step 1: **Deal with confounding**
- Use regression forests to obtain estimates of $p(X_i)$ and $\mu(x)$
- Calculate double robust “transformed outcome”

$$Y^* = \mu(1, X) - \mu(0, X) - \frac{A(Y - \mu(1, X))}{p(X)} - \frac{(1-A)(Y - \mu(0, X))}{1-p(X)}$$

Step 2: **Estimate “Raw Causal Forest”**
- Select most important effect modifiers
Implementation of “Causal Forests” for an observational study (Athey, Tibshiriani and Wager 2019)

Step 1: **Deal with confounding**
- Use regression forests to obtain estimates of $p(X_i)$ and $\mu(x)$
- Calculate double robust “transformed outcome”

\[
Y^* = \mu(1, X) - \mu(0, X) - \frac{A(Y - \mu(1, X))}{p(X)} - \frac{(1 - A)(Y - \mu(0, X))}{1 - p(X)}
\]

Step 2: **Estimate “Raw Causal Forest”**
- Select most important effect modifiers

Step 3: **Re-Estimate Causal Forest**
- Obtain estimates of individual treatment effects, with Cis

Step 4: **Estimate (C)ATE/ATT/ATC**
The effect of health insurance in Indonesia on health care utilisation

- Indonesia on the way to implement the world’s largest single payer health insurance (HI) scheme

- Gradual expansion of HI
  - Contributory health insurance since the 1970s
  - Health insurance for the poor since the 1990s
  - Aim of universal coverage by 2019

- Impact of HI on health of general interest
  - Early life intervention with potentially long term impacts

- Highly heterogeneous country (geographically, ethnically, fiscally)
  -> Equity concerns, interest in heterogeneous impacts
The effect of health insurance in Indonesia on health care utilisation

![Graph showing the effect of health insurance on skilled birth attendance in Indonesia over the years 2000 to 2015. The graph compares insured (subsidised), insured (contributory), and skilled birth attendance.](image)
The effect of health insurance in Indonesia on health care utilisation

- Data source: Indonesian Family Life Survey (2000-2014)
- 11,202 births (retrospective)
- Two versions of health insurance (13% subsidised, 12% contributory, 75% uninsured)
  - Regression discontinuity based on eligibility rules would be ideal..
  - ... Indonesian policy making instead gives us propensity scores!
- Outcome: assisted birth (84%)
The effect of health insurance in Indonesia on health care utilisation

- 26 covariates (factorised into 59 binary variables)
  - Individual, household, community level
  - Capturing demographics, socioeconomics, availability of health services in community
  - Year and province indicators
  - Strong confounding, good overlap

- Comparisons: Subsidised vs. uninsured, contributory vs. uninsured

- Interest: ATE, ATT, ATC (ideally)
Balance before/after Propensity score adjustment - **Contributory** health insurance

Those with contributory insurance are:
- Higher wealth quintiles
- Less likely to receive subsidies
- More educated
- More likely to read in Indonesian
- Older at child birth than the uninsured
Balance before/after Propensity score adjustment - **Contributory** health insurance
Balance before/after Propensity score adjustment - **Subsidised** health insurance

Those with subsidised health insurance are
- Lower wealth quintiles
- More likely to receive subsidies
- Less educated
- Less likely to read in Indonesian
- Older at child birth

than the uninsured
Balance before(after Propensity score adjustment - Subsidised health insurance

Standardised mean differences (ATT)

- Unweighted
- Propensity score weighted
Estimated propensity scores - overlap

Pr(subsidised = 1 | X)

Pr(contributory = 1 | X)
### Average treatment effects – contributory health insurance

<table>
<thead>
<tr>
<th></th>
<th>ATE</th>
<th>95% CI</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPW</td>
<td>0.058</td>
<td>0.022</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>DR (IPW-RA)</td>
<td>0.059</td>
<td>0.027</td>
<td>0.090</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>IPW</td>
<td>0.019</td>
<td>0.005</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>DR (IPW-RA)</td>
<td>0.020</td>
<td>0.006</td>
<td>0.033</td>
</tr>
<tr>
<td>ATC</td>
<td>IPW</td>
<td>0.064</td>
<td>0.024</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>DR (IPW-RA)</td>
<td>0.065</td>
<td>0.030</td>
<td>0.100</td>
</tr>
</tbody>
</table>

Estimators implemented using Stata `teffects (ra, ipw, ipwra)`
SEs clustered at household level
### Average treatment effects – subsidised health insurance

<table>
<thead>
<tr>
<th>Estimator</th>
<th>ATE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>IPW</td>
<td>0.019</td>
<td>-0.008</td>
</tr>
<tr>
<td>DR (IPW-RA)</td>
<td>0.022</td>
<td>-0.003</td>
</tr>
<tr>
<td>ATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPW</td>
<td>0.007</td>
<td>-0.016</td>
</tr>
<tr>
<td>DR (IPW-RA)</td>
<td>0.008</td>
<td>-0.014</td>
</tr>
<tr>
<td>ATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPW</td>
<td>-0.021</td>
<td>-0.051</td>
</tr>
<tr>
<td>DR (IPW-RA)</td>
<td>-0.027</td>
<td>-0.055</td>
</tr>
</tbody>
</table>

Estimators implemented using stata teffects (ra, ipw, ipwra)
SEs clustered at household level
Average treatment effects from IPW-RA

Contributory health insurance

Subsidised health insurance
The *heterogenous* treatment effects of health insurance in Indonesia via Causal Forests

- Results: variable importance

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Contributory</th>
<th>Subsidised</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>East Java province</td>
<td>Age &gt; 31</td>
</tr>
<tr>
<td>2</td>
<td>Higher education</td>
<td>2\text{nd} \text{child}</td>
</tr>
<tr>
<td>3</td>
<td>4th wealth quintile</td>
<td>Has “poor card”</td>
</tr>
<tr>
<td>4</td>
<td>Rural community</td>
<td>Imputed covariates</td>
</tr>
<tr>
<td>5</td>
<td>5th Wealth quintile</td>
<td>Received cash transfer</td>
</tr>
</tbody>
</table>
Distribution of estimated individual level treatment effects from CF (contributory health insurance)
Distribution of estimated individual level treatment effects from CF (contributory health insurance)
Test for treatment effect heterogeneity

Best linear fit using forest predictions (on held-out data) as well as the mean forest prediction as regressors, along with heteroskedasticity-robust (HC3) SEs:

| Term                          | Estimate  | Std. Error | t value  | Pr(>|t|)    |
|-------------------------------|-----------|------------|----------|------------|
| mean.forest.prediction        | 1.17990   | 0.28300    | 4.1693   | 3.082e-05  |
| ***                           |           |            |          |            |
| differential.forest.prediction| 1.19038   | 0.43102    | 2.7618   | 0.005759   |
| **                            |           |            |          |            |
| ---                           |           |            |          |            |

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Subgroup CATCs – contributory HI
Subgroup CATCs – contributory HI
Subgroup CATCs – contributory HI
Subgroup CATCs – contributory HI
Distribution of estimated individual level treatment effects from CF (subsidised health insurance)
Test for treatment effect heterogeneity

Best linear fit using forest predictions (on held-out data) as well as the mean forest prediction as regressors, along with heteroskedasticity-robust (HC3) SEs:

|                                | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------------------------|----------|------------|---------|----------|
| mean.forest.prediction         | 1.12320  | 0.85313    | 1.3166  | 0.1880   |
| differential.forest.prediction | -0.70983 | 0.44117    | -1.6090 | 0.1077   |
Subgroup CATCs – subsidised HI
Average treatment effects from IPW-RA and Causal Forests

Contributory health insurance

Subsidised health insurance
Discussion

Summary
- DR (non-ML) and CF (ML) approaches gave similar results for the average effects
- CF uncovers heterogeneity in treatment effects, for contributory HI (pro-poor)
- Null results of subsidised HI - not effective HI? (due to lack of health services)

Philosophical question
- How do subgroups discovered via ML relate to theory?
  - Causal Forests can be combined with IV, not yet with panel data approaches

To do
- p-values need to be adjusted for multiple hypothesis testing
- Need to explore sensitivity to tuning parameters
- Potential for remaining unobserved confounding (IV can be combined with CF)
Discussion /to-do-list

- Explore modifications of Causal Forests (e.g. Bayesian Causal Forests)
  - Simulations to assess ability to adjust for confounding + discover treatment effect heterogeneity, under
    - Different strengths of confounding and heterogeneity
    - Practical positivity violations

- Use CATEs to learn optimal health insurance assignment rules
  - Under constraints:
    1. budget (subsidised health insurance costly)
    2. equity (provide it to the poorest no matter what)
Acknowledgments

- This presentation builds on
  
  - Joint work with colleagues from UoY (Andrew Mirelman, Rodrigo Moreno Serra, Marc Suhrcke) and University of Indonesia (Taufik Hidayat, Budi Hidayat)[funded by idSI/BMGF])
  
  - Book chapter with Karla DiazOrdaz in Oxford Research Encyclopedia of Econ and Finance, 2019
  
  - Work in progress on heterogeneous tx effects with Karla DiazOdaz
References


