Heterogeneous Treatment Effects of Medicaid and Efficient Policies

Shishir Shakya

John Chambers College of Business and Economics
West Virginia University

June 4, 2020

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.
- Medicaid for **non-categorically** eligible population: able-bodied adults within 138% of Federal Poverty Line

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.
- Medicaid for **non-categorically** eligible population: able-bodied adults within 138% of Federal Poverty Line
- Federal and state government jointly fund medicaid, it cost \$8000 per person per year.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.
- Medicaid for **non-categorically** eligible population: able-bodied adults within 138% of Federal Poverty Line
- Federal and state government jointly fund medicaid, it cost $8000 per person per year.
- State are suffering budget hole.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.
- Medicaid for **non-categorically** eligible population: able-bodied adults within 138% of Federal Poverty Line
- Federal and state government jointly fund medicaid, it cost \$8000 per person per year.
- State are suffering budget hole.
- Policy makers reform Medicaid with Waivers and plan amendments.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.
- Medicaid for **non-categorically** eligible population: able-bodied adults within 138% of Federal Poverty Line
- Federal and state government jointly fund medicaid, it cost $8000 per person per year.
- State are suffering budget hole.
- Policy makers reform Medicaid with Waivers and plan amendments.
- These waivers/reforms **tweak** the pre-existing eligibility criteria.

Introduction

- Medicaid is health insurance for low-income American and 1 in 5 are covered with Medicaid.
- Public in US have favorable opinions toward Medicaid.
- Medicaid for **categorically** eligible population: welfare recipients, low income pregnant women, children and parents.
- Medicaid for **non-categorically** eligible population: able-bodied adults within 138% of Federal Poverty Line
- Federal and state government jointly fund medicaid, it cost $8000 per person per year.
- State are suffering budget hole.
- Policy makers reform Medicaid with Waivers and plan amendments.
- These waivers/reforms **tweak** the pre-existing eligibility criteria.
- Setting eligibility criteria = **segmentation and targeting the population**.

Research Questions

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?
- **Data**: Oregon Health Insurance Experiment (OHIE) – random assignment (lottery) of Medicaid – which is likely to circumvents selection bias.

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?
- **Data**: Oregon Health Insurance Experiment (OHIE) – random assignment (lottery) of Medicaid – which is likely to circumvents selection bias.
- **Dependent/outcome variables**:
  - **Health care utilization**

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?
- **Data**: Oregon Health Insurance Experiment (OHIE) – random assignment (lottery) of Medicaid – which is likely to circumvents selection bias.
- **Dependent/outcome variables**:
  - **Health care utilization**, Preventive care utilization, Financial strains, Self-reported health and well beings, and several other mechanisms.

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?
- **Data**: Oregon Health Insurance Experiment (OHIE) – random assignment (lottery) of Medicaid – which is likely to circumvents selection bias.
- **Dependent/outcome variables**:
  - **Health care utilization**, Preventive care utilization, Financial strains, Self-reported health and well beings, and several other mechanisms.
- **Treatment variable**: Random assignment of Medicaid.

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?
- **Data**: Oregon Health Insurance Experiment (OHIE) – random assignment (lottery) of Medicaid – which is likely to circumvents selection bias.
- **Dependent/outcome variables**:
  - **Health care utilization**, Preventive care utilization, Financial strains, Self-reported health and well beings, and several other mechanisms.
- **Treatment variable**: Random assignment of Medicaid.
- **Control variables**: Federal Poverty Level, Age, Household size, Insurance status of last 6-months, Education status, Employment status.

Research Questions

- **Main question**: What are the efficient Medicaid eligibility criterion?
- **Auxiliary question**: What are the heterogeneous treatment effects of Medicaid?
- **Data**: Oregon Health Insurance Experiment (OHIE) – random assignment (lottery) of Medicaid – which is likely to circumvents selection bias.
- **Dependent/outcome variables**:
  - **Health care utilization**, Preventive care utilization, Financial strains, Self-reported health and well beings, and several other mechanisms.
- **Treatment variable**: Random assignment of Medicaid.
- **Control variables**: Federal Poverty Level, Age, Household size, Insurance status of last 6-months, Education status, Employment status.
- **Drop** the race, gender and residency for non-discriminatory policy/waiver purpose.

Oregon Health Insurance Experiment

Oregon Health Insurance Experiment

- Oregon's Medicaid in 2000
  - **OHP Standard**: for non-categorically eligible Medicaid population.

Oregon Health Insurance Experiment

- Oregon's Medicaid in 2000
  - **OHP Standard**: for non-categorically eligible Medicaid population.

    Adults ages 19−64.
    who are Oregon residents and U.S. citizens or legal immigrants.
    and have incomes below the 100% federal poverty level.

Oregon Health Insurance Experiment

- Oregon's Medicaid in 2000
  - **OHP Standard**: for non-categorically eligible Medicaid population.

    Adults ages 19−64.
    who are Oregon residents and U.S. citizens or legal immigrants.
    and have incomes below the 100% federal poverty level.
    and/or who have been without health insurance for at least six months,
    and/or have less than $2,000 in assets.

Oregon Health Insurance Experiment

- Oregon's Medicaid in 2000
  - **OHP Standard**: for non-categorically eligible Medicaid population.

    Adults ages 19−64.
    who are Oregon residents and U.S. citizens or legal immigrants.
    and have incomes below the 100% federal poverty level.
    and/or who have been without health insurance for at least six months,
    and/or have less than $2,000 in assets.
- OHP Standard covers doctors, hospitals, drugs, mental health, etc., with no consumer cost-sharing and low or no premiums.

Oregon Health Insurance Experiment

- OHP Standard was started in 2000 and terminated in 2004 due to budgetary shortfall.

Oregon Health Insurance Experiment

- OHP Standard was started in 2000 and terminated in 2004 due to budgetary shortfall.
- In 2008, funds for 10,000 OHP Standard were available, but about 90,000 Oregonians were non-categorically eligible.

Oregon Health Insurance Experiment

- OHP Standard was started in 2000 and terminated in 2004 due to budgetary shortfall.
- In 2008, funds for 10,000 OHP Standard were available, but about 90,000 Oregonians were non-categorically eligible.
- DHS got approved to lottery/randomize OHP standard from CMS.

Oregon Health Insurance Experiment

- OHP Standard was started in 2000 and terminated in 2004 due to budgetary shortfall.
- In 2008, funds for 10,000 OHP Standard were available, but about 90,000 Oregonians were non-categorically eligible.
- DHS got approved to lottery/randomize OHP standard from CMS.
- Data comprises 74,922 individuals (representing 66,385 households).

Oregon Health Insurance Experiment

- OHP Standard was started in 2000 and terminated in 2004 due to budgetary shortfall.
- In 2008, funds for 10,000 OHP Standard were available, but about 90,000 Oregonians were non-categorically eligible.
- DHS got approved to lottery/randomize OHP standard from CMS.
- Data comprises 74,922 individuals (representing 66,385 households).
- At baseline, data on lottery (W) and various demographics (x) were collected, then after a year, data on outcome variable (Y) were collected.

Heterogeneous Treatment Effects: Why and How?

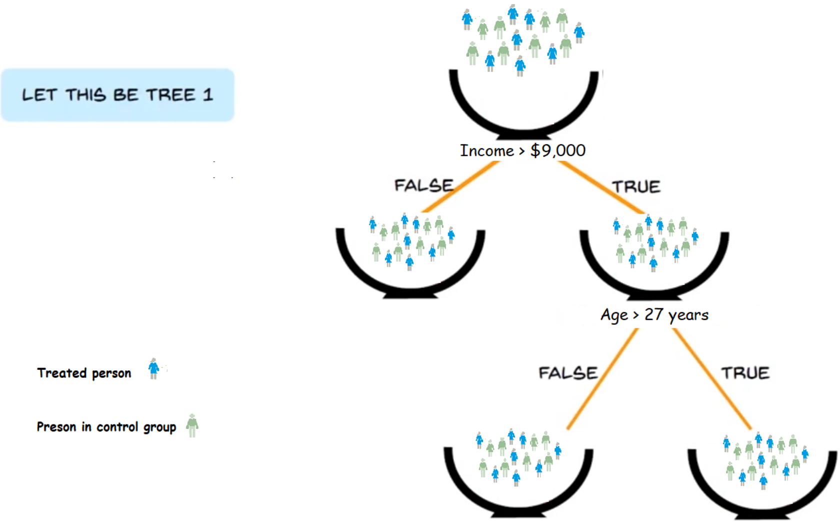
Why Heterogeneous Treatment Effects?

- Mixed effects of Medicaid $\Longrightarrow$ gap in literature
- Systematically identify subpopulations and estimate treatment effects, valid inference.
- HTEs are mandatory step to understand mechanisms and identify efficient policy.

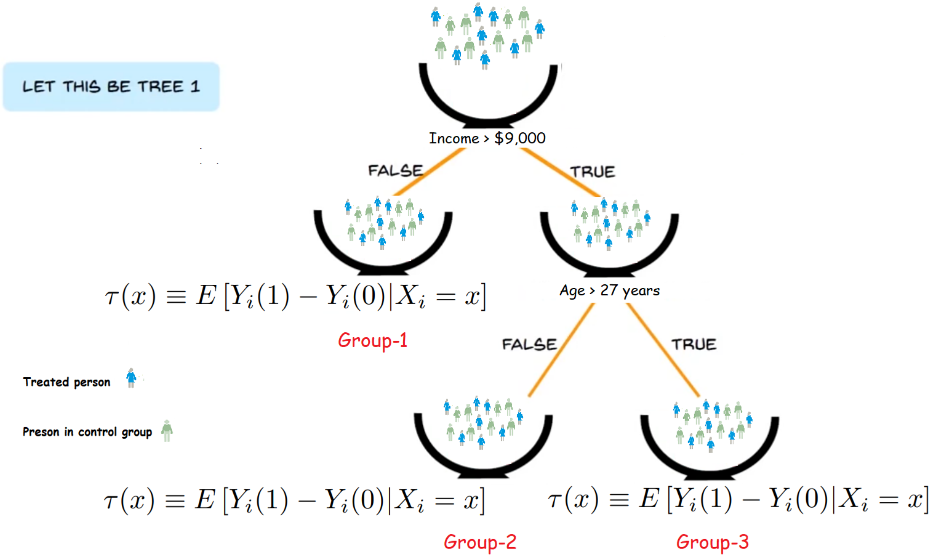
Causal ML: Causal Tree (Athey and Imbens, 2016)

Causal Tree (Athey and Imbens, 2016)

- Recursively classifying data into sub-groups that have similar treatment effect τ.



Causal Tree (Athey and Imbens, 2016)



$$\tau(x) \equiv E\left[Y_i(1) - Y_i(0) | X_i = x\right]$$

$$\tau(x) \equiv E\left[Y_i(1) - Y_i(0) | X_i = x\right]$$

$$\tau(x) \equiv E\left[Y_i(1) - Y_i(0) | X_i = x\right]$$

Causal Tree (Athey and Imbens, 2016)

- Recursively classifying data into sub-groups that have similar treatment effect τ.
- Goal: Minimize *MSE* of treatment effects on each leaf i.e.

$$-E_{S^{tr}}\left[\sum_{i \in S^{tr}} \left(\tau_i - \hat{\tau}(X_i)\right)^2\right]$$

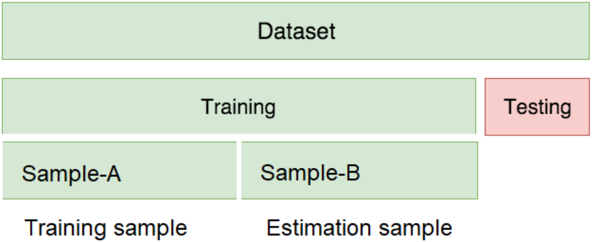
- Cross-validate to prune the tree-depths
- Problem: this is infeasible (τ_i or the true treatment effect unobserved)

Causal Tree (Athey and Imbens, 2016)

- Recursively classifying data into sub-groups that have similar treatment effect τ.
- Goal: Minimize *MSE* of treatment effects on each leaf i.e.

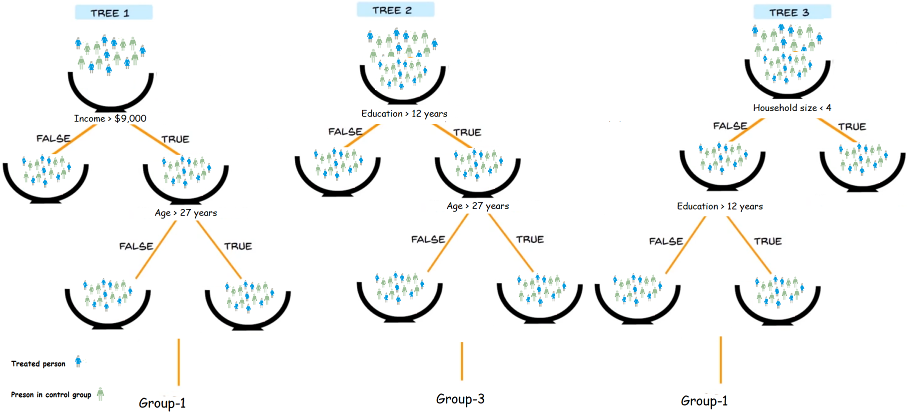
$$-E_{S^{tr}}\left[\sum_{i\in S^{tr}}(\tau_i-\hat{\tau}(X_i))^2\right]$$

- Cross-validate to prune the tree-depths
- Problem: this is infeasible (τ_i or the true treatment effect unobserved)
- Sample splitting and Honest estimation.



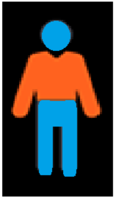
Causal ML: Causal Forest $=$ Many Causal Trees

Causal Forest = Many Causal Trees



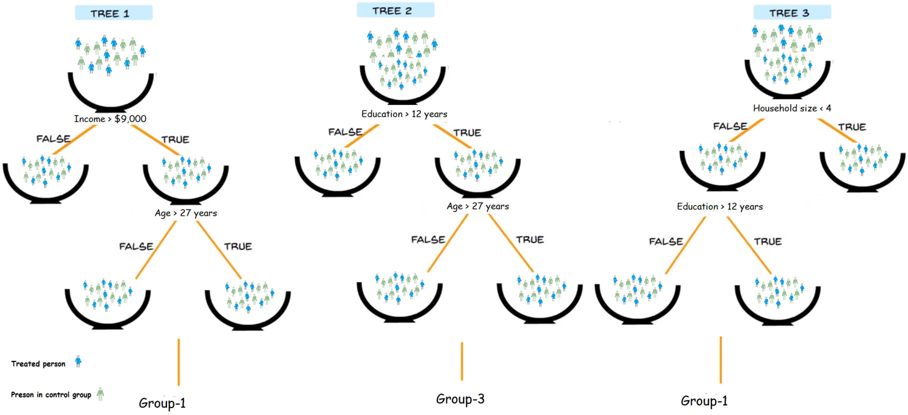
Out-of-Sample Prediction

NOW LETS TRY TO
CLASSIFY THIS

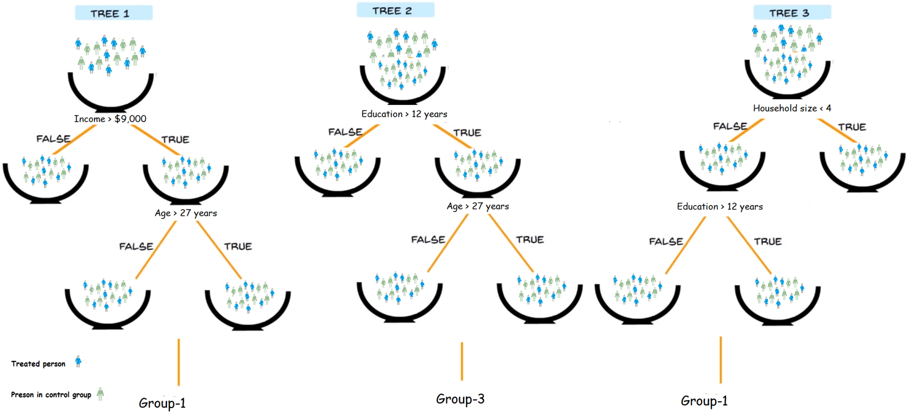


Age = 30
HH income = $12000
HH size = 5
Education = 12

Out-of-Sample Prediction



Out-of-Sample Prediction



$$\hat{\Gamma}_i = \hat{\tau}_{-i}(X_i) + \frac{1\{W_i = \pi(X_i)\}}{\hat{e}_{-i}(X_i, W_i)} \cdot (Y_i - \hat{\mu}_{-i}(X_i, W_i))$$

Standard errors are Cluster-robust at household level.

Establish asymptotic normality: $\frac{\hat{\tau}(x) - \tau(x)}{\sigma(x)} \Rightarrow N(0, 1)$.

Bootstrap Aggregation (Bagging) | Honest Estimation | Cross-fitting | R-Learner | Causal Random Forest

Results: Part 1 of 3

Results

Table 2: Health Care Utilization

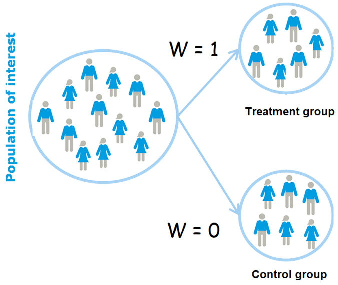
Outcome variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Panel A: Health care utilization						
Extensive margins						
Currently taking any prescription medications	0.021**	0.067**	0.007	-0.018	0.801	-0.494
	(0.009)	(0.03)	(0.009)	(0.018)	(1.015)	(0.734)
Outpatient visits last six months	0.07***	0.224***	0.062***	0.055***	1.028***	1.316***
	(0.009)	(0.027)	(0.009)	(0.017)	(0.145)	(0.312)
ER visits last six months	0.009	0.029	0.005	-0.014	0.696	-3.331
	(0.008)	(0.024)	(0.008)	(0.015)	(1.172)	(1.816)
Inpatient hospital admissions last six months	0.002	0.005	0.001	-0.006	0.272	-0.626
	(0.004)	(0.014)	(0.005)	(0.009)	(2.322)	(1.4)
Intensive margins						
Number of prescription medications currently taking	0.104*	0.342*	0.042	-0.119	0.899	-0.383
	(0.055)	(0.177)	(0.055)	(0.109)	(1.219)	(1.005)
Number of Outpatient visits last six months	0.335***	1.087***	0.304***	0.426***	1.037***	1.502***
	(0.052)	(0.166)	(0.055)	(0.11)	(0.188)	(0.373)
Number of ER visits last six months	0.006	0.018	-0.003	-0.115***	1.97	-10.89
	(0.016)	(0.053)	(0.017)	(0.035)	(14.846)	(2.98)
Number Inpatient hospital admissions last six months	0.007	0.024	0.007	0.008	0.713	-2.071
	(0.007)	(0.021)	(0.007)	(0.014)	(0.661)	(1.974)

Results

Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)
Outpatient visits last six months	0.07*** (0.009)	0.224*** (0.027)	0.062*** (0.009)

$$\text{Outpatient}_{i,h} = \beta_0 + \beta_1 \text{Lottery}_{i,h} + x_{ih}\beta_2 + \varepsilon_{it}$$



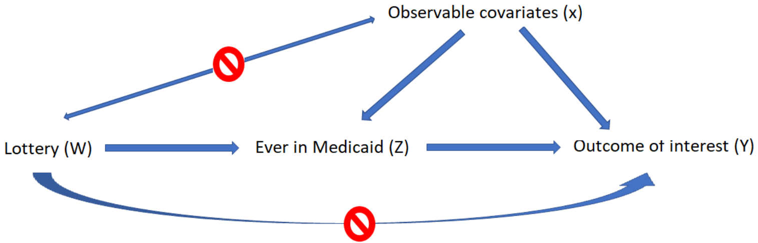
Results

Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)
Outpatient visits last six months	0.07***	0.224***	0.062***
	(0.009)	(0.027)	(0.009)

$$Medicaid_{i,h} = \delta_0 + \delta_1 \text{Lottery}_{i,h} + x_{ih}\delta_2 + \mu_{it}$$

$$\text{Outpatient}_{i,h} = \phi_0 + \phi_1 \widehat{Medicaid}_{i,h} + x_{ih}\phi_2 + \nu_{it}$$



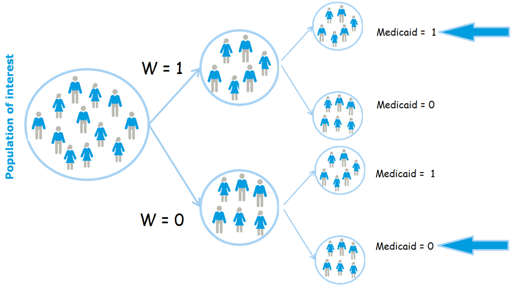
Results

Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)
Outpatient visits last six months	0.07*** (0.009)	0.224*** (0.027)	0.062*** (0.009)

$$Medicaid_{i,h} = \delta_0 + \delta_1 \text{Lottery}_{i,h} + x_{ih}\delta_2 + \mu_{it}$$

$$\text{Outpatient}_{i,h} = \phi_0 + \phi_1 \widehat{Medicaid}_{i,h} + x_{ih}\phi_2 + \nu_{it}$$

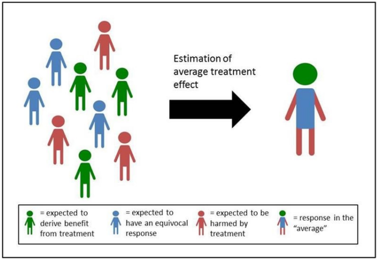


Results

Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)
Outpatient visits last six months	0.07***	0.224***	0.062***
	(0.009)	(0.027)	(0.009)

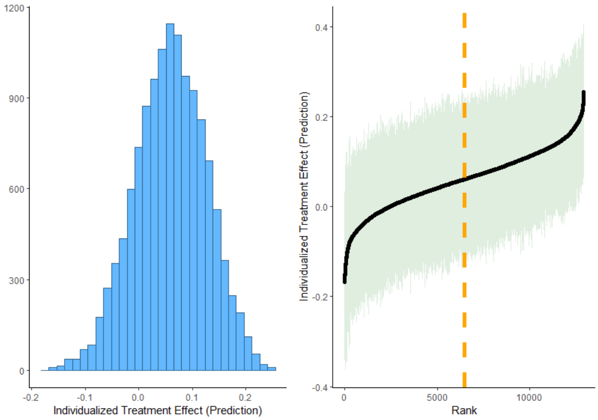
A Average Treatment Effect Assessed in a Heterogeneous Population



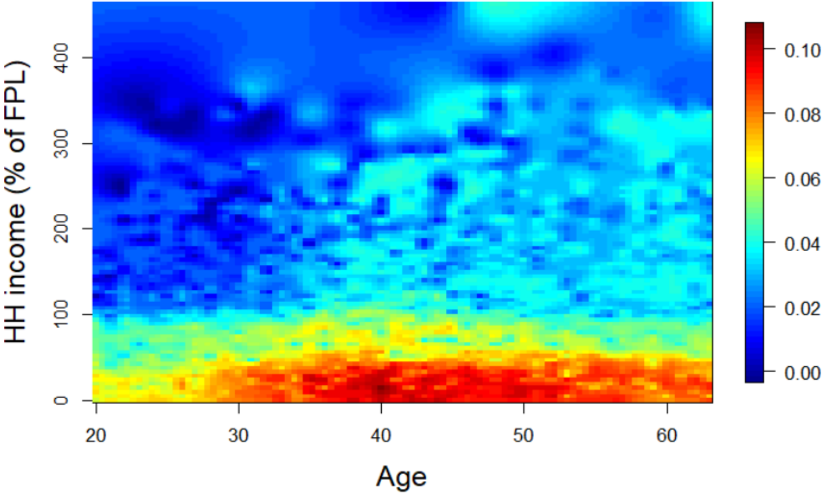
Heuristic-test Mean Forest Prediction Differential Forest Prediction

Tests for Heterogeneity.

Heterogeneity: Graphical Intuition



Heterogeneity: Graphical Intuition



Outpatient visits last six months

Results: Part 2 of 3

Accessing Treatment Heterogeneity

Table 2: Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Outpatient visits last six months	0.07***	0.224***	0.062***	0.055***	1.028***	1.316***
	(0.009)	(0.027)	(0.009)	(0.017)	(0.145)	(0.312)

Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The regressions in Columns (1) and (2) include household size dummies, survey wave dummies, and survey wave interacted with household size dummies. For the LATE estimates in Column (2), the instrumental variable is lottery assignment, and the endogenous variable is "Ever in Medicaid". The ITT and LATE estimates are base on the double-selection post-LASSO.

- Measure of Treatment Heterogeneity: Heuristic*, $MFP = 1$, and $DFP > 0$.
- Once we know there is treatment heterogeneity, we would like to know efficient policy that allocates scares resources.

Efficient Policy Learning

Efficient Policy Learning

- Efficient policy $\Longrightarrow$ cost $\downarrow$ & health outcomes $\uparrow$.

Efficient Policy Learning

- Efficient policy $\Longrightarrow$ cost $\downarrow$ & health outcomes $\uparrow$.
- Find policy (tree) $\pi : \chi \rightarrow W$ that maximize expected outcomes (welfare).

Efficient Policy Learning

- Efficient policy $\Longrightarrow$ cost $\downarrow$ & health outcomes $\uparrow$.
- Find policy (tree) $\pi : \chi \rightarrow W$ that maximize expected outcomes (welfare).

$$\pi^* \in \arg \max_{\pi \in \prod} E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

Efficient Policy Learning

- Efficient policy $\Longrightarrow$ cost $\downarrow$ & health outcomes $\uparrow$.
- Find policy (tree) $\pi : \chi \rightarrow W$ that maximize expected outcomes (welfare).

$$\pi^* \in \arg \max_{\pi \in \prod} E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

- Any other non-optimal policy leads to less welfare or regret $R\left(\pi\right)$.

Efficient Policy Learning

- Efficient policy $\Longrightarrow$ cost $\downarrow$ & health outcomes $\uparrow$.
- Find policy (tree) $\pi : \chi \rightarrow W$ that maximize expected outcomes (welfare).

$$\pi^* \in \arg \max_{\pi \in \prod} E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

- Any other non-optimal policy leads to less welfare or regret $R(\pi)$.

$$R(\pi) = E\left[Y_i\left(\pi^*\left(X_i\right)\right)\right] - E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

Efficient Policy Learning

- Efficient policy $\Longrightarrow$ cost $\downarrow$ & health outcomes $\uparrow$.
- Find policy (tree) $\pi : \chi \rightarrow W$ that maximize expected outcomes (welfare).

$$\pi^* \in \arg \max_{\pi \in \prod} E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

- Any other non-optimal policy leads to less welfare or regret $R(\pi)$.

$$R(\pi) = E\left[Y_i\left(\pi^*\left(X_i\right)\right)\right] - E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

- We would like to minimize the regret.

Efficient Policy Learning

- Estimate regret by Q-learning:

$$\hat{Q}(\pi) = n^{-1} \sum_i \pi(X_i) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

Efficient Policy Learning

- Estimate regret by Q-learning:

$$\hat{Q}(\pi) = n^{-1} \sum_{i} \pi(X_i) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

  - Given a policy $\pi(X_i)$,

Efficient Policy Learning

- Estimate regret by Q-learning:

$$\hat{Q}(\pi) = n^{-1} \sum_i \pi(X_i) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

  - Given a policy $\pi(X_i)$, we generate score $\hat{\Gamma}_i$.

Efficient Policy Learning

- Estimate regret by Q-learning:

$$\hat{Q}(\pi) = n^{-1} \sum_{i} \pi(X_i) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

  - Given a policy $\pi(X_i)$, we generate score $\hat{\Gamma}_i$.
  - The $sign\left(\hat{\Gamma}_i\right)$ show whether to intervene or not.

Efficient Policy Learning

- Estimate regret by Q-learning:

$$\hat{Q}(\pi) = n^{-1} \sum_i \pi(X_i) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

  - Given a policy $\pi(X_i)$, we generate score $\hat{\Gamma}_i$.
  - The $sign\left(\hat{\Gamma}_i\right)$ show whether to intervene or not.
  - The magnitude show $|\hat{\Gamma}_i|$ show much/less we wish to intervene.

Efficient Policy Learning

- Estimate regret by Q-learning:

$$\hat{Q}(\pi) = n^{-1} \sum_{i} \pi(X_i) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

  - Given a policy $\pi(X_i)$, we generate score $\hat{\Gamma}_i$.
  - The $sign\left(\hat{\Gamma}_i\right)$ show whether to intervene or not.
  - The magnitude show $|\hat{\Gamma}_i|$ show much/less we wish to intervene.
- The regret converges as:

$$\sqrt{n}\left(\hat{R}_{DML}(\pi) - R(\pi)\right) \xrightarrow{d} N\left(0, \sigma^2(\pi)\right)$$

Double Machine Learning VC Dimension

Results: Part 3 of 3

Results

Estimate of the utility improvement of various policies over a random assignment baseline.

Variable	Baseline (1)	Probability rule (2)	CATE rule (3)	Shallow tree (4)	Deeper tree (5)
Panel A: Health care utilization					
Outpatient visits last six months	0.604*** (0.002)	4.74*** (0.182)	5.119*** (0.17)	4.228*** (0.197)	2.898*** (0.177)

- **Baseline**: On average 60% of population have outpatient visit in last 6 month.

Results

Estimate of the utility improvement of various policies over a random assignment baseline.

Variable	Baseline (1)	Probability rule (2)	CATE rule (3)	Shallow tree (4)	Deeper tree (5)
Panel A: Health care utilization					
Outpatient visits	0.604***	4.74***	5.119***	4.228***	2.898***
last six months	(0.002)	(0.182)	(0.17)	(0.197)	(0.177)

Probability rule: If the propensity is below average, assign the treatment.

$$\hat{\Gamma}_i = \hat{e}_i^{(-i)}(X_i, W_i)$$

Results

Estimate of the utility improvement of various policies over a random assignment baseline.

Variable	Baseline (1)	Probability rule (2)	CATE rule (3)	Shallow tree (4)	Deeper tree (5)
Panel A: Health care utilization					
Outpatient visits last six months	0.604*** (0.002)	4.74*** (0.182)	5.119*** (0.17)	4.228*** (0.197)	2.898*** (0.177)

CATE rule: If the CATE is non-zero and positive, assign the treatment.

$$\hat{\Gamma}_i = \hat{\tau}_i^{(-i)}(X_i)$$

Results

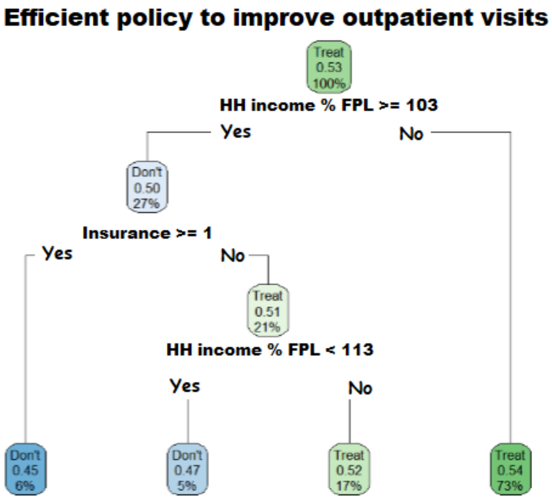
Estimate of the utility improvement of various policies over a random assignment baseline.

Variable	Baseline (1)	Probability rule (2)	CATE rule (3)	Shallow tree (4)	Deeper tree (5)
Panel A: Health care utilization					
Outpatient visits	0.604***	4.74***	5.119***	4.228***	2.898***
last six months	(0.002)	(0.182)	(0.17)	(0.197)	(0.177)

Shallow tree: treatment assignment rule base on a simpler tree (2-Level depths).

Deeper tree: treatment assignment rule base on complex or more in-depth tree (up to 6-Level depths).

$$\hat{\Gamma}_i = \hat{\tau}_{-i}(X_i) + \frac{1\{W_i = \pi(X_i)\}}{\hat{e}_{-i}(X_i, W_i)} \cdot (Y_i - \hat{\mu}_{-i}(X_i, W_i))$$



Matching story and policy. VC Dimension

Discussion and Conclusion

Discussion

Policy evaluation.

- Positive economics
- "What is" the impact of treatment on outcome?
- Random assignment of treatment.
- Ex-post causal claims.

Policy design/recommendation.

- Normative economics.
- "Who should" get the treatment?
- A blanket policy focusing on outcomes.
- Ex-ante causal claims.

Discussion

Policy evaluation.

- Positive economics
- "What is" the impact of treatment on outcome?
- Random assignment of treatment.
- Ex-post causal claims.
- At least, properly control the observable, like: race, gender, residency, etc.
- Complex modeling

Policy design/recommendation.

- Normative economics.
- "Who should" get the treatment?
- A blanket policy focusing on outcomes.
- Ex-ante causal claims.
- Non-discriminatory, politically correct, administratively/legally feasible.
- Simple or thumb-rule to implement.

Discussion

Policy evaluation.

- Positive economics
- "What is" the impact of treatment on outcome?
- Random assignment of treatment.
- Ex-post causal claims.
- At least, properly control the observable, like: race, gender, residency, etc.
- Complex modeling
- Local Average Treatment Effect (LATE) – effect among compliers.

Policy design/recommendation.

- Normative economics.
- "Who should" get the treatment?
- A blanket policy focusing on outcomes.
- Ex-ante causal claims.
- Non-discriminatory, politically correct, administratively/legally feasible.
- Simple or thumb-rule to implement.
- Intent to Treat (ITT) & consumer sovereignty.

Discussion

Policy evaluation.

- Positive economics
- "What is" the impact of treatment on outcome?
- Random assignment of treatment.
- Ex-post causal claims.
- At least, properly control the observable, like: race, gender, residency, etc.
- Complex modeling
- Local Average Treatment Effect (LATE) – effect among compliers.
- Focus on treatment.

Policy design/recommendation.

- Normative economics.
- "Who should" get the treatment?
- A blanket policy focusing on outcomes.
- Ex-ante causal claims.
- Non-discriminatory, politically correct, administratively/legally feasible.
- Simple or thumb-rule to implement.
- Intent to Treat (ITT) & consumer sovereignty.
- Focus on welfare.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.
- Higher effect among impoverished households, suggesting standard adverse selection theory or maybe a pent up demand.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.
- Higher effect among impoverished households, suggesting standard adverse selection theory or maybe a pent up demand.
- Oregonian population $\neq$ the low-income U.S. adults.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.
- Higher effect among impoverished households, suggesting standard adverse selection theory or maybe a pent up demand.
- Oregonian population $\neq$ the low-income U.S. adults.
- Equally weighted observation to generalize.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.
- Higher effect among impoverished households, suggesting standard adverse selection theory or maybe a pent up demand.
- Oregonian population $\neq$ the low-income U.S. adults.
- Equally weighted observation to generalize.
- Intent-to-treat approach based on eligibility criteria.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.
- Higher effect among impoverished households, suggesting standard adverse selection theory or maybe a pent up demand.
- Oregonian population $\neq$ the low-income U.S. adults.
- Equally weighted observation to generalize.
- Intent-to-treat approach based on eligibility criteria.
- False Negative: Not assigning the Medicaid to those who need it.

Conclusion

- Proposed policies would improve selected outcomes by a range of 2%-9% against status-quo for out-of-sample. Policy Table
- One-year impact of expanding Medicaid access.
- Higher effect among impoverished households, suggesting standard adverse selection theory or maybe a pent up demand.
- Oregonian population $\neq$ the low-income U.S. adults.
- Equally weighted observation to generalize.
- Intent-to-treat approach based on eligibility criteria.
- False Negative: Not assigning the Medicaid to those who need it.
- Does this method work in an observational setting or panel or IV?

The End

Q&A

Email: ss0088@mix.wvu.edu

Rubin's Potential Outcome Framework back

We observe a sequence of triples $\{(W_i, Y_i, X_i)\}_i^N$ for N individuals, where:

- W_i represents if individual i is treated with lottery insurance or not.
- Y_i is the outcome variable.
- In Rubin (1974) potential outcomes framework,
  - $Y_i(1)$ represents potential outcome of subject i if he had received the treatment and
  - $Y_i(0)$ represents potential outcome of subject i if he had not had received the treatment
- X_i is vector of observable characteristics.
- Individual treatment effect for i^{th} is: $Y_i(1) - Y_i(0)$
- The average treatment effect (ATE) is: $\tau = E[Y_i(1) - Y_i(0)]$
- Unfortunately, in our data we of course can only observe one of these two potential outcomes, so actually computing this difference for everyone is impossible. This is **fundamental problem of causal inference**.

Rubin's Potential Outcome Framework back

The Naive estimator of average treatment effect (ATE) can be expressed as:

$$\tau = E\left[\tau\right] = E\left[Y_i\left(1\right) - Y_i\left(0\right)\right]$$

With linearity of expectation:

$$\tau = E\left[Y_i\left(1\right)\right] - E\left[Y_i\left(0\right)\right]$$

With independence assumption

$$\tau = E\left[Y_i\left(1\right)|W_i = 1\right] - E\left[Y_i\left(0\right)|W_i = 1\right]$$

Which is just the simple difference of means:

$$\hat{\tau} = \frac{1}{n_1}\sum_{i|W_i=1}^{n_1} y_i - \frac{1}{n_0}\sum_{i|W_i=0}^{n_0} y_i$$

item This is feasible only if there is no selection and heterogeneity biases.

Rubin's Potential Outcome Framework back

When, λ portion of population is exposed to treatment, the estimator of average treatment effect (ATE) can be expressed as:

$$\underbrace{E[\tau]}_{e} = \lambda \left\{ \underbrace{E[Y_i(1)|W_i = 1]}_{a} - \underbrace{E[Y_i(0)|W_i = 1]}_{c} \right\} - (1-\lambda) \left\{ \underbrace{E[Y_i(0)|W_i = 0]}_{d} - \underbrace{E[Y_i(1)|W_i = 0]}_{b} \right\}$$

where,

- a and d can be observed in from the data.
- c is counter factual of a.
- b is counter factual of d.
- λ is percentage of treated population.
- $Odom = (a - d)$ is Observed difference of mean.
- $(a - c)$ is Average treatment effect on treated (ATT).
- $(b - d)$ is Average treatment effect on untreated (ATU).

Rubin's Potential Outcome Framework back

The ATE can be expressed as:

$$e = \lambda (a - c) - (1 - \lambda)(d - b)$$

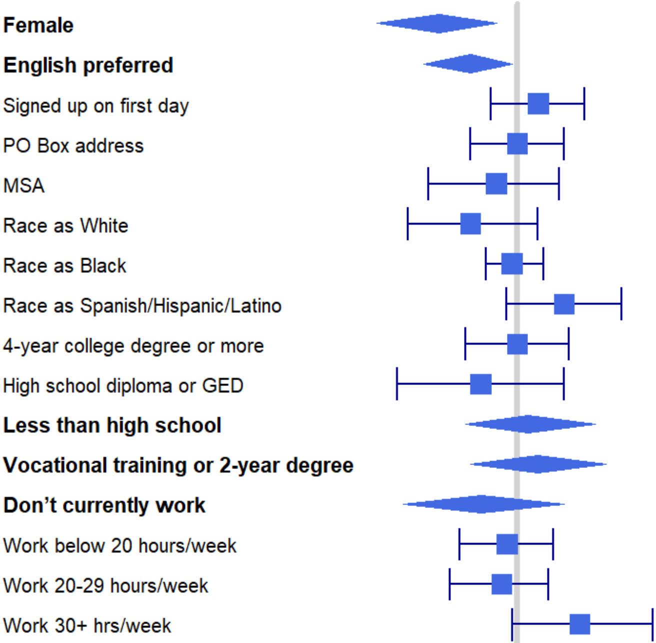
The solution can be expressed as:

$$\underbrace{(a - d)}_{Odom} = \underbrace{e}_{ATE} + \underbrace{(c - d)}_{selection\ \ bias} + \underbrace{(1 - \lambda)\{(a - c) - (b - d)\}}_{heterogeneous\ \ treatment\ \ effect\ \ bias}$$

- Under assumption of unconfoundedness[1], selection bias and heterogeneous treatment effect bias nullify and causal ATE is simple difference of mean. Technically: $Y_i(1), Y_i(0) \perp W_i | X_i$.
- The overlap assumption guarantees that no sub–population is entirely located in only one of control or treatment groups. Technically: $\forall x \in \text{supp}(X),\ 0 < P(W = 1 | X = x) < 1$.

[1]Unconfoundedness implies treatment is randomly assigned, and knowing observable characteristics of individual i, and then treatment status gives no information on the potential outcomes.

Pre-treatment Mean Comparison back



Double Selection Post-LASSO back

- LASSO simultaneously performs model selection and coefficient estimation by minimizing the sum of squared residuals plus a penalty term.
- Consider a following linear model

$$\tilde{y}_i = \Theta_i \beta_1 + \varepsilon_i$$

- where Θ is high-dimensional covariates, the LASSO estimator is defined as the solution to:

$$\min_{\beta_1 \in \mathbb{R}^p} E_n \left[(\tilde{y}_i -_i \Theta \beta_1)^2 \right] + \frac{\lambda}{n} \|\beta_1\|_1$$

- The penalty level λ is a tuning parameter to regularize/controls the degree of penalization and to guard against over-fitting. The penalty term penalizes the size of the model through the sum of absolute values of coefficients.
- The cross-validation technique chooses the best λ in prediction models and $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$.
- The kinked nature of penalty function induces $\hat{\beta}$ to have many zeros; thus LASSO solution feasible for model selection.

Double Selection Post-LASSO back

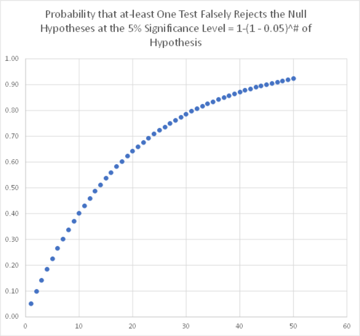
- The double-post-LASSO procedure comprises the following steps:
  - **First**, run LASSO of dependent variables on a large list of potential covariates to select a set of predictors for the dependent variable.
  - **Second**, run LASSO of treatment variable on a large list of potential covariates to select a set of predictors for treatment [2].
  - **Third**, run OLS regression of dependent variable on treatment variable, and the union of the sets of regressors selected in the two LASSO runs to estimate the effect of treatment on the dependent variable then correct the inference with usual heteroscedasticity robust OLS standard error.

[2] If the treatment is truly exogenous, I should expect this second step should not select any variables.

Why Causal-ML to estimate heterogeneous effects? back

- Problem of **the multiple hypothesis testing**[3].

# of Hypot hesis	least One Test Falsely Rejects the Null Hypotheses at the 5% Significance Level
1	0.05
2	0.10
3	0.14
4	0.19
5	0.23
6	0.26
7	0.30
8	0.34
9	0.37
10	0.40
11	0.43
12	0.46
13	0.49
14	0.51
15	0.54
16	0.56
17	0.58
18	0.60
19	0.62
20	0.64
21	0.66
22	0.68
23	0.69



[3] The "multiple hypothesis testing problems" leads to the so-called "ex-post selection problem," which is widely recognized in the program evaluation literature. For example, for fifty single hypotheses tests, the probability that at least one test falsely rejects the null hypotheses at the 5% significance level (assuming independent test statistics as an extreme case) is $1 - 0.95^{50} = 0.92$ or 92%.

Why Causal-ML to Estimate Heterogeneous Effects? back

- Performing ad-hoc searches or $p-$**hacking**[4] to detect the responsive subgroups may lead to false discoveries or may mistake noise for an actual treatment effect.
- To avoid many of the issues associated with data mining or $p-$hacking, researchers can commit in advance to study only a subgroup by **a preregistered analysis plan**[5]. However, this may also prevent discovering unanticipated results and developing new hypotheses.
- Using ML algorithms like CARTs, R.F., and N.N, etc. seems practical, but these algorithms are designed for prediction and not for causal inference.

[4] The $p-$hacking is an exhaustive search for statistically significant relations from combinations of variables or combinations of interactions of variables or subgroups. The $p-$hacking could lead to discovering the statistically significant relationship, when, in fact, there could have no real underlying effect.

[5] A preregistered analysis plan is sets of analyses plans released in the public domain by the researchers in advance prior they collect the data and learn about outcomes.

Athey and Imben Causal Tree back

Consider a tree or partitioning $\prod$ to a partition of feature space $\mathbb{X}$, with $\#\left(\prod\right)$ the numbers of elements in the partition given as:

$$\prod = \left\{\ell_1, \ldots, \ell_{\#(\tau)}\right\}$$

Then given a partition $\prod$, for each observations $\left(Y_i^{obs}, X_i, W_i\right)$, the population average condition mean function $\mu\left(x; \prod\right)$ is given as:

$$\mu\left(w, x; \prod\right) \equiv \mathbb{E}\left[Y_i\left(w\right)|X_i \in \ell\left(x; \prod\right)\right]$$

And its average causal effect is given as:

$$\tau\left(x; \prod\right) \equiv \mathbb{E}\left[Y_i\left(1\right) - Y_i\left(0\right)|X_i \in \ell\left(x; \prod\right)\right]$$

Then the goal is to construct $\pi\left(.\right)$ that maximizes the following honest criterion:

$$Q^H\left(\pi\right) = -\mathbb{E}_{S^{te}, S^{est}}\left[MSE_\tau\left(S^{te}, S^{est}, \pi\left(S^{tr}\right)\right)\right]$$

Athey and Imben Causal Tree back

The above equation can be rearranged as:

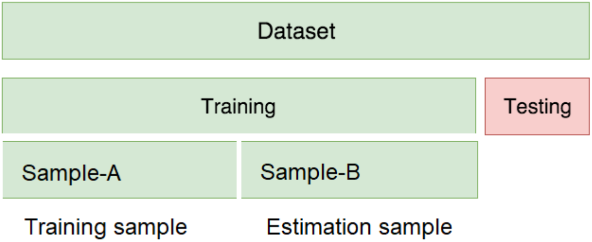
$$-EMSE_{\tau}\left(S^{tr}, N^{est}, \prod\right) = \alpha \frac{1}{N^{tr}} \sum_{i \in S^{tr}} \left(X_i; S^{tr}, \prod, \alpha\right)$$

$$-(1-\alpha)\left(\frac{1}{N^{tr}} + \frac{1}{N^{est}}\right) \cdot \sum_{\ell \in \prod} \left(\frac{s^2_{S^{tr}_{treat}}(\ell)}{p} + \frac{s^2_{S^{tr}_{control}}(\ell)}{1-p}\right)$$

Which provides a causal tree, where $s^2_{S^{tr}_{treat}}(\ell)$ is the within-leaf variance on outcomes Y for $S^{tr}_{control}$in leaf ℓ; $s^2_{S^{tr}_{control}}(\ell)$ is the counterpart for S^{tr}_{treat}; $p = N_{treat}/N$ is the treatment probability, $\alpha \in (0, 1)$ and is a parameter to adjust the portion of MSE and the variance of $EMSE$.

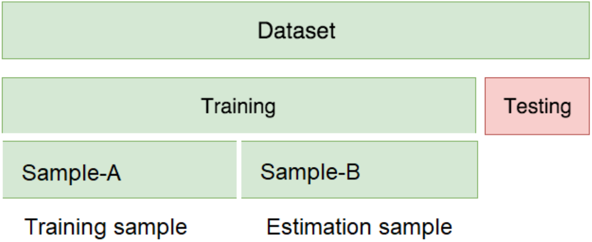
Sample Splitting, Honest Estimation and Cross-fitting back

- Sample splitting and Honest estimation.

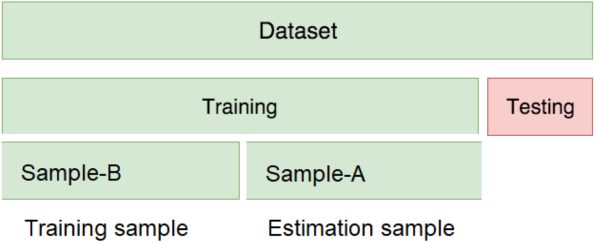


Sample Splitting, Honest Estimation and Cross-fitting back

- Sample splitting and Honest estimation.



- Cross-fitting



Bootstrapping and Aggregating (Boosting) back

Aggregation of Information

Then probability of x out of n choosing correctly follows a bi-nominal distribution given as: $P_n = \sum_{x=\frac{n+1}{2}}^{n} \binom{n}{x} p^x (1-p)^{n-x}$

Extremely large n makes correct choice

$\lim_{n\to\infty} P_n \to 1.$

This means that the limit results that group competence approaches one as group size approaches infinity. In other words, extremely large committees almost certainty make the correct choice.

Causal Random Forest back

- Random Forest approach makes prediction from an average of b CARTs or trees, as follow:

  (1) for each tree $b = 1, \ldots, B$, draw a subsample $S_b \subseteq \{1, \ldots, n\}$

  (2) grow a tree via recursive partitioning on each such subsample of the data; and

  (3) make a prediction by averaging the prediction made by individual tree as:

$$\hat{\mu}(x) = \frac{1}{B} \sum_{b=1}^{B} \sum_{n=1}^{n} \frac{Y_i \mathbf{1}\left(\{X_i \in L_b(x), i \in S_b\}\right)}{|\{i : X_i \in L_b(x), i \in S_b\}|}$$

- where, $L_b(x)$ denotes the leaf of the b^{th} tree containing the training sample x.
- For out-of-bag prediction, one can estimate the average as $\hat{\mu}^{(-i)}(x)$ by only considering those trees b for which $i \notin S_b$. $(-i)$ superscript denote "out-of-bag" or "out-of-fold" prediction.

R−Learner Objective Function back

- "$R-$learner" objective function for heterogeneous treatment effect estimation as:

$$\hat{\tau}(\cdot) = \arg\min_{\tau} \left\{ \sum_{i=1}^{n} \left(\left(Y_i - \hat{m}^{(-i)}(X_i) \right) - \tau(X_i) \left(W_i - \hat{e}^{(-i)}(X_i) \right) \right)^2 + \right.$$

- where, $\lambda_n(\tau(\cdot))$ is a "regularizer" that controls the complexity of the learned conditional average treatment effect $\hat{\tau}(\cdot)$ function.
- $e(x) = P[W_i | X_i = x]$ is the propensity score or probability of being treated.
- $m(x) = E[Y_i | X_i = x]$ is expected outcomes marginalizing over treatment; $(-i)$ superscript denote "out-of-bag" or "out-of-fold" prediction.

Causal Random Forest back

- Random Forest ensembles of many trees and provides prediction as an average prediction made by many individual trees.
- A Random Forest can be equivalent as an adaptive kernel method and re-express the random forest from equation:

$$\hat{\mu}(x) = \sum_{i=1}^{n} a_i(x) Y_i; \quad a_i(x) = \frac{1}{B} \sum_{b=1}^{B} \frac{Y_i \mathbf{1}(\{X_i \in L_b(x), i \in S_b\})}{|\{i : X_i \in L_b(x), i \in S_b\}|}$$

- where, $a_i(x)$ is a data-adaptive kernel or simply weights that measure how often the i^{th} training example appears in the same leaf as the test point x.
- The kernel-based perspective on forests suggests a natural way to use them for treatment estimation by first growing a forest to get weights $a_i(x)$, and then set

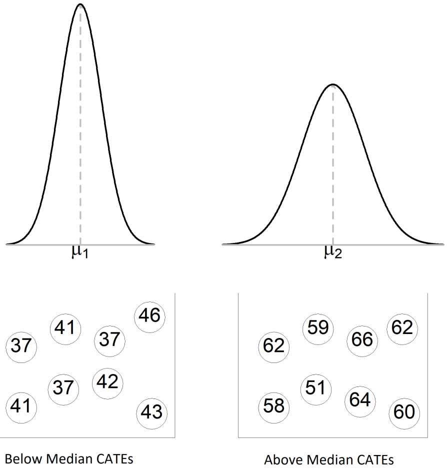
$$\hat{\tau} = \frac{\sum_{i=1}^{n} a_i(x_i)\left(Y_i - \hat{m}^{(-i)}(X_i)\right)\left(W_i - \hat{e}^{(-i)}(X_i)\right)}{\sum_{i=1}^{n} a_i(x_i)\left(W_i - \hat{e}^{(-i)}(X_i)\right)}$$

Causal Random Forest back

- At the implementation level, the causal forest starts by fitting two separate regression forests to estimate $\hat{m}(\cdot)$ and $\hat{e}(\cdot)$ and making out-of-bag predictions using these two first-stage forests.
- Then the model uses these out-of-bag predictions as inputs to the causal forest where cross-validation on the "$R-$learner" objective function, chooses the tuning parameters for the causal forest.

Accessing Treatment Heterogeneity back

- **Heuristic** test:
  - Group observation based on CATE (above or below median CATE).
  - Test if the ATE of these two groups is different from each other or not.
  - It provides qualitative insights about the strength of heterogeneity.



Accessing Treatment Heterogeneity back

- Mean Forest Prediction (**MFP**) by Chernozhukov et al. (2018)
  - Define: $B_i = Y_i - \hat{y}_i^{(-i)}$
  - Define: $C_i = \bar{\tau}(W_i - \hat{e}_i^{(-i)})$
  - $\bar{\tau}$ is out-of-sample ATE, and $\hat{e}_i^{(-i)}$ is propensity.
  - $\frac{dB_i}{dC_i} = 1$, for calibrated model.
- Differential Forest Prediction (**DFP**) by Chernozhukov et al. (2018)
  - $D_i = (\hat{\tau}^{(-i)}(X_i) - \bar{\tau})(W_i - \hat{e}_i^{(-i)})$
  - $\frac{dB_i}{dD_i} > 0$, for existence of treatment heterogeneity.

Accessing Treatment Heterogeneity back

Table 2: Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Outpatient visits last six months	0.07***	0.224***	0.062***	0.055***	1.028***	1.316***
	(0.009)	(0.027)	(0.009)	(0.017)	(0.145)	(0.312)

Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The regressions in Columns (1) and (2) include household size dummies, survey wave dummies, and survey wave interacted with household size dummies. For the LATE estimates in Column (2), the instrumental variable is lottery assignment, and the endogenous variable is "Ever in Medicaid". The ITT and LATE estimates are base on the double-selection post-LASSO.

- Measure of Treatment Heterogeneity: Heuristic*, $MFP = 1$, and $DFP > 0$.

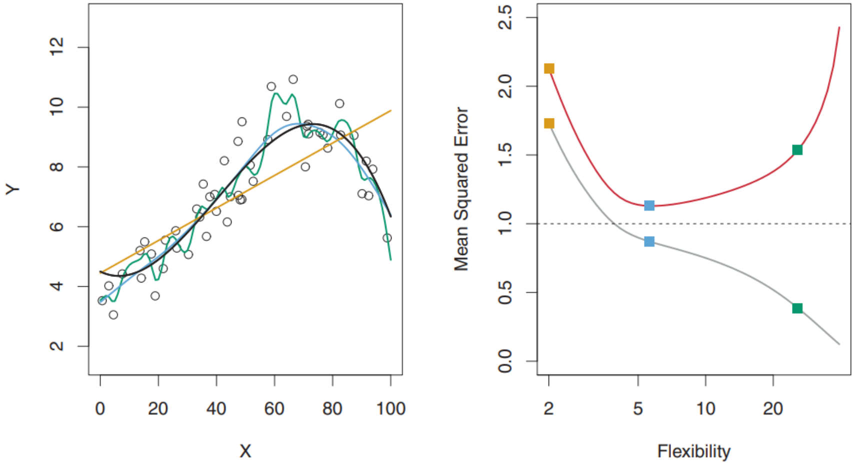
Results: Graphical Intuition (3D Plot) back

3D Visualization.

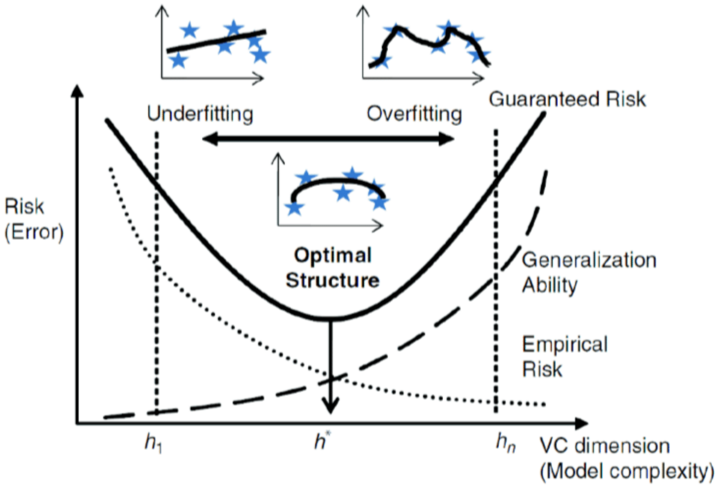
Frisch–Waugh–Lovell (FWL) theorem and Double Machine Learning (DML) back

- $y = f(d, x)$.
- Imagine f is a linear function.
- $y = f(x)$ and get residual $e_1 = y - \hat{y}$.
- $d = f(x)$ and get residual $e_2 = d - \hat{d}$.
- $e_1 = f(e_2)$ the coefficient yields how y changes w.r.t. the variable d.
- This is FWL theorem.
- Now, imagine f as some ML algorithm, hence this is is DML.

Training and Testing Error and VC Dimension back

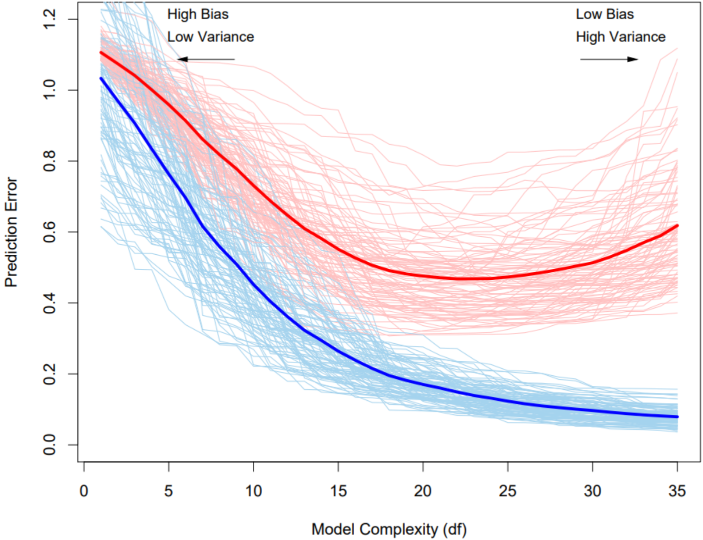


Training and Testing Error and VC Dimension back



$$TestingError \leq TrainingError + \sqrt{VC(\Pi)/n}$$

Training and Testing Error and VC Dimension back



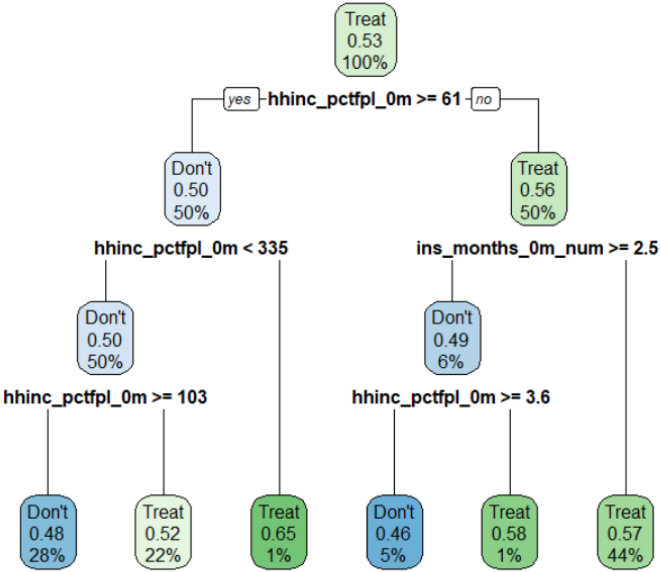
$$TestingError \leq TrainingError + \sqrt{VC\left(\Pi\right)/n}$$

Efficient Policy back

Table 6: Estimate of the utility improvement of various policies over a random assignment baseline.

Variable	Random assignment (1)	Probability rule (2)	CATE rule (3)	Shallow tree (4)	Deeper tree (5)
Panel A: Health care utilization					
Outpatient visits last six months	0.604***	4.74***	5.119***	4.228***	2.898***
	(0.002)	(0.182)	(0.17)	(0.197)	(0.177)
Panel B: Preventive care utilization					
Blood cholesterol checked (ever)	0.659***	0.575***	3.023***	1.934***	1.59***
	(0.005)	(0.176)	(0.146)	(0.166)	(0.154)
Blood tested for high blood sugar/diabetes (ever)	0.625***	1.066***	3.059***	2.665***	2.068***
	(0.003)	(0.157)	(0.124)	(0.137)	(0.178)
Mammogram within last 12 months (women + 40)	0.331***	7.008***	10.228***	9.26***	5.75***
	(0.002)	(0.482)	(0.398)	(0.552)	(0.42)
Pap test within last 12 months (women)	0.411***	3.489***	5.682***	4.955***	4.058***
	(0.003)	(0.286)	(0.24)	(0.315)	(0.316)
Panel C: Self-reported health					
Self-reported health good/very good/excellent (not fair or poor)	0.579***	1.952***	4.186***	4.225***	2.588***
	(0.003)	(0.174)	(0.145)	(0.201)	(0.195)
Panel D: Potential mechanism					
Have usual place of clinic-based care	0.558***	5.462***	7.44***	7.305***	4.718***
	(0.002)	(0.227)	(0.203)	(0.237)	(0.202)
Have personal doctor	0.544***	6.114***	6.432***	6.144***	4.576***
	(0.003)	(0.192)	(0.207)	(0.244)	(0.181)
Happiness, very happy or pretty happy (vs. not too happy)	0.629***	2.137***	4.883***	5.042***	3.306***
	(0.002)	(0.196)	(0.174)	(0.218)	(0.166)

Efficient Policy back



Efficient Policy

back