Safe Policy Learning through Extrapoation

Application to Pre-trial Risk Assessment

Eli Ben-Michael

Harvard University

(joint work with Kosuke Imai, Jim Greiner, and Zhichao Jiang)

Johns Hopkins Causal Inference Seminar
October 2021

Algorithms are making consequential decisions

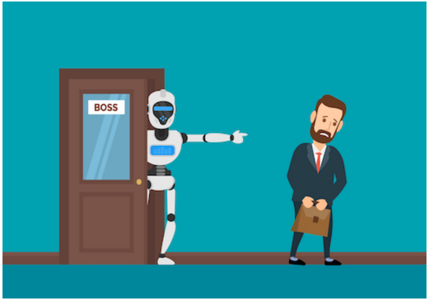
- Often assisting humans make decisions
- Criminal justice, social policy, medicine

Key promises of ML and AI

- Learn when and how to intervene algorithmically
- Design systems to aid human decision makers

Often based on known, **deterministic** rules

- Transparency and interpretability gives accountability
- Lack of overlap: difficult to learn new algorithms



First appearance hearings

- Judge decides pre-trial release conditions
- Cash bail? How much? Monitoring?
- Short, many in one day

Presumption of innocence: judges balance between

- Risk of new crime or failing to appear
- Costs of pre-trial detention

Assessment scores designed to help judges

- Use arrestee features to classify risks
- Recommend a course of action

We use a unique RCT evaluating pre-trial risk assessment to learn new algorithms

The PSA-DMF System

Public Safety Assessment (PSA) classifies 3 risks

1. Failure To Appear in court (FTA)
2. New Criminal Activity (NCA)
3. **New Violent Criminal Activity (NVCA)**

The PSA-DMF System

Public Safety Assessment (PSA) classifies 3 risks

1. Failure To Appear in court (FTA)
2. New Criminal Activity (NCA)
3. **New Violent Criminal Activity (NVCA)**

New Violent Criminal Arrest: Points		
PSA FACTOR	RESPONSE	POINTS
Current violent offense	No	0
	Yes	2
Current violent offense and 20 years old or younger	No	0
	Yes	1
Pending charge at the time of arrest	No	0
	Yes	1
Prior conviction (misdemeanor or felony)	No	0
	Yes	1
Prior violent conviction	No	0
	Yes, 1 or 2	1
	Yes, 3 or more	2

The PSA-DMF System

Public Safety Assessment (PSA) classifies 3 risks

1. Failure To Appear in court (FTA)
2. New Criminal Activity (NCA)
3. **New Violent Criminal Activity (NVCA)**

Decision Making Framework (DMF)

- Combines scores for bail recommendation

New Violent Criminal Arrest: Points		
PSA FACTOR	RESPONSE	POINTS
Current violent offense	No	0
	Yes	2
Current violent offense and 20 years old or younger	No	0
	Yes	1
Pending charge at the time of arrest	No	0
	Yes	1
Prior conviction (misdemeanor or felony)	No	0
	Yes	1
Prior violent conviction	No	0
	Yes, 1 or 2	1
	Yes, 3 or more	2

The PSA-DMF System

Public Safety Assessment (PSA) classifies 3 risks

1. Failure To Appear in court (FTA)
2. New Criminal Activity (NCA)
3. **New Violent Criminal Activity (NVCA)**

Decision Making Framework (DMF)

- Combines scores for bail recommendation

Randomized control trial on risk assessment

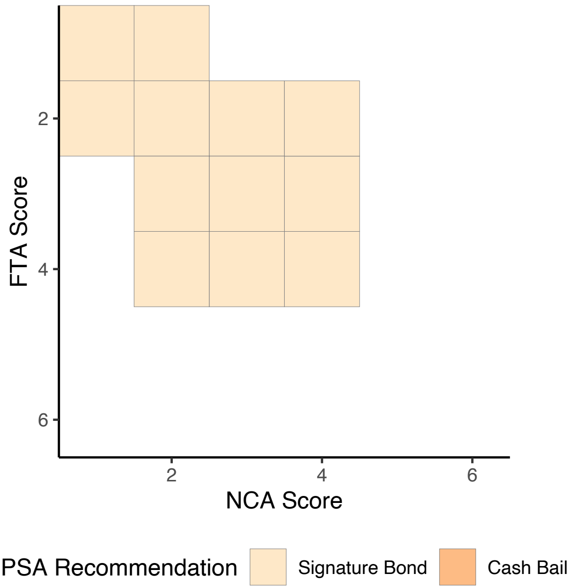
[Greiner et al., 2020; Imai et al., 2020]

- 1891 first arrests in Dane County, WI
- Randomly make the PSA-DMF scores and recommendations available to judges
- Record decision (signature bond or cash bail)
- Record outcome (NVCA)

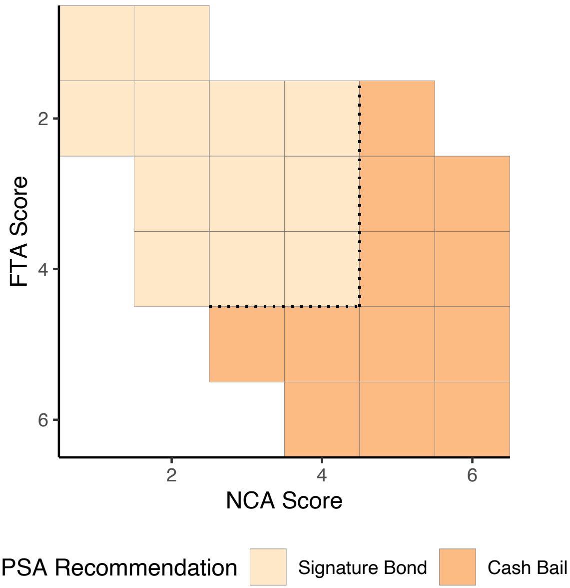
New Violent Criminal Arrest: Points

PSA FACTOR	RESPONSE	POINTS
Current violent offense	No	0
	Yes	2
Current violent offense and 20 years old or younger	No	0
	Yes	1
Pending charge at the time of arrest	No	0
	Yes	1
Prior conviction (misdemeanor or felony)	No	0
	Yes	1
Prior violent conviction	No	0
	Yes, 1 or 2	1
	Yes, 3 or more	2

The DMF matrix



The DMF matrix



Safe policy learning through extrapolation

Deterministic policies pose thorny identification issues

- Without overlap, we can't identify the counterfactual
- Would things be better under a different course of action?

Existing status quo algorithms are an important benchmark

- Already implemented with institutional support
- The data speaks directly to quality

Safe policy learning through extrapolation

Deterministic policies pose thorny identification issues

- Without overlap, we can't identify the counterfactual
- Would things be better under a different course of action?

Existing status quo algorithms are an important benchmark

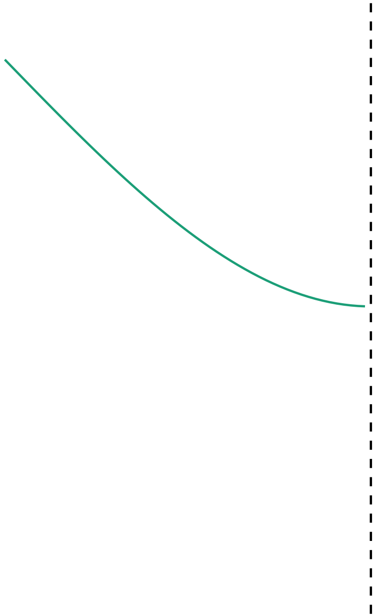
- Already implemented with institutional support
- The data speaks directly to quality

We extrapolate from the data in a safe way by finding the **best policy in the worst case**

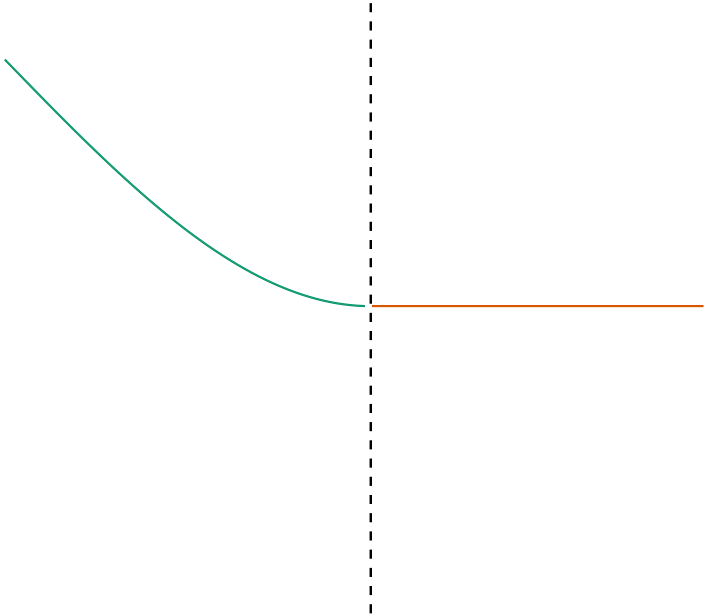
$$\max_{\text{policies } \pi} \ \min_{\text{models } \mathcal{M}} \ \text{Value}(\pi, m)$$

- A **robust optimization** approach that partially identifies the value
- Optimizing for the worst-case across observationally equivalent models
- A statistical safety guarantee of improvement over the status quo

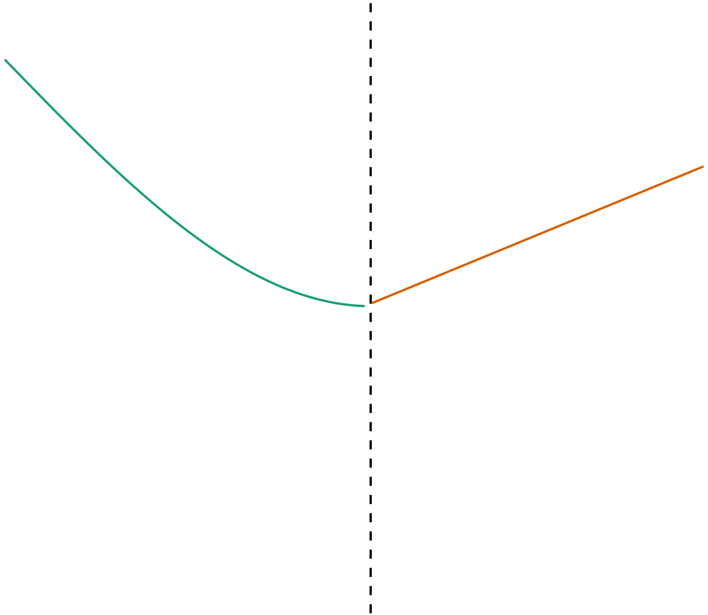
Many ways to extrapolate from observable data



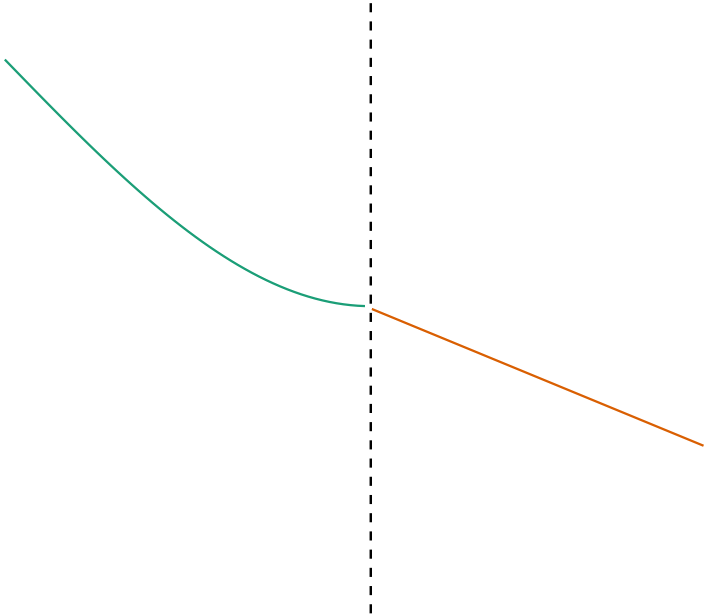
Many ways to extrapolate from observable data



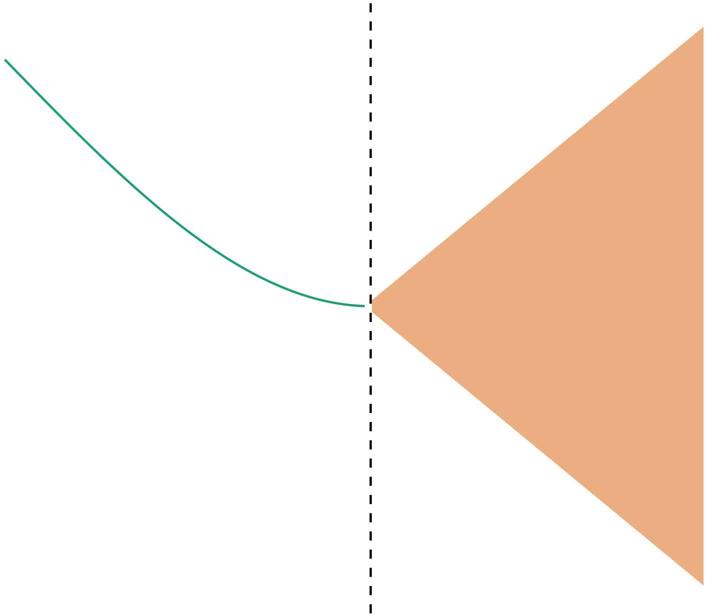
Many ways to extrapolate from observable data



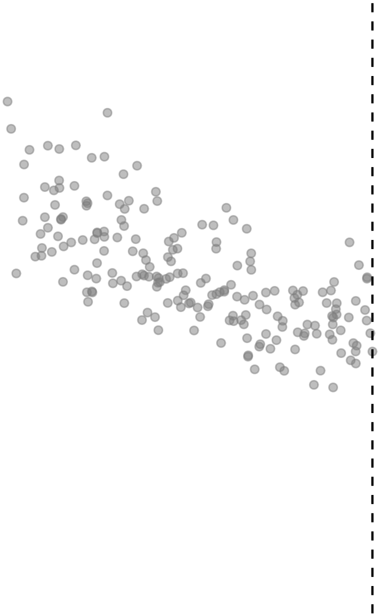
Many ways to extrapolate from observable data



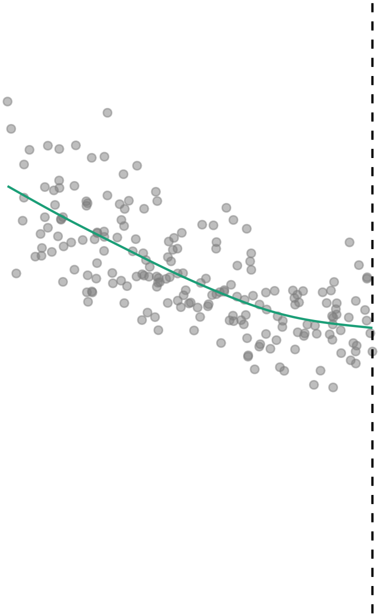
Many ways to extrapolate from observable data



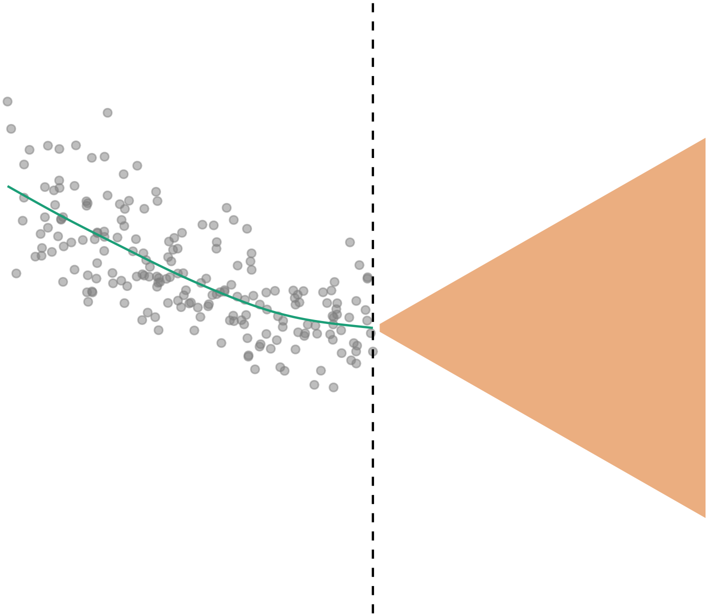
Noise creates additional uncertainty



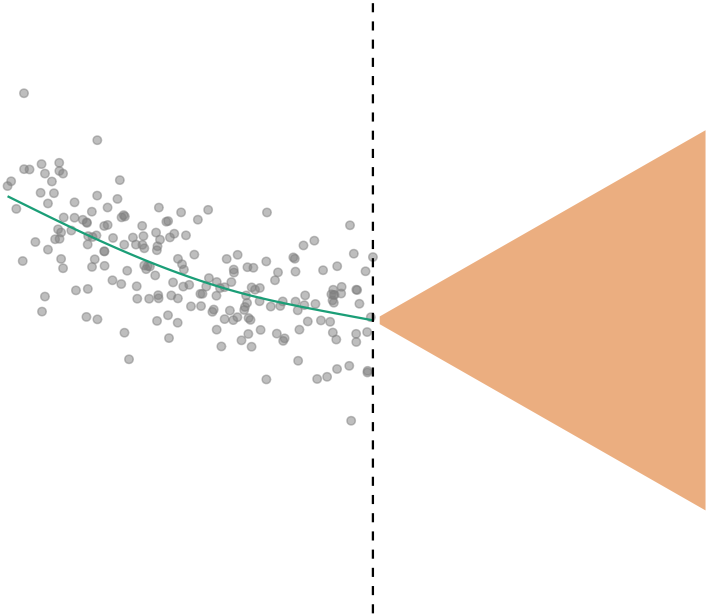
Noise creates additional uncertainty



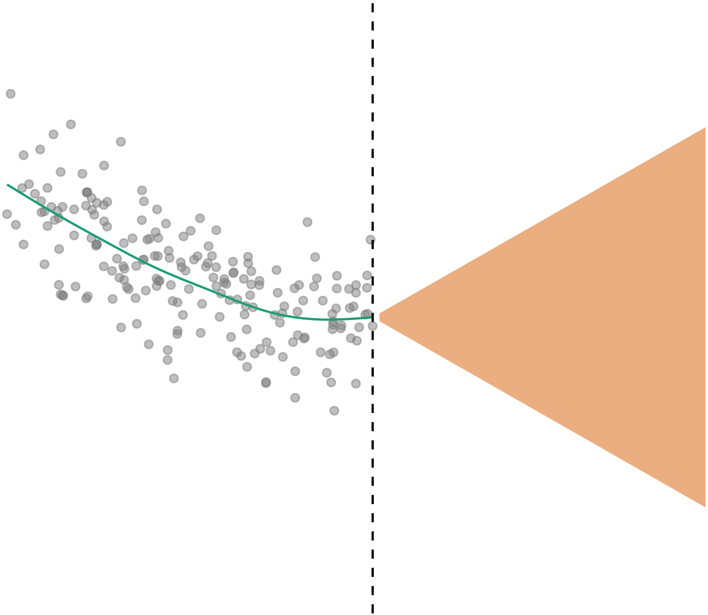
Noise creates additional uncertainty



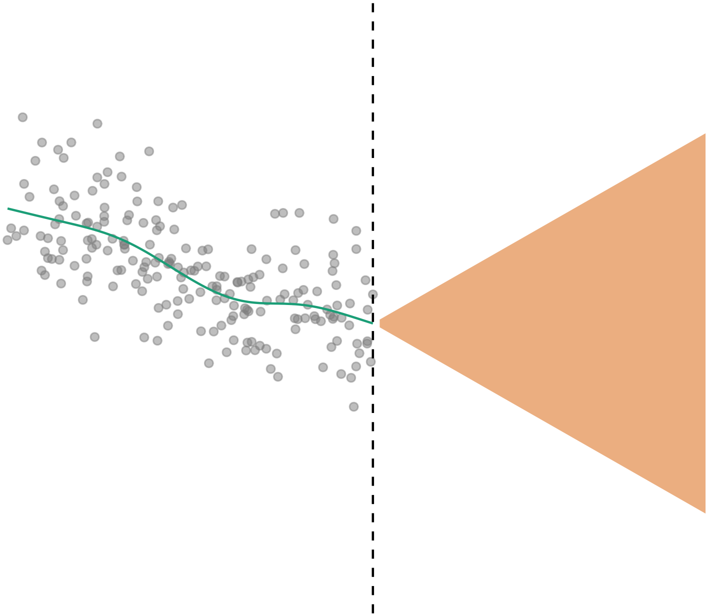
Noise creates additional uncertainty



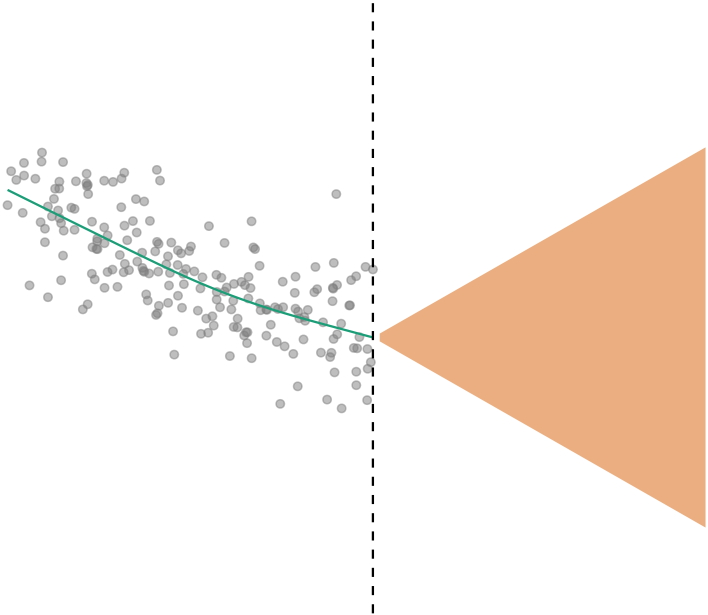
Noise creates additional uncertainty



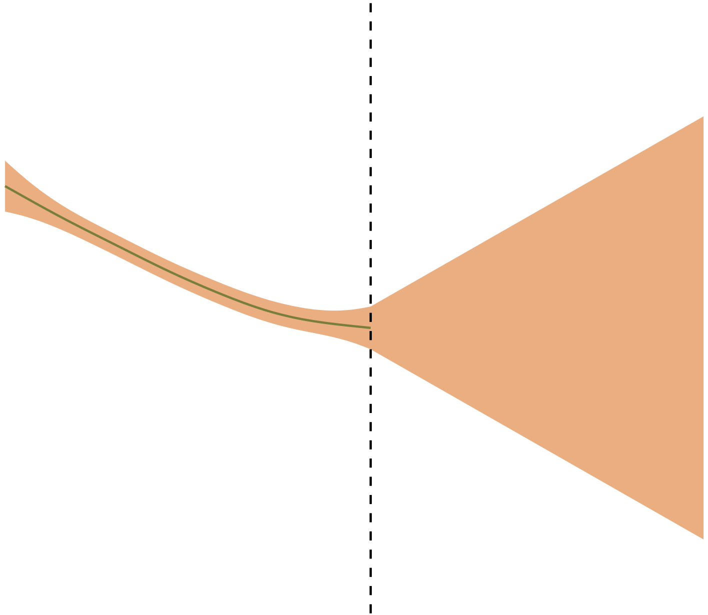
Noise creates additional uncertainty



Noise creates additional uncertainty



Noise creates additional uncertainty



Robust optimization defends against perils of extrapolation

Need to account for both forms of uncertainty

1. Epistemic uncertainty: many ways to extrapolate
2. Statistical uncertainty: only see noisy outcomes

Robust optimization gives a safety guarantee across both forms of uncertainty

Only changes the status quo if data and models support improvement

The Population Safe Policy

Setup

For each individual i, observe

- Covariates $X_i \in \mathcal{X}$
- Action taken $A_i \in \mathcal{A}$
- Binary outcome $Y_i \in \{0, 1\}$

Don't observe potential outcome under action a, $Y(a)$

- Conditional expectation $m(a, x) = \mathbb{E}[Y(a) \mid X = x]$

Setup

For each individual i, observe

- Covariates $X_i \in \mathcal{X}$
- Action taken $A_i \in \mathcal{A}$
- Binary outcome $Y_i \in \{0, 1\}$

Don't observe potential outcome under action a, $Y(a)$

- Conditional expectation $m(a, x) = \mathbb{E}[Y(a) \mid X = x]$

Deterministic baseline policy $\tilde{\pi}$

- Observed outcomes are $Y_i = Y_i(\tilde{\pi}(X_i))$
- Partitions the covariate space $\mathcal{X}_a = \{x \in \mathcal{X} \mid \tilde{\pi}(x) = a\}$

Setup

For each individual i, observe

- Covariates $X_i \in \mathcal{X}$
- Action taken $A_i \in \mathcal{A}$
- Binary outcome $Y_i \in \{0, 1\}$

Don't observe potential outcome under action a, $Y(a)$

- Conditional expectation $m(a, x) = \mathbb{E}[Y(a) \mid X = x]$

Deterministic baseline policy $\tilde{\pi}$

- Observed outcomes are $Y_i = Y_i(\tilde{\pi}(X_i))$
- Partitions the covariate space $\mathcal{X}_a = \{x \in \mathcal{X} \mid \tilde{\pi}(x) = a\}$

Utility of actions and outcomes

$$c(a) + uY(a)$$

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + uY(a))\right] = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + um(a, X))\right]$$

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + uY(a))\right] = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + um(a, X))\right]$$

But how do we identify the counterfactuals?

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(a \mid X)\,(c(a) + uY(a))\right] = \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(a \mid X)\,(c(a) + um(a,X))\right]$$

But how do we identify the counterfactuals?

$$\text{When } \tilde{\pi}(x) = a \qquad \mathbb{E}[Y(a) \mid X = x] \quad =$$

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + uY(a))\right] = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + um(a, X))\right]$$

But how do we identify the counterfactuals?

$$\text{When } \tilde{\pi}(x) = a \qquad \mathbb{E}[Y(a) \mid X = x] \quad = \quad \mathbb{E}[Y \mid X = x]$$

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + uY(a))\right] = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + um(a,X))\right]$$

But how do we identify the counterfactuals?

$$\begin{aligned} &\text{When } \tilde{\pi}(x) = a \qquad \mathbb{E}[Y(a) \mid X = x] &&= \qquad \mathbb{E}[Y \mid X = x] \\ &\text{When } \tilde{\pi}(x) \neq a \qquad \mathbb{E}[Y(a) \mid X = x] &&= \end{aligned}$$

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\left(c(a) + uY(a)\right)\right] = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\left(c(a) + um(a, X)\right)\right]$$

But how do we identify the counterfactuals?

$$\begin{array}{lccc} \text{When } \tilde{\pi}(x) = a & \mathbb{E}[Y(a) \mid X = x] & = & \mathbb{E}[Y \mid X = x] \\ \text{When } \tilde{\pi}(x) \neq a & \mathbb{E}[Y(a) \mid X = x] & = & ? \end{array}$$

Identification issues get in the way

Our goal: Find a policy with high **value**

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + uY(a))\right] = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + um(a, X))\right]$$

But how do we identify the counterfactuals?

$$\begin{array}{llcl} \text{When } \tilde{\pi}(x) = a & \mathbb{E}[Y(a) \mid X = x] & = & \mathbb{E}[Y \mid X = x] \\ \text{When } \tilde{\pi}(x) \neq a & \mathbb{E}[Y(a) \mid X = x] & = & ? \end{array}$$

Existing work uses **stochastic** policies for identification

[e.g. Beygelzimer and Langford, 2009; Qian and Murphy, 2011; Zhang et al., 2012; Zhao et al., 2012; Swaminathan and Joachims, 2015; Kitagawa and Tetenov, 2018; Kallus, 2018]

$$\mathbb{E}[Y(a) \mid X = x] = \mathbb{E}\left[\frac{Y\mathbb{1}\{A = a\}}{P(A = a \mid X = x)} \mid X = x\right]$$

Double robust methods as well

[e.g. Dudik and Langford, 2011; Luedtke and Van Der Laan, 2016; Athey and Wager, 2021]

Instead, optimize for the worst case

Decompose the value into

$$V(\pi, m) =$$

Instead, optimize for the worst case

Decompose the value into identifiable

$$V(\pi, m) = \underbrace{\mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\left(c(a) + uY\tilde{\pi}(a \mid X)\right)\right]}_{\pi \text{ and } \tilde{\pi} \text{ agree}}$$

Instead, optimize for the worst case

Decompose the value into identifiable and unidentifiable components

$$V(\pi, m) = \underbrace{\mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\left(c(a) + uY\tilde{\pi}(a \mid X)\right)\right]}_{\pi \text{ and } \tilde{\pi} \text{ agree}} + \underbrace{\mathbb{E}\left[\sum_{a \in \mathcal{A}} u\pi(a \mid X)(1 - \tilde{\pi}(a \mid X))m(a, X)\right]}_{\pi \text{ and } \tilde{\pi} \text{ disagree}}$$

Instead, optimize for the worst case

Decompose the value into identifiable and unidentifiable components

$$V(\pi, m) = \underbrace{\mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(a \mid X)\,(c(a) + uY\tilde{\pi}(a \mid X))\right]}_{\pi \text{ and } \tilde{\pi} \text{ agree}} + \underbrace{\mathbb{E}\left[\sum_{a\in\mathcal{A}} u\pi(a \mid X)(1 - \tilde{\pi}(a \mid X))m(a, X)\right]}_{\pi \text{ and } \tilde{\pi} \text{ disagree}}$$

If we can partially identify $m \in \mathcal{M}$, then find the best policy in the worst case

Instead, optimize for the worst case

Decompose the value into identifiable and unidentifiable components

$$V(\pi, m) = \underbrace{\mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(a \mid X)\,(c(a) + uY\tilde{\pi}(a \mid X))\right]}_{\pi \text{ and } \tilde{\pi} \text{ agree}} + \underbrace{\mathbb{E}\left[\sum_{a\in\mathcal{A}} u\pi(a \mid X)(1 - \tilde{\pi}(a \mid X))m(a, X)\right]}_{\pi \text{ and } \tilde{\pi} \text{ disagree}}$$

If we can partially identify $m \in \mathcal{M}$, then find the best policy in the worst case

$$\pi^{\text{inf}} \in \operatorname*{argmax}_{\pi\in\Pi} \min_{m\in\mathcal{M}} V(\pi, m)$$

Instead, optimize for the worst case

Decompose the value into identifiable and unidentifiable components

$$V(\pi, m) = \underbrace{\mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(a \mid X)\,(c(a) + uY\tilde{\pi}(a \mid X))\right]}_{\pi \text{ and } \tilde{\pi} \text{ agree}} + \underbrace{\mathbb{E}\left[\sum_{a \in \mathcal{A}} u\pi(a \mid X)(1 - \tilde{\pi}(a \mid X))m(a, X)\right]}_{\pi \text{ and } \tilde{\pi} \text{ disagree}}$$

If we can partially identify $m \in \mathcal{M}$, then find the best policy in the worst case

$$\pi^{\text{inf}} \in \operatorname*{argmax}_{\pi \in \Pi} \min_{m \in \mathcal{M}} V(\pi, m) \iff \pi^{\text{inf}} \in \operatorname*{argmin}_{\pi \in \Pi} \max_{m \in \mathcal{M}} V(\tilde{\pi}) - V(\pi, m).$$

Instead, optimize for the worst case

Decompose the value into identifiable and unidentifiable components

$$V(\pi, m) = \underbrace{\mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(a \mid X)\,(c(a) + uY\tilde{\pi}(a \mid X))\right]}_{\pi \text{ and } \tilde{\pi} \text{ agree}} + \underbrace{\mathbb{E}\left[\sum_{a\in\mathcal{A}} u\pi(a \mid X)(1 - \tilde{\pi}(a \mid X))m(a, X)\right]}_{\pi \text{ and } \tilde{\pi} \text{ disagree}}$$

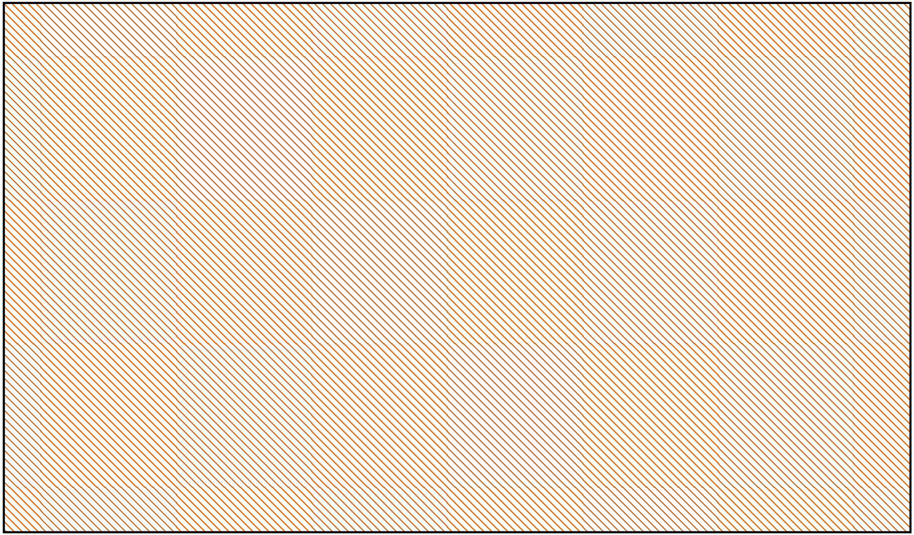
If we can partially identify $m \in \mathcal{M}$, then find the best policy in the worst case

$$\pi^{\text{inf}} \in \operatorname*{argmax}_{\pi\in\Pi} \min_{m\in\mathcal{M}} V(\pi, m) \iff \pi^{\text{inf}} \in \operatorname*{argmin}_{\pi\in\Pi} \max_{m\in\mathcal{M}} V(\tilde{\pi}) - V(\pi, m).$$

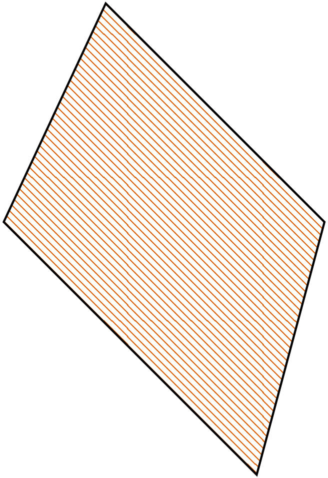
This is a safe policy

- A robust optimization approach [Bertsimas et al., 2011; Kallus and Zhou, 2021; Pu and Zhang, 2021]
- Conservative, “pessimistic” principle [Manski, 2005; Cui, 2021]
- Falls back on status quo if there is too much uncertainty

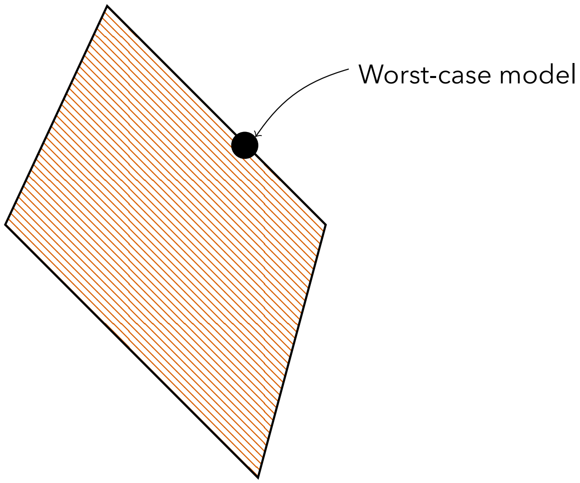
Maximin principle



Maximin principle



Maximin principle



Finding observationally equivalent models

To partially identify the conditional expectation $\mathbb{E}[Y(a) \mid X = x]$

1. Put restrictions on the class of possible models (necessary to make progress)

Finding observationally equivalent models

To partially identify the conditional expectation $\mathbb{E}[Y(a) \mid X = x]$

1. Put restrictions on the class of possible models (necessary to make progress)
2. Compute the set of functions f in the model class that agree with the observable data

$$\tilde{\pi}(a \mid x)\left(f(a, x) - \mathbb{E}\left[Y \mid X = x\right]\right) = 0 \quad \forall\, x \in \mathcal{X}_a$$

Finding observationally equivalent models

To partially identify the conditional expectation $\mathbb{E}[Y(a) \mid X = x]$

1. Put restrictions on the class of possible models (necessary to make progress)
2. Compute the set of functions f in the model class that agree with the observable data

$$\tilde{\pi}(a \mid x)\left(f(a,x) - \mathbb{E}\left[Y \mid X = x\right]\right) = 0 \quad \forall x \in \mathcal{X}_a$$

Many model classes result in pointwise bounds

$$B_\ell(a,x) \leq m(a,x) \leq B_u(a,x)$$

- Lipschitz functions, additive models, linear models

Finding observationally equivalent models

To partially identify the conditional expectation $\mathbb{E}[Y(a) \mid X = x]$

1. Put restrictions on the class of possible models (necessary to make progress)
2. Compute the set of functions f in the model class that agree with the observable data

$$\tilde{\pi}(a \mid x)\left(f(a,x) - \mathbb{E}\left[Y \mid X = x\right]\right) = 0 \quad \forall x \in \mathcal{X}_a$$

Many model classes result in pointwise bounds

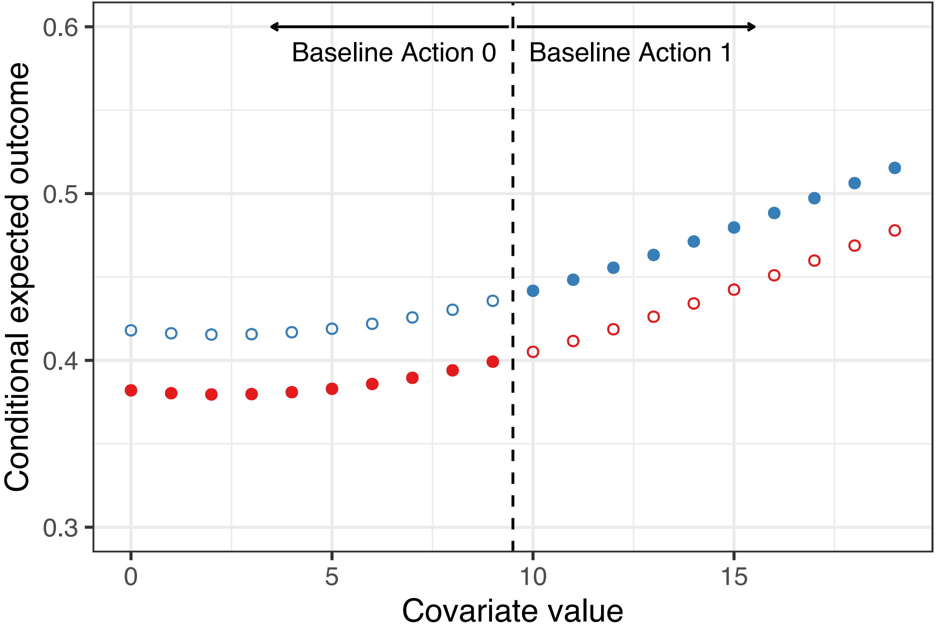
$$B_\ell(a,x) \leq m(a,x) \leq B_u(a,x)$$

- Lipschitz functions, additive models, linear models

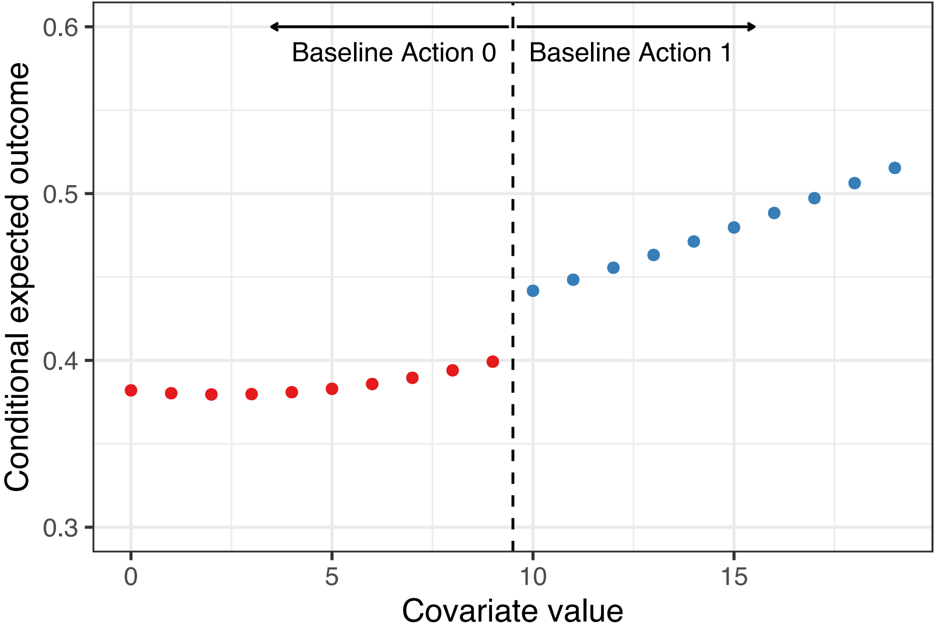
Easy to compute!

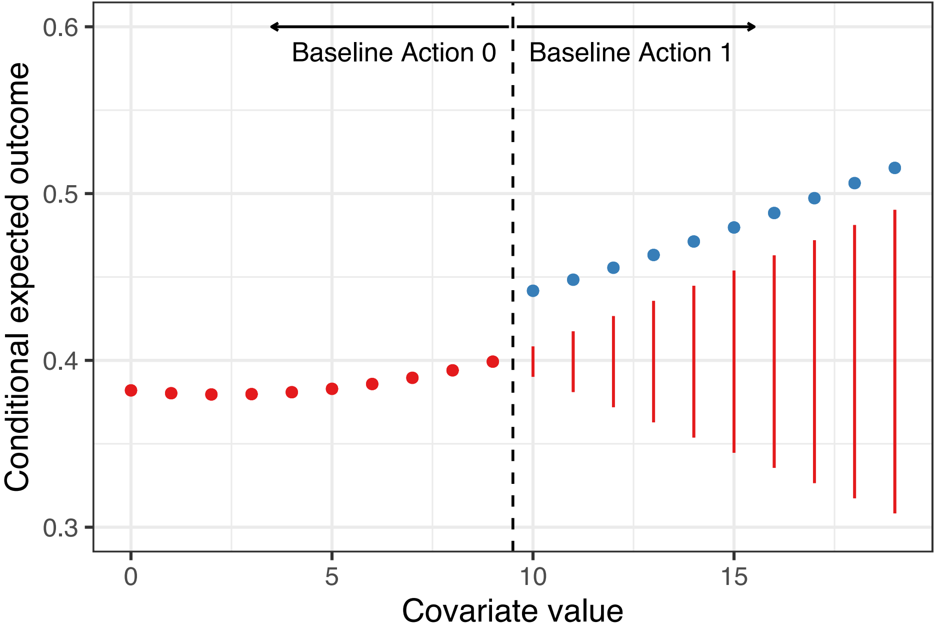
- Use the worst-case bound in place of the missing counterfactual [Pu and Zhang, 2021]

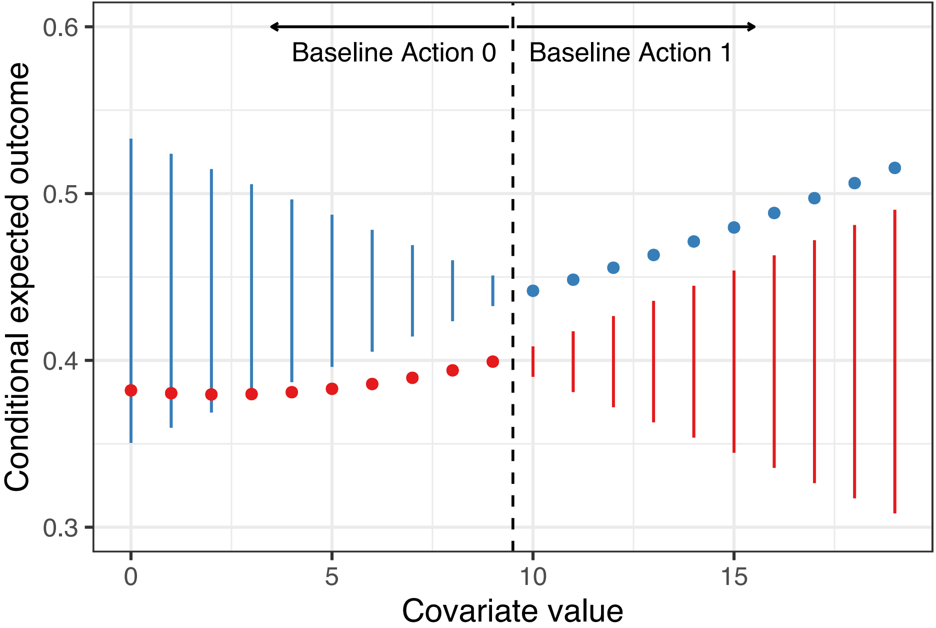
$$\Upsilon(a) = \tilde{\pi}(a \mid X)Y + (1 - \tilde{\pi}(a \mid X))B_\ell(a,X)$$

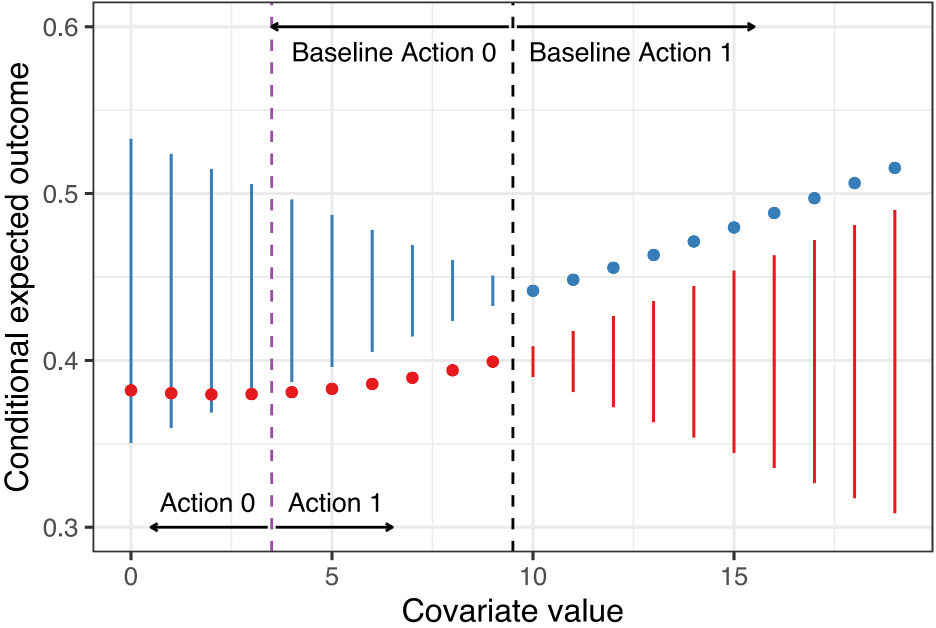


Baseline Action 0
Baseline Action 1
Conditional expected outcome
0.6
0.5
0.4
0.3
0
5
10
15
Covariate value


Baseline Action 0
Baseline Action 1
Conditional expected outcome
Covariate value
0.6
0.5
0.4
0.3
0
5
10
15


Baseline Action 0
Baseline Action 1
Conditional expected outcome
0.6
0.5
0.4
0.3
0
5
10
15
Covariate value


Baseline Action 0
Baseline Action 1
Conditional expected outcome
0.6
0.5
0.4
0.3
0
5
10
15
Covariate value


Baseline Action 0
Baseline Action 1
Action 0
Action 1
Conditional expected outcome
Covariate value
0.6
0.5
0.4
0.3
0
5
10
15

How does this compare?

Compare to baseline policy $\tilde{\pi}$ and oracle policy $\pi^* \in \operatorname{argmax}_{\pi \in \Pi} V(\pi)$

How does this compare?

Compare to baseline policy $\tilde{\pi}$ and oracle policy $\pi^* \in \operatorname{argmax}_{\pi \in \Pi} V(\pi)$

The value of the safe policy is at least as high as the baseline

$$V(\tilde{\pi}) - V\left(\pi^{\text{inf}}\right) \leq 0$$

How does this compare?

Compare to baseline policy $\tilde{\pi}$ and oracle policy $\pi^* \in \operatorname{argmax}_{\pi \in \Pi} V(\pi)$

The value of the safe policy is at least as high as the baseline

$$V(\tilde{\pi}) - V\left(\pi^{\text{inf}}\right) \leq 0$$

Optimality gap controlled by size of $\mathcal{M}$

$$V(\pi^*) - V\left(\pi^{\text{inf}}\right) \leq u\mathbb{E}\left[\max_{a \in \mathcal{A}} B_u(a, X) - B_\ell(a, X)\right]$$

- Safety comes at the cost of a potentially suboptimal policy
- The tighter the partial identification, the lower the cost

The Empirical Safe Policy

Finding the safe policy empirically from data

Construct a **larger** empirical model class $\widehat{\mathcal{M}}_n(\alpha)$

$$P\left(\mathcal{M} \in \widehat{\mathcal{M}}_n(\alpha)\right) \geq 1 - \alpha$$

Finding the safe policy empirically from data

Construct a **larger** empirical model class $\widehat{\mathcal{M}}_n(\alpha)$

$$P\left(\mathcal{M} \in \widehat{\mathcal{M}}_n(\alpha)\right) \geq 1 - \alpha$$

Using simultaneous confidence bands for $\mathbb{E}[Y \mid X = x]$, get pointwise bounds

$$\widehat{B}_{\alpha\ell}(a, x) \leq m(a, x) \leq \widehat{B}_{\alpha u}(a, x)$$

Impute missing counterfactuals from bound

$$\widehat{\Upsilon}_i(a) = \tilde{\pi}(a \mid X)Y + (1 - \tilde{\pi}(a \mid X))\widehat{B}_{\alpha\ell}(a, X)$$

Finding the safe policy empirically from data

Construct a **larger** empirical model class $\widehat{\mathcal{M}}_n(\alpha)$

$$P\left(\mathcal{M} \in \widehat{\mathcal{M}}_n(\alpha)\right) \geq 1 - \alpha$$

Using simultaneous confidence bands for $\mathbb{E}[Y \mid X = x]$, get pointwise bounds

$$\widehat{B}_{\alpha\ell}(a, x) \leq m(a, x) \leq \widehat{B}_{\alpha u}(a, x)$$

Impute missing counterfactuals from bound

$$\widehat{\Upsilon}_i(a) = \tilde{\pi}(a \mid X)Y + (1 - \tilde{\pi}(a \mid X))\widehat{B}_{\alpha\ell}(a, X)$$

Solve an empirical welfare maximization problem

$$\hat{\pi} \in \underset{\pi \in \Pi}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \sum_{a \in \mathcal{A}} \pi(a \mid X_i)(c(a) + u\widehat{\Upsilon}_i(a))$$

Statistical properties

Value is probably, approximately at least as high as baseline

$$V(\tilde{\pi}) - V(\hat{\pi}) \lesssim \text{Complexity}(\Pi) \qquad \text{with probability at least } \gtrsim 1 - \alpha$$

- Conservative approach gives a statistical safety guarantee with level α
- If policy class Π is complex, need more samples to avoid overfitting

Statistical properties

Value is probably, approximately at least as high as baseline

$$V(\tilde{\pi}) - V(\hat{\pi}) \lesssim \text{Complexity}(\Pi) \qquad \text{with probability at least } \gtrsim 1 - \alpha$$

- Conservative approach gives a statistical safety guarantee with level α
- If policy class Π is complex, need more samples to avoid overfitting

Empirical optimality gap controlled by size of $\widehat{\mathcal{M}}_n(\alpha)$ and complexity of Π

$$V(\pi^*) - V(\hat{\pi}) \lesssim \frac{u}{n} \sum_{i=1}^{n} \max_{a \in \mathcal{A}} \widehat{B}_{\alpha u}(a, X_i) - \widehat{B}_{\alpha \ell}(a, X_i) + \text{Complexity}(\Pi)$$

with probability at least $\gtrsim 1 - \alpha$

- Tradeoff between safety and optimality

Learning a new PSA-DMF system

Two necessary adaptations

Incorporating experiments evaluating a deterministic policy

- In our study, judges randomly receive the "null policy" $\emptyset$, no access to PSA

Allows us to work with **treatment effects** instead of outcomes

$$\tau(a,x) = \mathbb{E}[Y(a) - Y(\emptyset) \mid X = x]$$

- Treatment effects are often considered to be simpler than baseline outcomes
  [Künzel et al., 2019; Hahn et al., 2020; Nie and Wager, 2021]

Two necessary adaptations

Incorporating experiments evaluating a deterministic policy

- In our study, judges randomly receive the "null policy" $\emptyset$, no access to PSA

Allows us to work with **treatment effects** instead of outcomes

$$\tau(a,x) = \mathbb{E}[Y(a) - Y(\emptyset) \mid X = x]$$

- Treatment effects are often considered to be simpler than baseline outcomes
  [Künzel et al., 2019; Hahn et al., 2020; Nie and Wager, 2021]

Incorporating human decisions from algorithmic recommendations

- Have to incorporate uncertainty in judge's potential decision $D(a)$

Value includes two unidentified components, outcomes **and decisions**

- Need to find the worst case potential decision and outcome for cost and benefit

Learning a new NVCA flag

Construct a new NVCA flag using the same risk factors

- Current violent offense, $\geq$ 20 years old, prior convictions

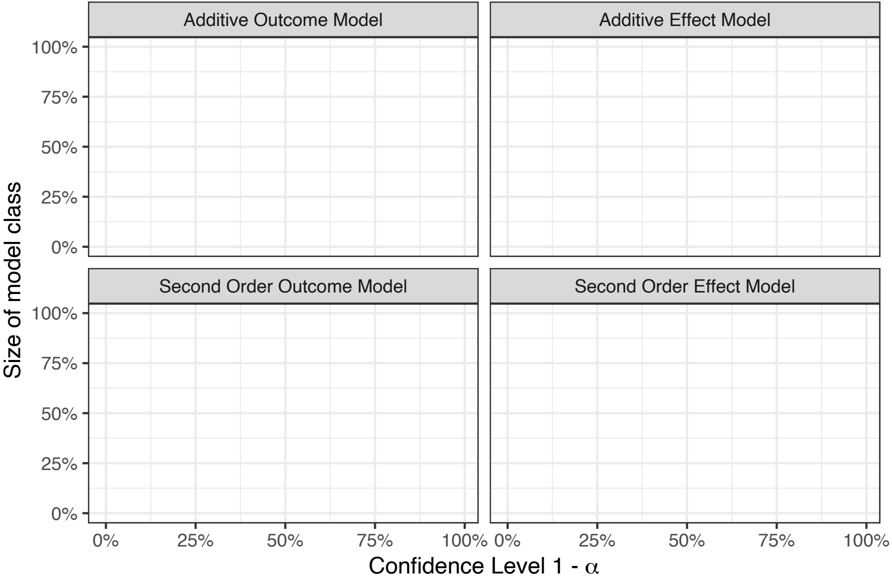
How do we weigh the costs of flagging defendants vs an NVCA?

- Monetary cost of triggering the flag is zero
- But fiscal costs on jurisdiction and socioeconomic costs on individual and community
- **Presumption of innocence**, so limit pre-trial detention

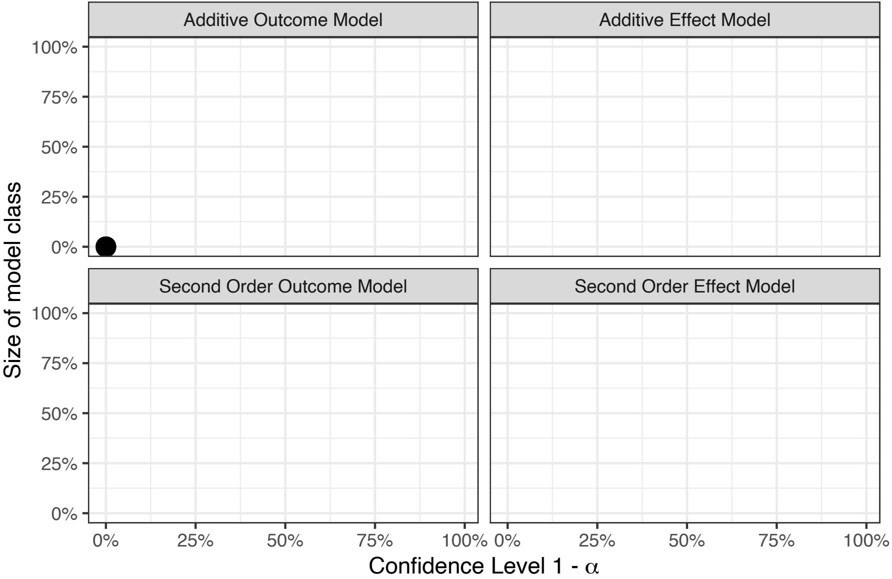
Use a single parameterization:

- Pin fiscal and societal costs to be 1
- Cost of an NVCA starts at 1 and grows

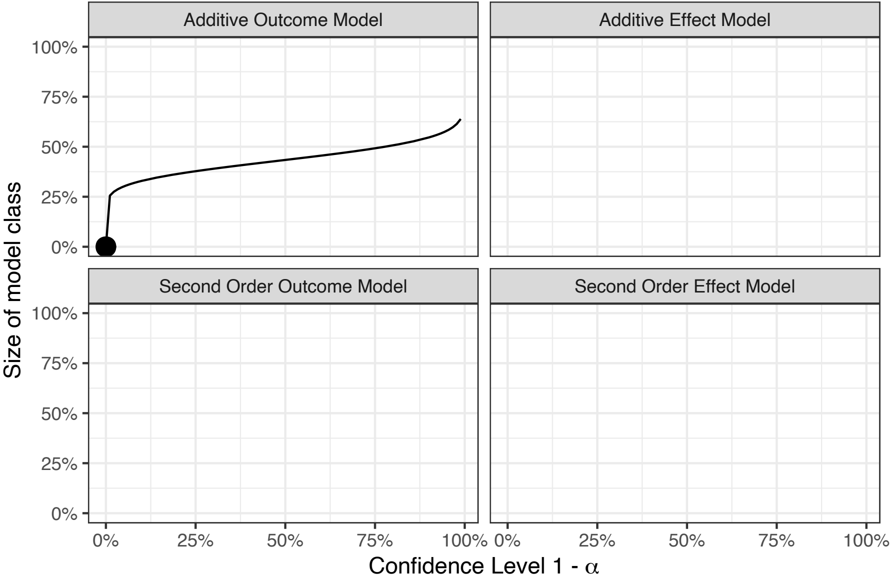
Suboptimality vs safety: diagnostic



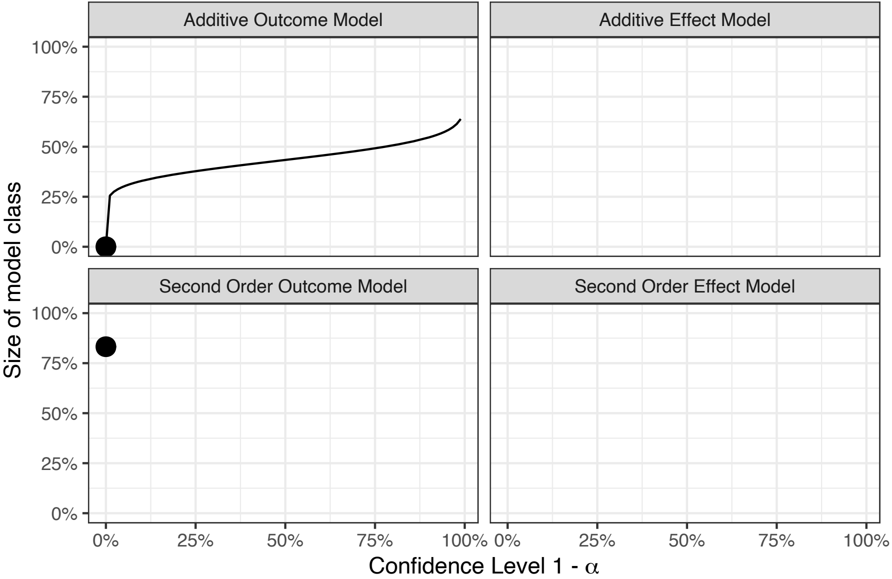
Suboptimality vs safety: diagnostic



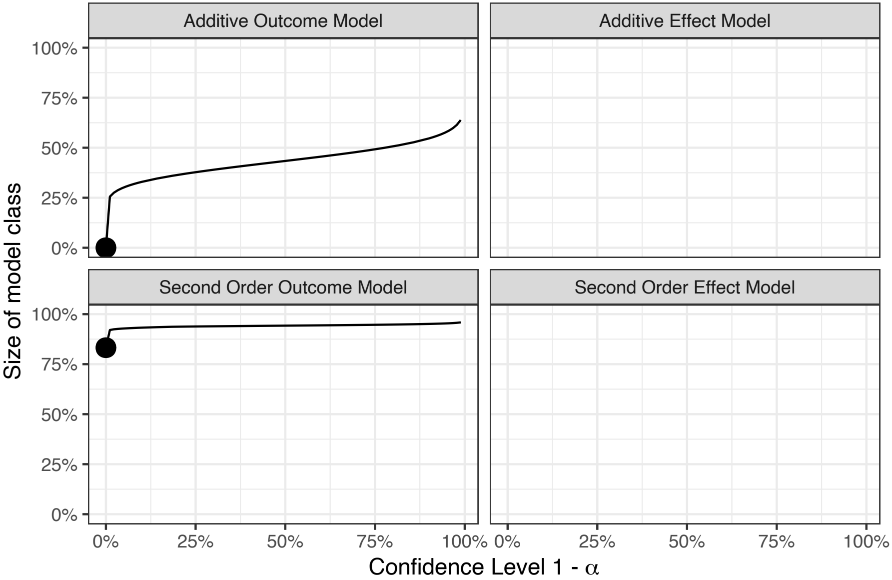
Suboptimality vs safety: diagnostic



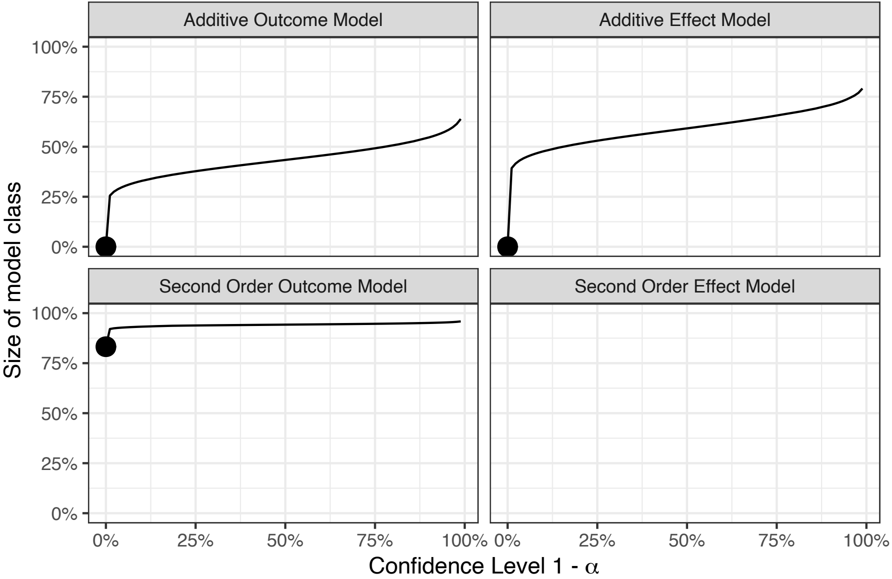
Suboptimality vs safety: diagnostic



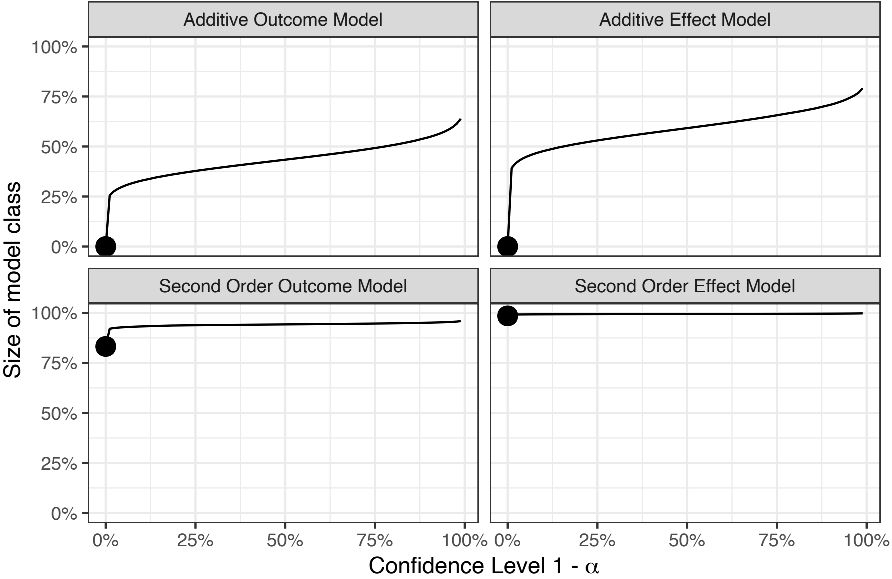
Suboptimality vs safety: diagnostic



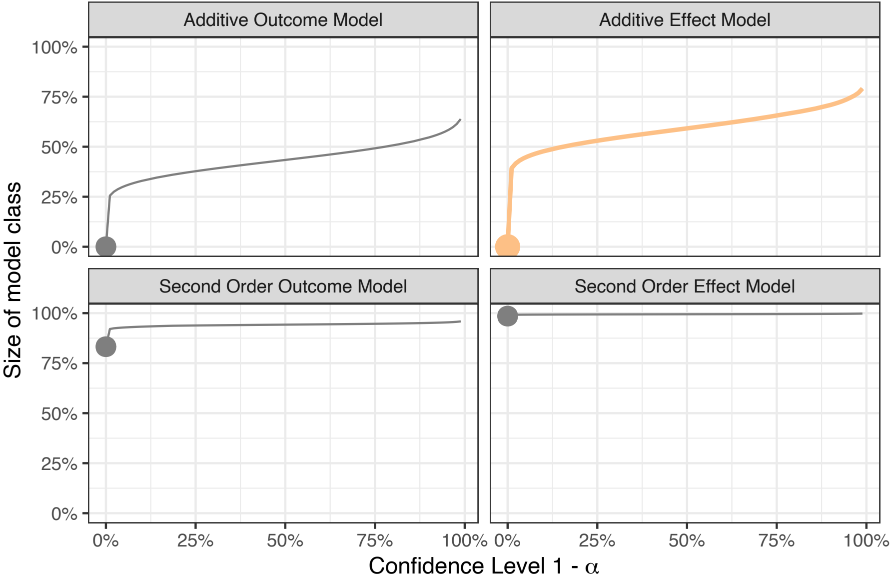
Suboptimality vs safety: diagnostic



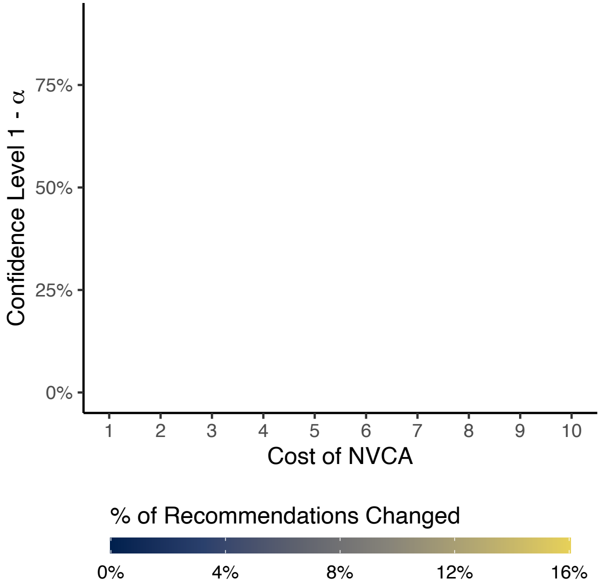
Suboptimality vs safety: diagnostic



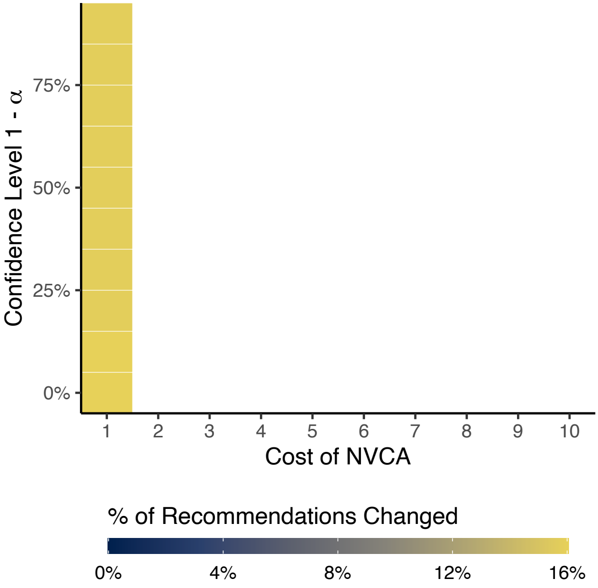
Suboptimality vs safety: diagnostic



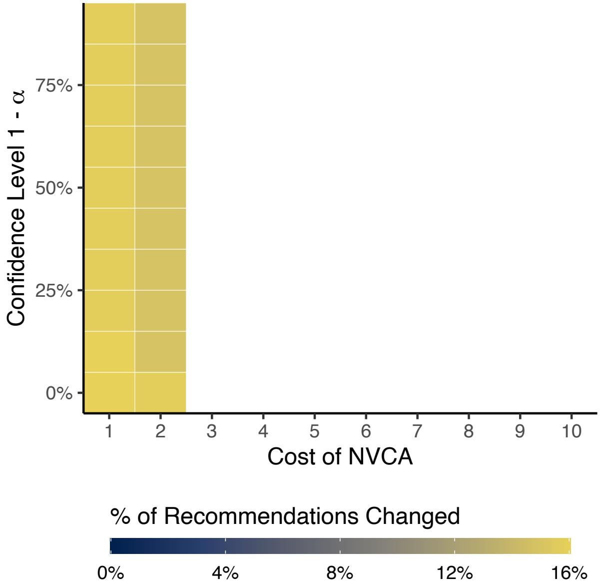
How the NVCA rules change



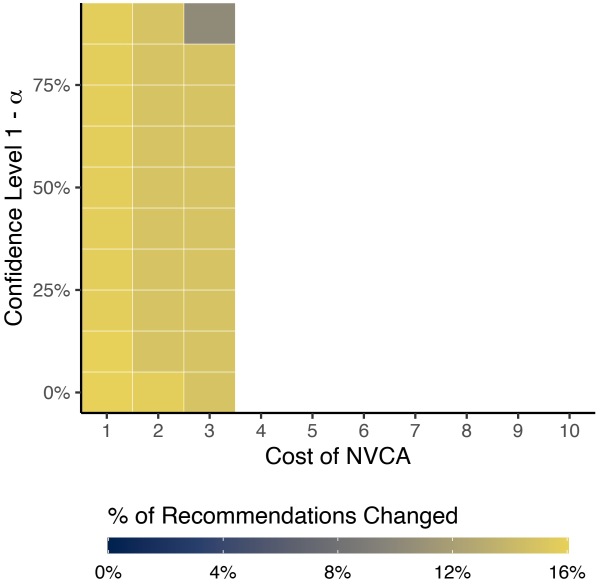
How the NVCA rules change



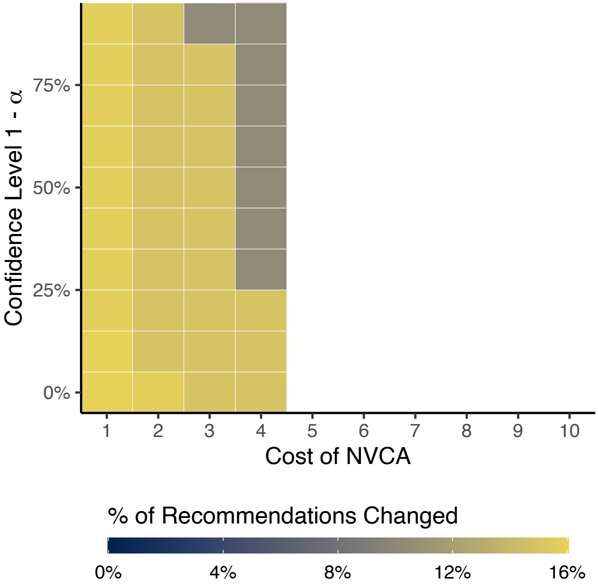
How the NVCA rules change



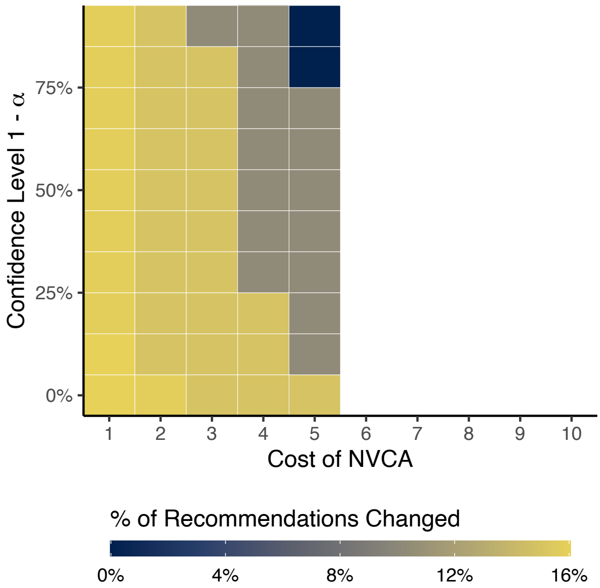
How the NVCA rules change



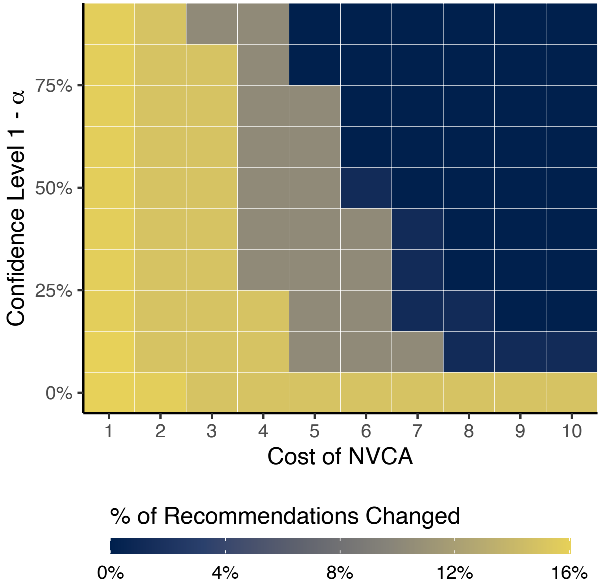
How the NVCA rules change



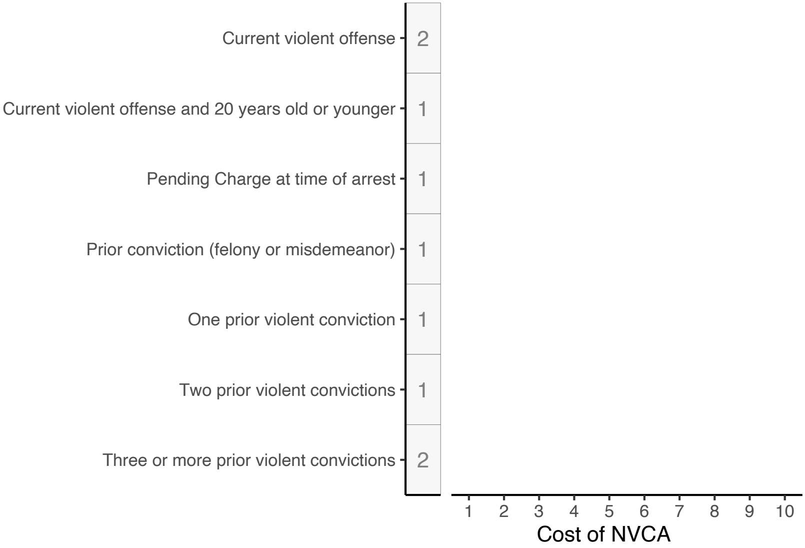
How the NVCA rules change



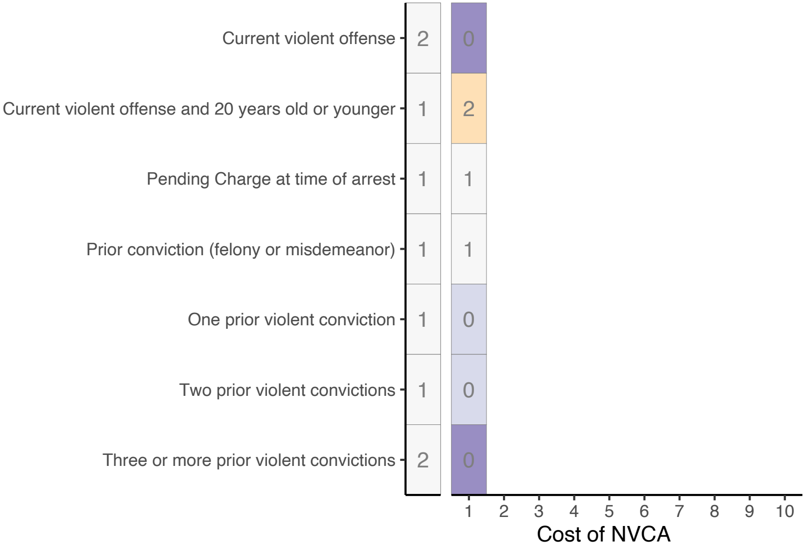
How the NVCA rules change



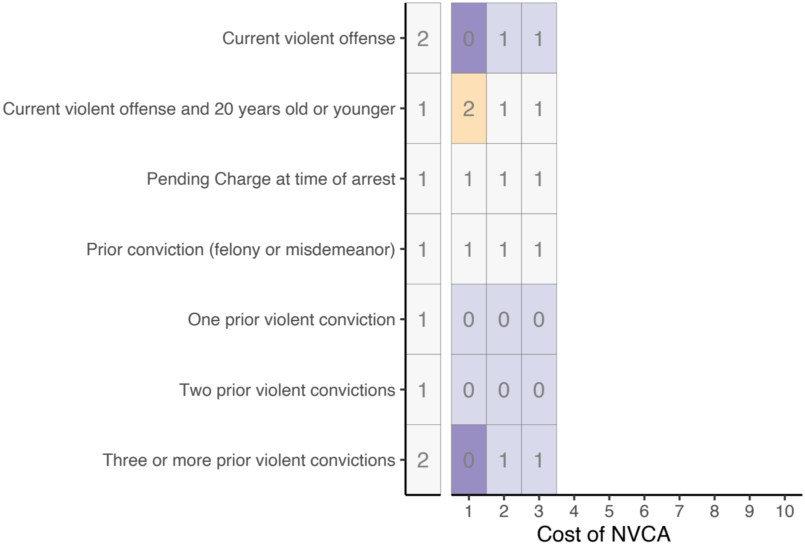
Changing weights on risk factors



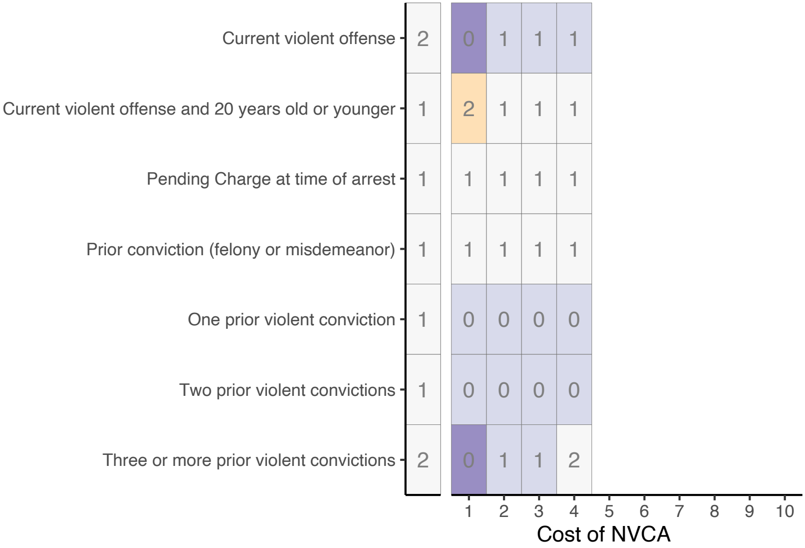
Changing weights on risk factors



Changing weights on risk factors



Changing weights on risk factors

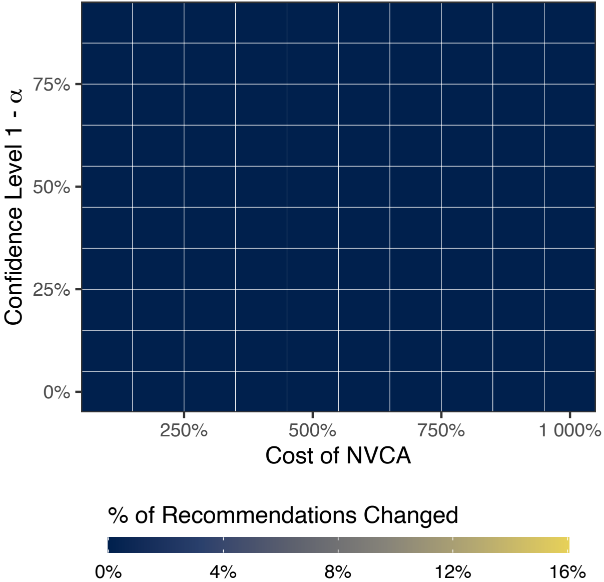


Changing weights on risk factors

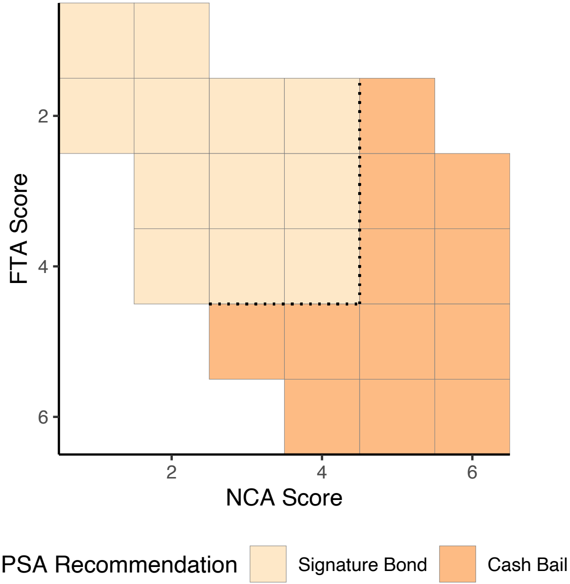
		1	2	3	4	5	6	7	8	9	10
Current violent offense	2	0	1	1	1	2	2	2	2	2	2
Current violent offense and 20 years old or younger	1	2	1	1	1	1	1	1	1	1	1
Pending Charge at time of arrest	1	1	1	1	1	1	1	1	1	1	1
Prior conviction (felony or misdemeanor)	1	1	1	1	1	1	1	1	1	1	1
One prior violent conviction	1	0	0	0	0	1	1	1	1	1	1
Two prior violent convictions	1	0	0	0	0	1	1	1	1	1	1
Three or more prior violent convictions	2	0	1	1	2	2	2	2	2	2	2

Cost of NVCA

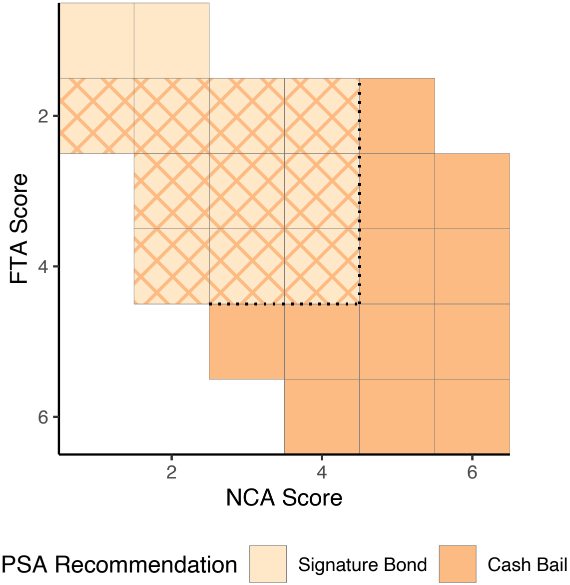
Too much noise to learn from Judges' decisions



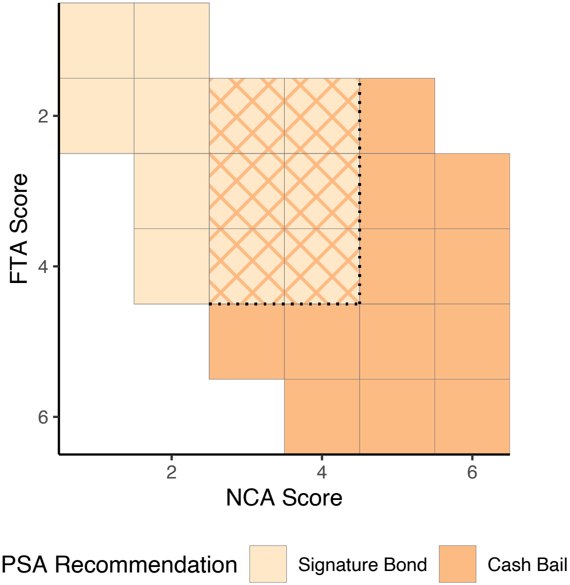
Identifying the DMF matrix with an additive model



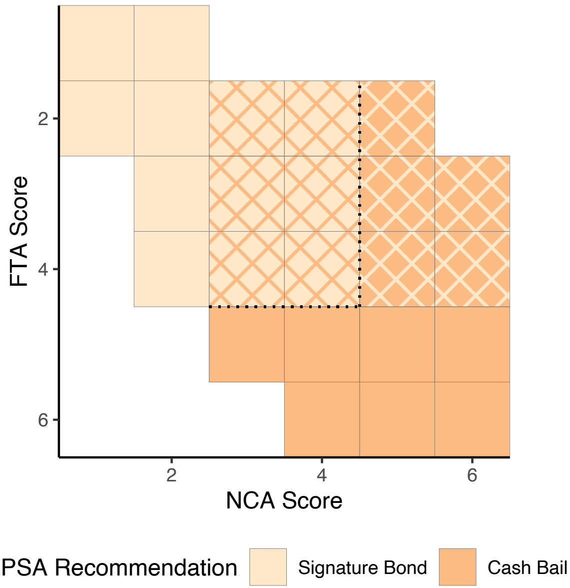
Identifying the DMF matrix with an additive model



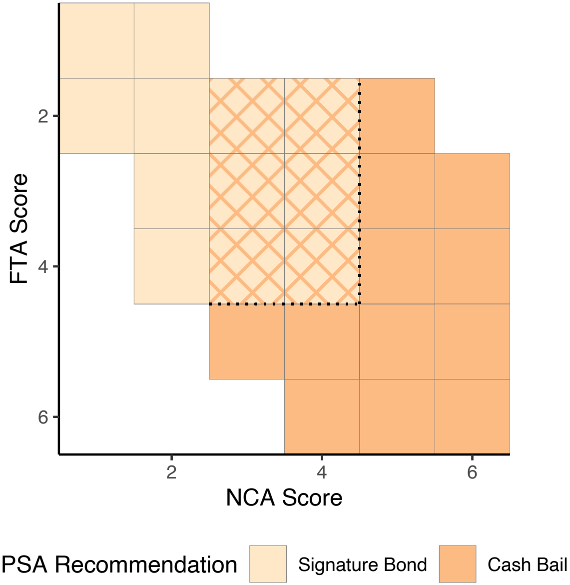
Identifying the DMF matrix with an additive model



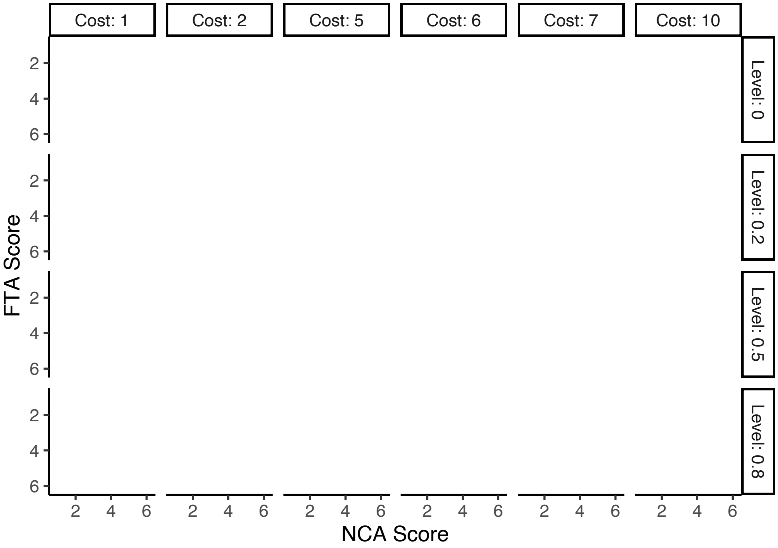
Identifying the DMF matrix with an additive model



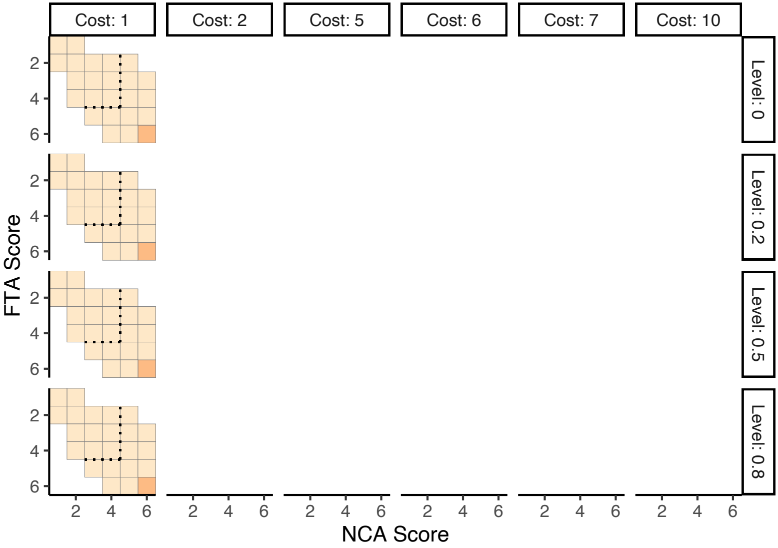
Identifying the DMF matrix with an additive model



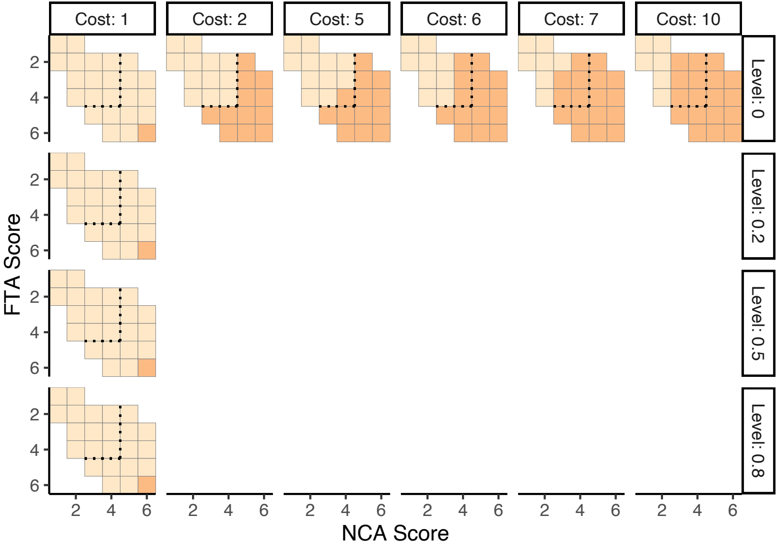
Learning a new monotone DMF matrix



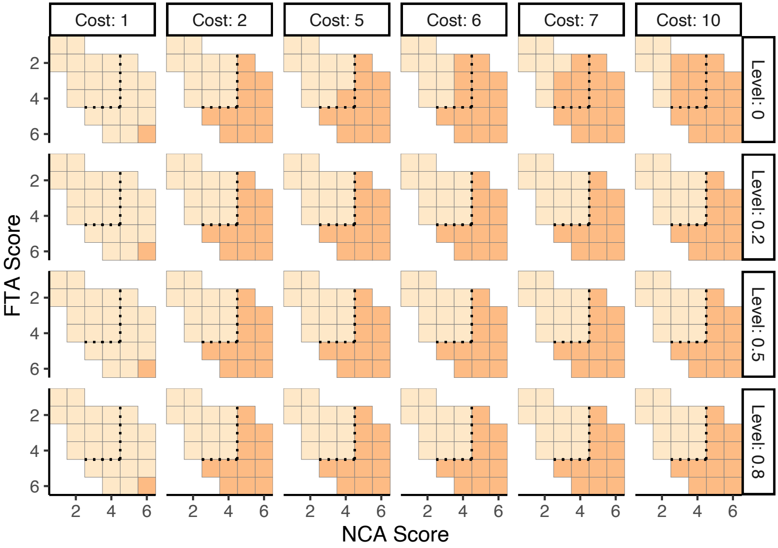
Learning a new monotone DMF matrix



Learning a new monotone DMF matrix



Learning a new monotone DMF matrix



Recap: Safe policy learning through extrapolation

Deterministic decision and recommendation algorithms touch many parts of our lives

- Generate a lot of data! But deterministic nature poses thorny identification issues

Recap: Safe policy learning through extrapolation

Deterministic decision and recommendation algorithms touch many parts of our lives

- Generate a lot of data! But deterministic nature poses thorny identification issues

This paper: Extrapolate and use robust optimization to learn a new algorithm

- Partially identify the counterfactuals and find the policy that is the best in the worst case
- Gives a statistical safety guarantee relative to the status quo
- Some evidence we can improve the PSA, but noisy. Need more data!

Recap: Safe policy learning through extrapolation

Deterministic decision and recommendation algorithms touch many parts of our lives

- Generate a lot of data! But deterministic nature poses thorny identification issues

This paper: Extrapolate and use robust optimization to learn a new algorithm

- Partially identify the counterfactuals and find the policy that is the best in the worst case
- Gives a statistical safety guarantee relative to the status quo
- Some evidence we can improve the PSA, but noisy. Need more data!

Many more questions on designing algorithms to assist human decision makers

- Asymmetric utility functions lead to unidentifiable objectives
- Optimizing for long term outcomes when we only can measure short term outcomes
- Learning policies when human decisions mediate future outcomes and decisions

Recap: Safe policy learning through extrapolation

Deterministic decision and recommendation algorithms touch many parts of our lives

- Generate a lot of data! But deterministic nature poses thorny identification issues

This paper: Extrapolate and use robust optimization to learn a new algorithm

- Partially identify the counterfactuals and find the policy that is the best in the worst case
- Gives a statistical safety guarantee relative to the status quo
- Some evidence we can improve the PSA, but noisy. Need more data!

Many more questions on designing algorithms to assist human decision makers

- Asymmetric utility functions lead to unidentifiable objectives
- Optimizing for long term outcomes when we only can measure short term outcomes
- Learning policies when human decisions mediate future outcomes and decisions

Thank you!

`ebenmichael.github.io`
`arxiv.org/abs/2109.11679`

Appendix

References I

Athey, S. and Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, 89(1):133–161.

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501.

Beygelzimer, A. and Langford, J. (2009). The offset tree for learning with partial labels. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 129–137.

Cui, Y. (2021). Individualized decision making under partial identification: three perspectives, two optimality results, and one paradox. *Harvard Data Science Review*. Just accepted.

Dudik, M. and Langford, J. (2011). Doubly Robust Policy Evaluation and Learning. In *Proceedings of the 28th International Conference on Machine Learning*.

Greiner, D. J., Halen, R., Stubenberg, M., and Chistopher L. Griffen, J. (2020). Randomized control trial evaluation of the implementation of the psa-dmf system in dane county. Technical report, Access to Justice Lab, Harvard Law School.

Hahn, P. R., Murray, J. S., and Carvalho, C. M. (2020). Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects. *Bayesian Analysis*, pages 1–33.

References II

Imai, K., Jiang, Z., Greiner, D. J., Halen, R., and Shin, S. (2020). Experimental Evaluation of Computer-Assisted Human Decision-Making: Application to Pretrial Risk Assessment Instrument. arxiv preprint `https://arxiv.org/pdf/2012.02845.pdf`.

Kallus, N. (2018). Balanced policy evaluation and learning. *Advances in Neural Information Processing Systems*, 2018-December(1):8895–8906.

Kallus, N. and Zhou, A. (2021). Minimax-optimal policy learning under unobserved confounding. *Management Science*, 67(5):2870–2890.

Kitagawa, T. and Tetenov, A. (2018). Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, 86(2):591–616.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10):4156–4165.

Luedtke, A. R. and Van Der Laan, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics*, 44(2):713–742.

Manski, C. F. (2005). *Social Choice with Partial Knowledge of Treatment Response*. Princeton University Press.

References III

Nie, X. and Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319.

Pu, H. and Zhang, B. (2021). Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *Journal of the Royal Statistical Society Series B*, pages 1–28.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210.

Swaminathan, A. and Joachims, T. (2015). Batch learning from logged bandit feedback through counterfactual risk minimization. *Journal of Machine Learning Research*, 16:1731–1755.

Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. (2012). Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.