Safe Policy Learning through Extrapolation: Application to Pre-trial Risk Assessment

Kosuke Imai

Harvard University

Econometrics and Applied Micro Seminar
European University Institute
October 10, 2022

Joint work with Eli Ben-Michael, D. James Greiner, and Zhichao Jiang

Motivation

- Widespread use of algorithmic recommendation and decisions
- Fast growing literature on policy learning

- High-stake algorithmic recommendations/decisions in medicine and public policy
  - need for transparency and accountability
  - simple and deterministic rules

- Question: How can we learn new and better policies using the data based on existing deterministic policies?
- Prior policy learning methods require existing policies to be stochastic

- Goal: Develop a safe approach to policy learning through extrapolation

Overview of the Talk

- Methodology
  - find an improved policy over the status quo policy
  - maximize the expected utility in the worst case

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

  - robust optimization approach based on partial identification
  - statistical safety guarantee: limiting the probability for yielding a worse outcome than the existing policy
- Application
  - pre-trial risk assessment instruments in the US criminal justice system
  - experimental evaluation in Dane county, Wisconsin
  - learn new algorithmic scoring and recommendation rules while maintaining the transparency and structure of the existing rules
  - Imai *et al*. (2022). "Experimental Evaluation of Algorithm-Assisted Human Decision-Making: Application to Pretrial Public Safety Assessment." (with discussion) *Journal of the Royal Statistical Society, Series A* https://arxiv.org/pdf/2012.02845.pdf

Pretrial Public Safety Assessment (PSA)

- Algorithmic recommendations often used in US criminal justice system
- At the first appearance hearing, judges primarily make two decisions
  1. whether to release an arrestee pending disposition of criminal charges
  2. what conditions (e.g., bail and monitoring) to impose if released
- Goal: avoid predispositional incarceration while preserving public safety

- Judges are required to consider three risk factors along with others
  1. arrestee may fail to appear in court (FTA)
  2. arrestee may engage in new criminal activity (NCA)
  3. arrestee may engage in new violent criminal activity (NVCA)

- PSA as an algorithmic recommendation to judges
  - classifying arrestees according to FTA and NCA/NVCA risks
  - derived from an application of a machine learning algorithm to a training data set based on past observations
  - different from COMPAS score

A Field Experiment for Evaluating the PSA

- Dane County, Wisconsin
- PSA = weighted indices of ten factors
  - age as the single demographic factor: no gender or race
  - nine factors drawn from criminal history (prior convictions and FTA)
- PSA scores and recommendation
  1. two separate ordinal six-point risk scores for FTA and NCA
  2. one binary risk score for new violent criminal activity (NVCA)
  3. aggregate recommendation: signature bond, small and large cash bond
- Judges may have other information about an arrestee
  - affidavit by a police officer about the arrest
  - defense attorney may inform about the arrestee's connections to the community (e.g., family, employment)
- Field experiment
  - clerk assigns case numbers sequentially as cases enter the system
  - PSA is calculated for each case using a computer system
  - if the first digit of case number is even, PSA is given to the judge
  - mid-2017 – 2019 (randomization), 2-year follow-up for half sample

DANE COUNTY CLERK OF COURTS
Public Safety Assessment – Report

215 S Hamilton St #1000
Madison, WI 53703
Phone: (608) 266-4311

Name: [redacted]	**Spillman Name Number:** [redacted]
DOB: [redacted]	**Gender:** Male
Arrest Date: 03/25/2017	**PSA Completion Date:** 03/27/2017

New Violent Criminal Activity Flag

No

New Criminal Activity Scale

1	2	3	4	5	6

Failure to Appear Scale

1	2	3	4	5	6

Charge(s):

961.41(1)(D)(1) MFC DELIVER HEROIN <3 GMS F 3

Risk Factors:	**Responses:**
1. Age at Current Arrest	23 or Older
2. Current Violent Offense	No
a. Current Violent Offense & 20 Years Old or Younger	No
3. Pending Charge at the Time of the Offense	No
4. Prior Misdemeanor Conviction	Yes
5. Prior Felony Conviction	Yes
a. Prior Conviction	Yes
6. Prior Violent Conviction	2
7. Prior Failure to Appear Pretrial in Past 2 Years	0
8. Prior Failure to Appear Pretrial Older than 2 Years	Yes
9. Prior Sentence to Incarceration	Yes

Recommendations:

Release Recommendation - Signature bond

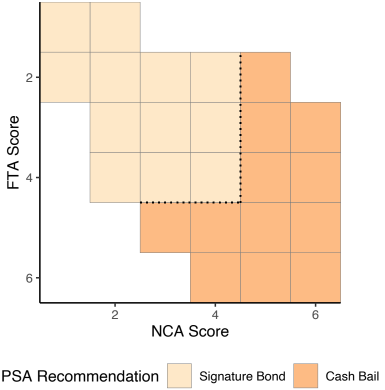
Conditions - Report to and comply with pretrial supervision

PSA Scoring Rule

Risk factor		FTA	NCA	NVCA
Current violent offense	$>$ 20 years old			2
	$\leq$ 20 years old			3
Pending charge at time of arrest		1	3	1
Prior conviction	misdemeanor or felony	1	1	1
	misdemeanor and felony	1	2	1
Prior violent conviction	1 or 2		1	1
	3 or more		2	2
Prior sentence to incarceration			2	
Prior FTA in past 2 years	only 1	2	1	
	2 or more	4	2	
Prior FTA older than 2 years		1		
Age	22 years or younger		2	

- FTA: $\{0 \to 1, 1 \to 2, 2 \to 3, (3,4) \to 4, (5,6) \to 5, 7 \to 6\}$
- NCA: $\{0 \to 1, (1,2) \to 2, (3,4) \to 3, (5,6) \to 4, (7,8) \to 5, (9,10,11,12,13) \to 6\}$
- NVCA: $\{(0,1,2,3) \to 0, (4,5,6,7) \to 1\}$

Decision Making Framework (DMF)



	No NVCA	NVCA	Total
Signature Bond	1130	80	1410
Cash Bail	452	29	481
Total	1782	109	1891

Setup

- For each individual i, observe
  - Covariates $X_i \in \mathcal{X}$
  - Action taken $A_i \in \mathcal{A}$
  - Binary outcome $Y_i \in \{0, 1\}$
- 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 $\mathcal{X}_a = \{x \in \mathcal{X} \mid \tilde{\pi}(x) = a\}$
- Cost of actions and utility of outcomes

$$\underbrace{c(a)}_{\text{cost}} + \underbrace{u}_{\text{utility}} Y(a)$$

Identification Problem

- Goal: Find a policy with high expected utility (value/welfare)

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

  where $\pi(a \mid X) = 1\{\pi(X) = a\}$

- But how do we identify the counterfactuals?

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

- Existing work uses stochastic policies for identification
  - inverse probability weighting

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

  - outcome model imputation, and double robust methods as well

Decomposition and Maxmin Principle

- Decompose the value into identifiable and unidentifiable components

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

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

- Partially identify $m \in \mathcal{M}$, then find the best policy in the worst case

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

- This is a safe policy based on robust optimization
  - Conservative, "pessimistic" principle
  - Falls back on the status quo policy if there is too much uncertainty

Partial Identification

- To partially identify the conditional expectation $\mathbb{E}[Y(a) \mid X = x]$
  1. Put restrictions on the class of possible models
  2. Compute the set of functions f in the selected model class that agree with the observable data

$$\mathcal{M} = \{f \in \mathcal{F} \mid f(\tilde{\pi}(x), x) = \mathbb{E}[Y \mid X = x] \; \forall x \in \mathcal{X}\}$$

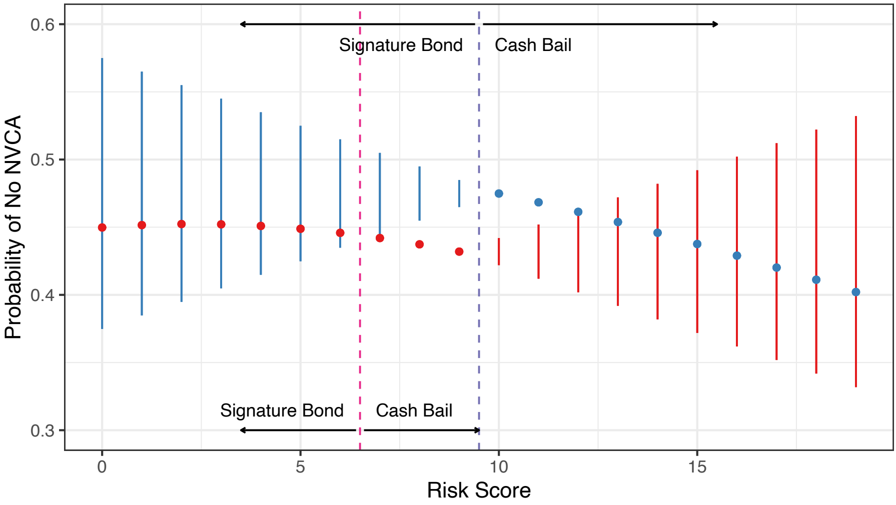
- Many model classes result in pointwise bounds

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

- Examples: Lipschitz functions, additive models, linear models
- Use the worst-case bound in place of the missing counterfactual:

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

Illustration with Single Discrete Covariate



Population Safe Policy

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

$$\underbrace{V\left(\tilde{\pi}\right) - V\left(\pi^{\text{inf}}\right)}_{\text{regret relative to baseline}} \leq 0$$

- Safety comes at the cost of a potentially suboptimal policy
- Compare to oracle policy $\pi^* \in \text{argmax}_{\pi \in \Pi} V(\pi)$

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

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

- The tighter the partial identification, the smaller the optimality gap

- 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

- Conservative approach gives a statistical safety guarantee with level α

Value is probably, approximately at least as high as baseline

$$V(\tilde{\pi}) - V(\hat{\pi}) \lesssim \text{Complexity}(\Pi)$$

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

- If policy class Π is complex, need more samples to avoid overfitting

Empirical optimality gap controlled by the size of the empirical model class and the complexity of policy class

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

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

- Same tradeoff between safety and optimality

Extensions

1. Incorporating experiments evaluating a deterministic policy
    - The control condition is 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 may be simpler than outcomes $\mathbb{E}[Y(a) \mid X = x]$
2. Incorporating human decisions with algorithmic recommendations
    - Incorporate uncertainty in judge's potential decision $D(a)$
    - Two unidentified components: outcomes and decisions

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

    - Need to find the worst case potential decision and outcome

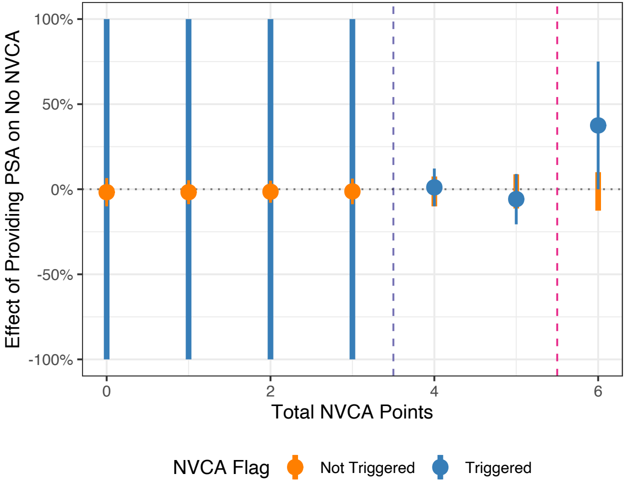
Learning a new NVCA Flag Threshold

- Find an improved NVCA flag threshold using the same risk factors
  - Status quo policy: $\tilde{\pi}(x_{\text{nvca}}) = 1\{x_{\text{nvca}} \geq 4\}$ where $x_{\text{nvca}} \in \{0, 1, \ldots, 6\}$
  - Policy class: $\Pi_{\text{thresh}} = \{\pi(x) = 1\{x_{\text{nvca}} \geq \eta\} \mid \eta \in \{0, \ldots, 7\}\}$

- Lipschitz constraint on the CATE $\tau(a, x_{\text{nvca}})$
- The Working–Hotelling–Scheffé simultaneous confidence intervals

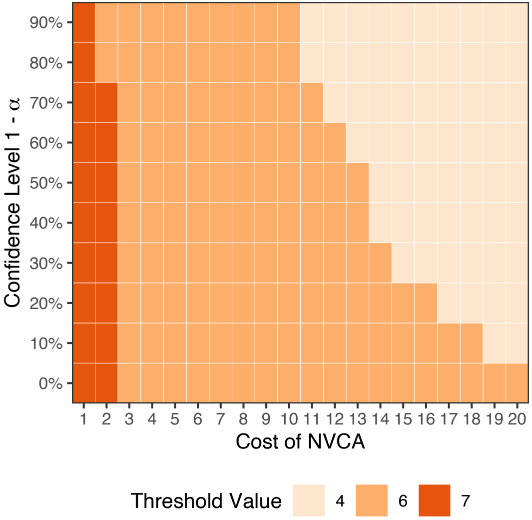
- Cost of triggering the NVCA flag is 1: $c(0) = 0$ and $c(1) = -1$
- Monetary cost is zero, but fiscal costs on jurisdiction and socioeconomic costs on individuals and community
- Equal utility $u(1) = u(0) = u$: cost of an NVCA is $-u$

Extrapolating the CATE



- More information when extrapolating the CATE for the case that the NVCA flag is *not triggered*

New NVCA Thresholds



- Higher cost of NVCA and greater confidence ⇝ fall back on the status quo policy

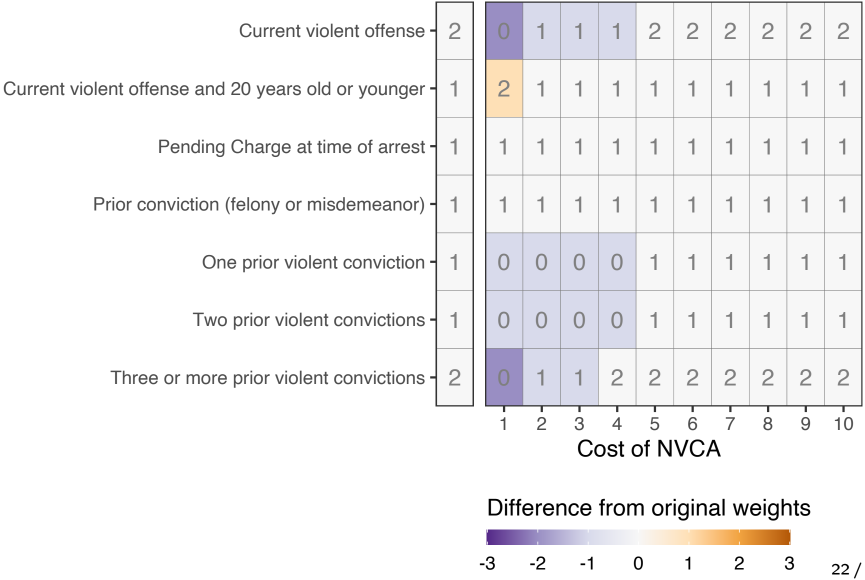
Learning a New NVCA Flag Point System

- Changing the integer weights applied to 7 binary risk factors
  - Status quo policy: $\tilde{\pi}(x) = 1\left\{\sum_{j=1}^{7} \tilde{\theta}_j x_j \geq 4\right\}$ where $x \in \{0,1\}^7$
  - Policy class: $\Pi_{\text{int}} = \left\{\pi(x) = 1\left\{\sum_{j=1}^{7} \theta_j x_j \geq 4\right\} \mid \theta_j \in \mathbb{Z}\right\}$

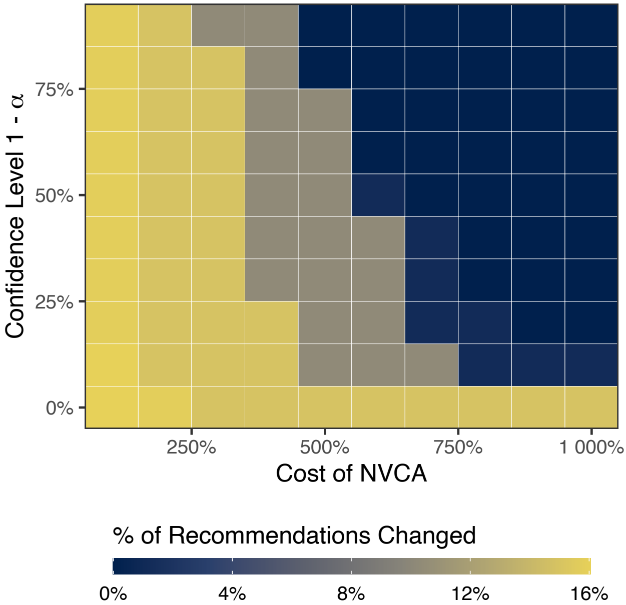
- Two model classes:
  1. additive models
  2. second-order interactive models
- For the outcome $m(a,x)$ as well as for the CATE $m(a,x) - m(\emptyset, x)$

Changes in the NVCA Flag Weights

(addtive effect model; confidence level = 80%)



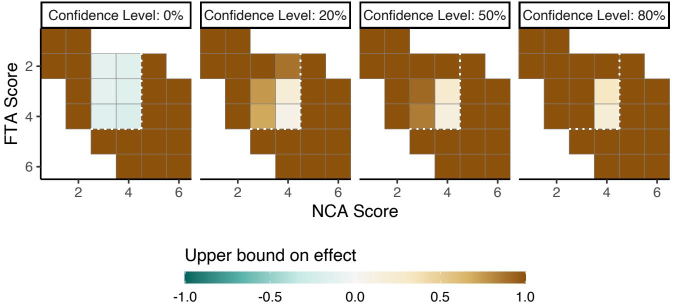
Changes over the Status Quo Policy



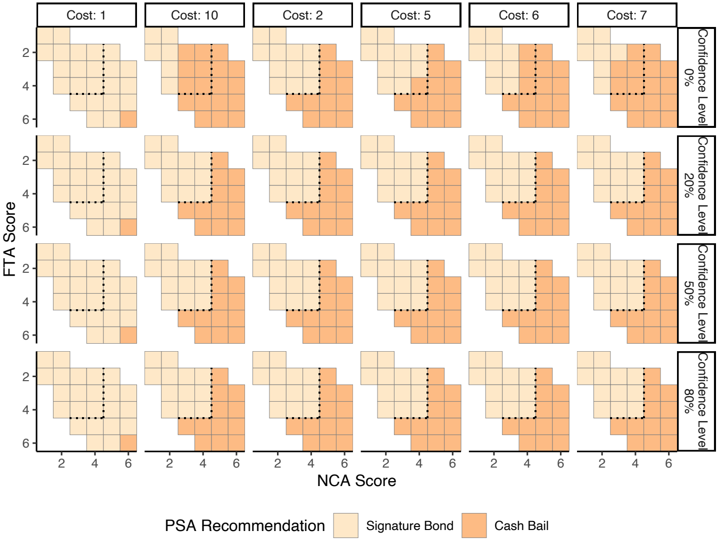
Learning a New DMF Matrix

- Aggregates the FTA and NCA scores into a single recommendation
  - two 6-point ordinal scores $(x_{\text{fta}}, x_{\text{nca}}) \in \{1, \ldots, 6\}^2$
  - additive treatment effect models:
    $\tau_{\text{add}}(a, x) = \tau_{\text{fta}}(a, x_{\text{fta}}) + \tau_{\text{nca}}(a, x_{\text{nca}})$
  - policy class: monotonically increasing in both x_{fta} and x_{nca}

- Upper bound on the treatment effect under the additive effect models



Changes in the DMF Matrix



Concluding Remarks

- Deterministic decisions and recommendation algorithms are ubiquitous
  - government policies and medical treatment decisions
  - transparency and simplicity
- Proposed methodology: extrapolate and use robust optimization to learn a new policy
  - 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!
- Potential applications:
  - Policy learning with asymmetric utility functions
  - Optimizing for long term outcomes using short term outcomes