Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

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

Eli Ben-Michael, D. James Greiner, Kosuke Imai & Zhichao Jiang

To cite this article: Eli Ben-Michael, D. James Greiner, Kosuke Imai & Zhichao Jiang (2025) Safe Policy Learning through Extrapolation: Application to Pre-trial Risk Assessment, Journal of the American Statistical Association, 120:551, 1386-1399, DOI: 10.1080/01621459.2025.2489135

To link to this article: https://doi.org/10.1080/01621459.2025.2489135

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, VOL. 120, NO. 551, 1386–1399: Applications and Case Studies
https://doi.org/10.1080/01621459.2025.2489135

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

Eli Ben-Michael[a], D. James Greiner[b], Kosuke Imai[c], and Zhichao Jiang[d]

[a]Department of Statistics & Data Science and Heinz College of Information Systems & Public Policy, Carnegie Mellon University, Pittsburgh, PA; [b]Honorable S. William Green Professor of Public Law, Harvard Law School, Cambridge, MA; [c]Department of Government and Department of Statistics, Harvard University, Institute for Quantitative Social Science, Cambridge, MA; [d]Professor, School of Mathematics, Sun Yat-sen University, Guangzhou Guangdong, China

ABSTRACT

Algorithmic recommendations and decisions have become ubiquitous in today's society. Many of these data-driven policies, especially in the realm of public policy, are based on known, deterministic rules to ensure their transparency and interpretability. We examine a particular case of algorithmic pre-trial risk assessments in the US criminal justice system, which provide deterministic classification scores and recommendations to help judges make release decisions. Our goal is to analyze data from a unique field experiment on an algorithmic pre-trial risk assessment to investigate whether the scores and recommendations can be improved. Unfortunately, prior methods for policy learning are not applicable because they require existing policies to be stochastic. We develop a maximin robust optimization approach that partially identifies the expected utility of a policy, and then finds a policy that maximizes the worst-case expected utility. The resulting policy has a statistical safety property, limiting the probability of producing a worse policy than the existing one, under structural assumptions about the outcomes. Our analysis of data from the field experiment shows that we can safely improve certain components of the risk assessment instrument by classifying arrestees as lower risk under a wide range of utility specifications, though the analysis is not informative about several components of the instrument. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received May 2023
Accepted March 2025

KEYWORDS
Algorithm-assisted decision-making; Decision-making under uncertainty; Optimal policy learning; Risk assessments; Robust optimization

1. Introduction

Algorithmic recommendations and decisions are ubiquitous in our daily lives. Many algorithmic policies are used for consequential decisions in high stakes settings such as criminal justice, social policy, and medical care. One common feature of such policies is that they are based on known, deterministic rules. This is often because transparency and interpretability are required to ensure accountability especially when algorithms are used for public policy-making.

In this article, we focus on a particular case: pre-trial risk assessment instruments (PRAI) in the American criminal justice system. The goal of a PRAI is to aid judges in deciding which arrestees should be released pending the disposition of any criminal charges. We consider a particular PRAI used in Dane County, Wisconsin, which includes the state capital, Madison (Section 2). This PRAI assigns scores to arrestees according to the risk that they are predicted to engage in undesirable behavior. It then aggregates these scores using a deterministic function and provides an overall release recommendation to the judge.

We analyze data from a unique field experiment on the PRAI (Greiner et al. 2020; Imai et al. 2023). Our goal is to learn new algorithmic scoring and recommendation rules that can lead to better overall outcomes while retaining the transparency of the existing instrument. Importantly, we focus on changing the algorithmic policies, which we can intervene on, rather than judge's decisions, which we cannot.

The large amounts of data collected after implementing *deterministic* policies such as PRAIs provide an opportunity to learn new policies that improve on the status quo. Unfortunately, prior approaches to policy learning are not applicable because they require existing policies to be *stochastic*, typically relying on inverse probability weighting (Section 3).

To address this challenge (Section 4), we partially identify the expected utility of a policy by calculating all potential values consistent with the observed data. This makes choosing an "optimal" policy ambiguous: a policy can perform well under some outcome models that are consistent with the data and poorly in others. We use the maximin criterion that finds a policy that maximizes the worst-case performance relative to the status quo. The resulting policy has a statistical *safety* property that limits the probability of yielding a worse outcome than the status quo policy, under the structural assumptions made about the outcomes. However, this safety property comes at the cost of potentially choosing a sub-optimal policy, though it is no worse than the status quo. We formally characterize the gap between this safe policy and the infeasible oracle policy.

We use this approach to explore whether the data from our field experiment support alterations to the existing PRAI (Section 5). We explore the three risk measures based on the

CONTACT Eli Ben-Michael *ebenmichael@cmu.edu* Department of Statistics & Data Science and Heinz College of Information Systems & Public Policy, Carnegie Mellon University, 4800 Forbes Avenue, Hamburg Hall, Pittsburgh PA 15213.

predicted likelihood that an arrestee, upon release, will (i) fail to appear in court (FTA), (ii) engage in new criminal activity (NCA), or (iii) engage in new violent criminal activity (NVCA). We also inspect the algorithm that recommends to the judge the level of cash bail and pre-trial supervision and monitoring conditions to impose.

We find that under several specifications of the utility function, it can be possible to improve safely upon the existing NVCA scoring rule by classifying arrestees as lower risk. However if the policy maker is primarily focused on avoiding NVCAs, the resulting safe policy falls back on the existing scoring rule. Our approach has limitations. Conducting our analysis requires several nontrivial choices that may be challenging in practice. In addition, our analysis does not provide meaningful insights about components of the instrument other than the NVCA scoring rule. This arises from identifiability issues caused by the structure of the underlying rules, as well as a high degree of statistical uncertainty due to small sample sizes for rare combinations of risk factors. We discuss these and other limitations in Section 6.

2. Pre-Trial Risk Assessment

We now briefly describe the particular PRAI, called the Public Safety Assessment (PSA), used in Dane County, Wisconsin. The PSA is an algorithmic recommendation designed to help judges make their pre-trial release decisions. We will also describe an original randomized experiment we conducted to evaluate the impact of the PSA on judges' decisions. In Section 5, we analyze this experimental dataset and consider how to improve outcomes by modifying certain aspects of the PSA system. Interested readers should consult Greiner et al. (2020) and Imai et al. (2023) for further details of the PSA and experiment; the study dataset has been made publicly available.

2.1. The PSA-DMF System

The goal of the PSA is to help judges decide, at first appearance hearings, whether to allow an arrestee's release without bail or release them only if the arrestee posts bail/bond (or meets other conditions). Because arrestees are presumed to be innocent, judges must avoid unnecessary incarceration. The PSA has several outputs. First, it returns three classification scores based on the predicted risk that each arrestee will engage in an FTA, NCA, or NVCA. Law requires judges to balance between these risks and the cost of incarceration. These three PSA scores are then combined via the so-called "Decision Making Framework" (DMF) into two overall recommendations: (i) whether to require a signature bond (i.e., release on their own recognizance) or some level of cash bail for release, and (ii) what, if any, monitoring conditions to place on release. Given the complexity of the system, our empirical analysis will focus on the question of how to improve each component separately (see Section 5).

FTA, NCA, and NVCA risk scores. These scores are deterministic functions of eight risk factors. The only demographic factor is the arrestee's age, and neither gender nor race is used. The other risk factors include the current offense and pending charges as well as measures of criminal history based on prior convictions and prior FTAs. These scores are constructed by assigning an integer-valued weight to each present risk factor, adding them together, and thresholding this value into a number of bins. For the sake of transparency, the foundation that funded the PSA's creation made these weights and thresholds publicly available (see https://advancingpretrial.org/psa/factors; Appendix Table G.1 summarizes the weights).

The FTA score has six levels and is based on four risk factors. The values range from 0 to 7, and the final score is thresholded into values between 1 (lowest risk) and 6 (highest risk) by assigning $\{0 \to 1, 1 \to 2, 2 \to 3, (3,4) \to 4, (5,6) \to 5, 7 \to 6\}$. The NCA score also has six levels, but is based on six risk factors and has a maximum value of 13 before being collapsed into six levels by assigning $\{0 \to 1, (1,2) \to 2, (3,4) \to 3, (5,6) \to 4, (7,8) \to 5), (9,10,11,12,13) \to 6\}$. Finally, the NVCA score is a binary flag based on five risk factors: if the sum of the weights is greater than or equal to 4, the PSA flags the arrestee as being at elevated risk of an NVCA. Otherwise, the NVCA score is 0, and the arrestee is not flagged as being at elevated risk.

Recommendations via the DMF. Next, the DMF transforms these three PSA risk scores into a recommendation regarding cash bail and one regarding additional monitoring conditions. For cases where the current charge is one of several serious violent offenses, the defendant was extradited, or the NVCA score is 1, the DMF automatically recommends cash bail with maximum supervision and monitoring conditions. For the remaining cases, the FTA and NCA risk scores are combined into one of 7 overall risk levels. If the FTA and NCA scores are both less than 5, and so the risk level is 3 or lower, then the recommendation is to only require a signature bond. Otherwise the recommendation is to require cash bail (limited to "modest" at levels 4–5 and "moderate" at level 6). Figure 1 visualizes the cash bail portion of the DMF. The risk levels similarly encode a recommendation for an increasing amount of pre-trial supervision and monitoring conditions, ranging from none (level 1) to maximum supervision with biweekly phone and face-to-face contacts (level 7). Appendix Figure G.10 shows these conditions along with the cash bail recommendations.

2.2. The Experimental Data

We analyze the data from a randomized controlled trial conducted in Dane County, Wisconsin. In this experiment, the PSA was computed for each first appearance hearing that a single judge oversaw during the study period. Across cases, we randomized whether the PSA was made available in its entirety to the judge. If a case is assigned to the treatment group, the judge received the three PSA scores, the DMF recommendations, and all of the risk factors that were used to construct them on a single sheet of paper. For the control group, the judge did not receive the PSA scores and DMF recommendations. Since the risk factors that go into the PSA were made available in other case files, the judge could, in principle, reconstruct the PSA output with enough time.

For each case, we observe the three scores (FTA, NCA, and NVCA) and the DMF recommendation, the underlying risk

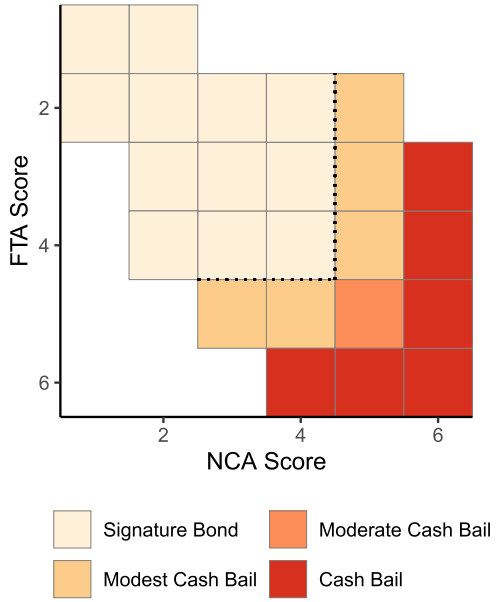

Figure 1. Decision Making Framework (DMF) matrix for cases where the current charge is not a serious violent offense, the NVCA flag is not triggered, and the defendant was not extradited. If the FTA score and the NCA score are both less than 5, then the recommendation is to only require a signature bond. Otherwise the recommendation is to require some amount of cash bail. The dashed line indicates this boundary. Unshaded areas indicate impossible combinations of FTA and NCA scores.

factors used to construct the scores, the binary decision by the judge (signature bond or cash bail), and three binary outcomes (FTA, NCA, and NVCA). We focus on first arrest cases in order to avoid spillover effects between cases. All told, there are 1,891 cases, 948 of which the judge was given access to the PSA.

Our goal is to improve the PSA recommendation system while taking into account the judicial decisions that partly result from the algorithmic recommendations; see Appendix D for further discussion on incorporating judicial decisions into the analysis. Crucially, each component of the PSA is *deterministic* and no aspect of it was randomized as part of the study. Therefore, there is a lack of *overlap*: the probability that any case would have had a different algorithmic recommendation than it actually received is exactly zero. This makes existing approaches to policy learning inapplicable because they rely on the inverse of this probability. Instead, learning a new recommendation policy in the absence of overlap requires *extrapolation*. Below, we will develop a methodological framework that provides a statistical property that the new, learned rules perform at least as well as the original recommendation.

3. Policy Learning with Observational Data

3.1. Notation and Setup

Suppose that we have a representative sample of n units independently drawn from a population $\mathcal{P}$. For each unit $i = 1, \ldots, n$, we observe a set of covariates $\boldsymbol{X}_i \in \mathcal{X} \subseteq \mathbb{R}^p$ (e.g., the risk factors from Appendix Table G.1) and a binary outcome $Y_i \in \{0, 1\}$. In our analysis presented in Section 5, we alternately consider the outcome $Y_i = 1$ as the *absence* of an FTA, NCA, or NVCA. We consider a set of K possible actions, denoted by $\mathcal{A} = \{0, 1, 2, \ldots, K-1\}$ that can be taken for each unit.

The actions correspond to the PSA recommendation: there are $K = 6$ possible actions when we consider the FTA and NCA risk scores, $K = 2$ for the NVCA flag, $K = 7$ for the overall DMF bail and monitoring recommendation, and $K = 2$ for the signature bond versus cash bail recommendation. In our experimental evaluation, we have access to the algorithm that generated the observed actions. Formally, we encode this as a known baseline deterministic policy $\tilde{\pi} : \mathcal{X} \to \mathcal{A}$ that generates the observed actions $A_i = \tilde{\pi}(\boldsymbol{X}_i)$. Throughout this article, we will also refer to the baseline policy as $\tilde{\pi}(\boldsymbol{x}, a) \equiv \mathbb{1}\{\tilde{\pi}(\boldsymbol{x}) = a\}$, the indicator of whether the baseline policy yields action a given the covariates $\boldsymbol{x}$.

We consider the effects of the algorithmic recommendation on the outcome, and assume that the algorithmic action A_i may affect its own unit's outcome Y_i but has no impact on the outcomes of other units (no interference between units; Rubin 1980). Then, we can write the potential outcome under each action $A_i = a$ as $Y_i(a)$ where $a \in \mathcal{A}$ and the observed outcome as $Y_i = Y_i(A_i) = Y_i(\tilde{\pi}(\boldsymbol{X}_i))$ (Neyman 1923). This setup focuses on the impacts of the algorithmic recommendation whose provision was randomized in our experimental evaluation. We marginalize over the potential human judicial decisions that may be influenced by the algorithmic recommendation (see Appendix D for further formalization). Finally, our setting implies that $(\{Y_i(a)\}_{a \in \mathcal{A}}, \boldsymbol{X}_i)$ are independent and identically distributed, so we sometimes drop the i subscript.

3.2. Optimal Policy Learning

Our primary goal is to find a new deterministic policy $\pi : \mathcal{X} \to \mathcal{A}$, that has a high expected utility. We will again use the notation $\pi(\boldsymbol{x}, a) \equiv \mathbb{1}\{\pi(\boldsymbol{x}) = a\}$ for the policy being equal to action a given the covariates $\boldsymbol{x}$. Let $u(y, a)$ denote the utility for outcome y under action a. Because the outcomes are binary, we can write this utility function as:[1]

$$\begin{aligned} &Y(a)u(1,a) + \{1 - Y(a)\}u(0,a) \\ &\quad = \{u(1,a) - u(0,a)\}Y(a) + u(0,a). \end{aligned}$$

The two key components of this utility function are (i) the utility change between the two outcomes for action a, $u(a) \equiv u(1, a) - u(0, a)$, which we assume is nonnegative without loss of generality, and (ii) the utility for an outcome of zero with an action a, $c(a) \equiv u(0, a)$. We will refer to the latter term as the "cost" because it denotes the utility under action a when the outcome event does not happen; $c(a) = 0$ corresponds to the action having no cost. We define the utility using both the outcome y and the action a to capture the fact that some actions are costly. For example, in Section 5, we will place a cost on triggering the NVCA flag, recommending cash bail, or assigning a high NCA, FTA, or overall risk score. We note, however, that our approach is agnostic to the particular choice of the utility function.

While this utility only takes into account the policy action and the outcome, policy makers may also be concerned about

[1] While we focus here on binary potential outcomes, this form of the utility function shows that we can extend our results to the case with continuous outcomes with utility functions that are linear in the (possibly transformed) outcomes.

the costs of subsequent human decisions that are possibly affected by algorithmic recommendations or actions. In Appendix D, we show how to incorporate such factors into the utility function.

The value of policy π is the expected utility under policy π in the population,

$$\begin{aligned} V(\pi, m^*) &= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(\boldsymbol{X}, a)\{u(a)Y(a) + c(a)\}\right] \\ &= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(\boldsymbol{X}, a)\{u(a)m^*(a, \boldsymbol{X}) + c(a)\}\right], \end{aligned} \tag{1}$$

where we have used the law of iterated expectations, with the first expection over $\boldsymbol{X}$ and $Y(a)$, and the second expectation over $\boldsymbol{X}$, to show the dependence on the true conditional expected potential outcome function $m^*(a, \boldsymbol{x}) \equiv \mathbb{E}[Y(a) \mid \boldsymbol{X} = \boldsymbol{x}]$. We explicitly denote the value under different potential models for our development below; in cases where it is not ambiguous, we omit the m^* argument to indicate the value under the true conditional expected potential outcome function.

Ideally, we would like to find a policy π that has the highest value within a policy class Π. We can write a population optimal policy as one that maximizes the value, that is, $\pi^* \in \text{argmax}_{\pi\in\Pi} V(\pi)$. The policy class Π is an important object both in the theoretical analysis and in applications. In Section 5, we discuss the substantive choice of policy class when applied to a PRAI.

To find an optimal policy, we need to point-identify the value $V(\pi, m^*)$ for all candidate policies $\pi \in \Pi$. Existing methods rely on an overlap assumption for identification. In our context, this would require that each case has a nonzero probability of being assigned algorithmic action $A = a$, that is $P(A = a \mid \boldsymbol{X}) > 0$ for all $a \in \mathcal{A}$. If the baseline policy were stochastic, satisfying the overlap assumption, we could directly use inverse probability weighting, model-based weighting, or a doubly robust approach to learn an optimal policy from data (e.g., Dudik, Langford, and Li 2011; Qian and Murphy 2011; Zhao et al. 2012; Kitagawa and Tetenov 2018; Athey and Wager 2021). In our application and many other settings, however, the baseline policy $\tilde{\pi}$ is a deterministic function of covariates, implying a lack of overlap. Thus, we cannot point-identify the value $V(\pi, m^*)$ for all policies $\pi \in \Pi$ and hence cannot use existing approaches. In Appendix E, we provide further discussion about this identification issue.

4. Safe Policy Learning through Extrapolation

To deal with the lack of overlap brought on by the deterministic policy, we propose to first partially identify the conditional expectation, and then use robust optimization to find the best policy under the worst-case model. We will develop our optimal safe policy approach in two parts. First, we show how to construct a safe policy if we had access to an infinite number of samples, that is, in the population. We then discuss how to construct policies empirically from data, and establish finite-sample statistical properties of the policies. Finally, we show how to incorporate the experimental control units to weaken the assumptions of our general approach and discuss the practical implementation of the procedure for our analysis.

4.1. Partially Identifying the Value of a Policy

To understand how the lack of overlap affects our ability to find a new policy, we will separate the value of a policy into identifiable and unidentifiable components. We will then consider scenarios where it is possible to at least *partially identify* the latter term. To do so, we write the value $V(\pi, m^*)$ in terms of the observed outcome Y when our policy π agrees with the baseline policy $\tilde{\pi}$, and the unidentifiable full model $m^*(a, \boldsymbol{x})$ when π disagrees with $\tilde{\pi}$:

$$\begin{aligned} V(\pi, m^*) = \mathbb{E}\Bigg[\sum_{a\in\mathcal{A}} \pi(\boldsymbol{X}, a)\Big\{u(a)\big[\tilde{\pi}(\boldsymbol{X}, a)Y \\ + \{1 - \tilde{\pi}(\boldsymbol{X}, a)\}\, m^*(a, \boldsymbol{X})\big] + c(a)\Big\}\Bigg]. \end{aligned} \tag{2}$$

Without further assumptions, we cannot point-identify the value of the conditional expectation when a is different from the baseline policy and so we cannot identify $V(\pi, m^*)$ for an arbitrary policy π. If we place restrictions on $m^*(a, \boldsymbol{x})$, however, we can partially identify a range of potential values for a given policy π (Manski 2005). Specifically, we encode the conditional expectation as a function $m : \mathcal{A} \times \mathcal{X} \to [0, 1]$, and restrict it to be in a particular model class $\mathcal{F}$. We then combine this with the fact that we have identified some function values, that is, the conditional expectation of the observed outcome under the baseline policy $\tilde{m}(\boldsymbol{x}) \equiv m^*(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x}) = \mathbb{E}[Y \mid \boldsymbol{X} = \boldsymbol{x}]$, to form a restricted model class:

$$\mathcal{M} = \{f \in \mathcal{F} \mid f(a, \boldsymbol{x}) = \tilde{m}(\boldsymbol{x})\ \forall \boldsymbol{x} \in \mathcal{X}, a = \tilde{\pi}(\boldsymbol{x})\}. \tag{3}$$

This restricted model class combines the structural information from the underlying class $\mathcal{F}$ (i.e., $f \in \mathcal{F}$) with the observable implications from the data (i.e., $f(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x}) = \tilde{m}(\boldsymbol{x})$). With this setup, a policy π can be associated with a range of possible values $\{V(\pi, m) \mid m \in \mathcal{M}\}$, one for each observationally indistinguishable model. We discuss particular choices of the model class $\mathcal{F}$ in our study (see Section 5), deferring computation to construct the associated restricted model class $\mathcal{M}$ to Appendix C.

4.2. Criteria for Decision-Making Under Ambiguity

The lack of identifiability leads to an ambiguity in choosing an "optimal" policy: a policy could have a high value under one model and a low value under another, and no amount of data can help to adjudicate between the two scenarios. However, the value of the baseline policy $\tilde{\pi}$ is point-identified using the observed policy values and outcomes:

$$V(\tilde{\pi}) = \mathbb{E}\left[\sum_{a\in\mathcal{A}} \tilde{\pi}(\boldsymbol{X}, a)\{u(a)Y + c(a)\}\right].$$

The baseline policy $\tilde{\pi}$ is also already implemented, so a natural requirement of a new policy is that it performs *at least as well* as the baseline.

To construct such a policy, we take a maximin approach by finding a policy that maximizes the improvement over the baseline in the worst case:

$$\pi^{\text{inf}} \in \underset{\pi\in\Pi}{\text{argmax}} \min_{m\in\mathcal{M}} \{V(\pi, m) - V(\tilde{\pi})\}. \tag{4}$$

Because the value of the baseline is point-identified, this is equivalent to finding a policy that maximizes the worst-case value across the set of potential models $\mathcal{M}$, that is, $\operatorname{argmax}_{\pi\in\Pi} \min_{m\in\mathcal{M}} V(\pi, m)$.

Such maximin criteria have been widely used for policy learning in various contexts with partial identification (e.g., Kallus and Zhou 2021; Pu and Zhang 2021). Other applications include decision problems with ambiguity more broadly, such as robust statistical learning and robust optimization (e.g., Bertsimas, Brown, and Caramanis 2011; Duchi and Namkoong 2021). In addition, Gilboa and Schmeidler (1989) show that the maximin expected utility criterion is equivalent to having a preference relation among policies that satisfies a notion of *uncertainty aversion* (in addition to other more standard properties).

A benefit of choosing the maximin criterion is that so long as the policy class Π includes the baseline policy $\tilde{\pi}$, and the underlying model lies in the restricted model class $\mathcal{M}$, the maximin optimal policy $\pi^{\inf}$ will be at least as good as the baseline. We formalize this as the following proposition.

Proposition 1 (Population safety). Let $\pi^{\inf}$ be a solution to (4). If $m^* \in \mathcal{M}$, and $\tilde{\pi} \in \Pi$, then $V(\tilde{\pi}, m^*) \leq V(\pi^{\inf}, m^*)$.

We call this a "safety" property because the baseline policy acts as a fallback option. If deviating from the baseline policy can lead to a worse expected utility, a maximin policy will stick to the baseline. In this way, the new policy will change the baseline only when there is sufficient evidence for improvement. We stress that this safety property only holds if the structural assumptions about the true model m^* are correct, that is, $m^* \in \mathcal{M}$. Furthermore, this notion of safety is from the point of view of the policy maker that sets the utility function: it says nothing about the expected utility for other stakeholders with different utility functions.

Furthermore, this safety property comes at a cost: maximin policies can be conservative and sub-optimal relative to the (infeasible) oracle policy, $\pi^* \in \operatorname{argmax}_{\pi\in\Pi} V(\pi, m^*)$ (e.g., Manski 2005; Cui 2021). Because the maximin criterion limits the downside risks of deviating from the baseline policy, it can miss situations where such deviations could lead to large utility gains. We bound this sub-optimality at the population level in Appendix Theorem A.1 and for policies learned empirically from finite samples in Theorem 2.

An alternative criterion that addresses this is the *minimax regret* criterion that measures the maximum value difference between the (infeasible) oracle and the chosen policy (e.g., Manski 2007; Stoye 2012; Song 2014). In addition, maximin policies can be sensitive to the existence of edge cases. Searching for the worst case across *all* possible models ignores the fact that we may find some models unlikely, even if they are possible. A Bayesian criterion that explicitly places a prior over models and computes the posterior expected utility given the observed data would counteract this (Jia, Ben-Michael, and Imai 2023).

4.3. The Empirical Safe Policy

Next, we show how to learn a policy from the observed data $\{\boldsymbol{X}_i, \tilde{\pi}(\boldsymbol{X}_i), Y_i(\tilde{\pi}(\boldsymbol{X}_i))\}_{i=1}^n$. We begin with a sample analog to the value function in (2):

$$\hat{V}(\pi, m) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} \pi(\boldsymbol{X}_i, a)\{u(a)[\tilde{\pi}(\boldsymbol{X}_i, a)Y_i + \{1-\tilde{\pi}(X_i, a)\} \times m(a, \boldsymbol{X}_i)] + c(a)\}. \tag{5}$$

With this, we could find the worst-case sample value across all models in the restricted model class $\mathcal{M}$ from (3). Unfortunately, since we do not have the *true* conditional expectation $\tilde{m}(x)$, we cannot compute the true restricted model class. One potential approach is to obtain an estimator of the conditional expectation function, $\hat{\tilde{m}}(\boldsymbol{x})$, and use the estimates in place of the true values. However, this fails to take into account the estimation uncertainty, and could lead to a policy that improperly deviates from the baseline due to noise, especially when the convergence rate of the estimated model $\hat{\tilde{m}}(\boldsymbol{x})$ is slow.

Instead, we construct a *larger*, empirical model class $\widehat{\mathcal{M}}_n(\alpha)$, based on the observed data, that contains the true restricted model class with a probability at least $1-\alpha$, that is, $P\left(\mathcal{M} \subseteq \widehat{\mathcal{M}}_n(\alpha)\right) \geq 1-\alpha$. Then, we construct our empirical policies by first finding the worst-case in-sample value improvement, then maximizing this objective across policies π:

$$\hat{\pi} \in \operatorname*{argmax}_{\pi\in\Pi} \min_{m\in\widehat{\mathcal{M}}_n(\alpha)} \left\{\hat{V}(\pi, m) - \hat{V}(\tilde{\pi})\right\}. \tag{6}$$

We refer to $\hat{\pi}$ as the empirical safe policy, as it is the empirical analog to the $\pi^{\inf}$. Note that since the empirical restricted model class is larger than the true restricted model class, a policy derived from it is more likely to fall back to the status quo rule.

To construct the empirical model class $\widehat{\mathcal{M}}_n(\alpha)$, we use a uniform $1-\alpha$ confidence band for the conditional expectation function $\tilde{m}(\boldsymbol{x})$, with lower and upper bounds $\widehat{C}_\alpha(\boldsymbol{x}) = [\widehat{C}_{\alpha\ell}(\boldsymbol{x}), \widehat{C}_{\alpha u}(\boldsymbol{x})]$ such that $P\left(\tilde{m}(\boldsymbol{x}) \in \widehat{C}_\alpha(\boldsymbol{x}) \ \forall\, \boldsymbol{x}\right) \geq 1-\alpha$. With such a confidence band, we construct the empirical restricted model class as

$$\widehat{\mathcal{M}}_n(\alpha) = \{f \in \mathcal{F} \mid f(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x}) \in \widehat{C}_\alpha(\boldsymbol{x}) \ \forall \boldsymbol{x} \in \mathcal{X}\}.$$

Throughout, we construct our confidence bands so that the 0% confidence band corresponds to the point estimate: $\widehat{C}_{\alpha\ell}(\boldsymbol{x}) = \widehat{C}_{\alpha u}(\boldsymbol{x}) = \hat{\tilde{m}}(\boldsymbol{x})$, and therefore setting $\alpha = 1$ creates the restricted model class directly from the point estimates as described above. In our analysis in Section 5, the covariates are all discrete. Thus, we first construct a point-wise confidence interval for each unique data point, and then create a uniform confidence band by using a Bonferonni correction for the number of unique data points. We discuss how to construct the empirical model class and solve this optimization problem in Section 5.

4.4. Finite Sample Statistical Properties

Compared to the population maximin problem, the empirical problem has an additional layer of uncertainty due to sampling error that arises in finite samples. First, we establish a statistical safety property: if the structural assumptions about the true model m^* are correct, the learned policy will perform at least as well as the baseline policy with probability approximately

$1-\alpha$. We then characterize how conservative the solution is via the *optimality gap*, $V(\pi^*) - V(\hat{\pi})$: the policy value difference between the infeasible oracle that knows the true model, and our data-driven maximin policy that uses the worst-case model.

The results below use the *population Rademacher complexity* of a function class $\mathcal{G}$:

$$\mathcal{R}_n(\mathcal{G}) \equiv \mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i g(\boldsymbol{X}_i)\right|\right]$$

where ε_i is an iid Rademacher random variable, that is, $\Pr(\varepsilon_i = 1) = \Pr(\varepsilon_i = -1) = 1/2$, and the expectation is taken over both ε_i and $\boldsymbol{X}_i$ (Wainwright 2019, sec. 4). We consider the maximum Rademacher complexity across the sub-policy classes for actions $a \in \mathcal{A}$: $\Pi_a \equiv \{\pi(\cdot, a) \mid \pi \in \Pi\}$. This measures the ability of the policy class to overfit. Using this measure, we establish a statistical safety property.

Theorem 1 (Statistical safety). If the baseline policy $\tilde{\pi} \in \Pi$ and the true conditional expectation $m^*(a, \boldsymbol{x}) \in \mathcal{M}$, for any $0 < \delta \leq e^{-1}$, the value of $\hat{\pi}$ relative to the baseline $\tilde{\pi}$ is,

$$V(\tilde{\pi}) - V(\hat{\pi}) \leq 6C(K-1)\left[\max_a \mathcal{R}_n(\Pi_a) + 2\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}}\right],$$

with probability at least $1-\alpha-\delta$, where

$$C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|.$$

Like Proposition 1, Theorem 1 is only meaningful if the assumptions about the true model m^* are correct. If they are, Theorem 1 shows that the empirical safe policy will not have a lower policy value than the baseline, up to standard empirical process terms: the Rademacher complexity of the policy class Π, and an error term due to sampling variability that decreases at a rate of $n^{-1/2}$. The complexity of the policy class Π_a controls the chance that the learned policy is worse than the baseline due to overfitting.

For many standard policy classes, we expect the Rademacher complexity to decrease to zero as the sample size increases, with the complexity determining the rate of convergence. For simple policy classes, the bound will quickly go toward zero for any level α; complex policy classes will require larger samples to ensure that the safety property is meaningful, regardless of the level α. By using the larger model class $\widehat{\mathcal{M}}_n(\alpha)$, the estimation error for the conditional expectation $\hat{\tilde{m}}(\boldsymbol{x}) - \tilde{m}(\boldsymbol{x})$ does not directly enter into the bound.[2] However, if we cannot estimate $\tilde{m}(\boldsymbol{x})$ well, the empirical restricted model class $\widehat{\mathcal{M}}_n(\alpha)$ will be large, and so the empirical safe policy may collapse to the baseline policy.

To quantify the optimality gap, we denote $\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}(\pi^*(1-\tilde{\pi}))$ as the width of the empirical model class $\widehat{\mathcal{M}}_n(\alpha)$ in the direction that π^* and $\tilde{\pi}$ disagree, where

$$\widehat{\mathcal{W}}_{\mathcal{F}}(g) = \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} f(a,\boldsymbol{X}_i)g(a,\boldsymbol{X}_i) - \inf_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} f(a,\boldsymbol{X}_i)g(a,\boldsymbol{X}_i)$$

is the usual notion of the width of a set, for the set defined by all possible values of a function $f \in \mathcal{F}$ at the data points $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ for actions $a \in \mathcal{A}$, in the direction defined by the vector of all values of another function $g(a, \boldsymbol{X}_i)$.

Theorem 2 (Optimality gap). Let $u(a) = u > 0$ for all actions. If the true conditional expectation $m^* \in \mathcal{M}$, then for any $0 < \delta \leq e^{-1}$ the optimality gap is

$$V(\pi^*) - V(\hat{\pi}) \leq 2C\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}\left(\pi^*(1-\tilde{\pi})\right) + 6C(K-1) \times \left[\max_a \mathcal{R}_n(\Pi_a) + 2\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}}\right],$$

with probability at least $1-\alpha-\delta$, where

$$C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|.$$

To simplify the statement, we have assumed that the utility gain across different actions is constant and, without loss of generality, positive.

The bound on the empirical optimality gap contains the width term $\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}(\pi^*(1-\tilde{\pi}))$, in addition to the standard empirical process terms found in Theorem 1. If the baseline policy is the oracle policy, then this width is zero, the bounds in Theorems 1 and 2 coincide, and the regret of $\hat{\pi}$ relative to the oracle π^* will converge to zero so long as the complexity of the policy class goes to zero. Otherwise, the width term does not necessarily converge to zero: if the baseline and oracle policies disagree for many cases, the empirical safe policy could perform substantially worse than the oracle.

This leads to a tradeoff between statistical safety (Theorem 1) and optimality (Theorem 2). Increasing the confidence level will yield a greater probability that the learned policy is safe relative to the baseline, but it will also widen the potential optimality gap when the baseline and oracle policies disagree. This is similar to the tradeoff between a low Type I error rate (α low) and high power ($\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}(\pi^*(1-\tilde{\pi}))$ low) in hypothesis testing. The tradeoff extends to the choice of model class as well: statistical safety requires that the model class contains the true conditional expectation function, that is, $m^* \in \mathcal{M}$. This is palatable if we choose a complex model class, but complex model classes may lead to a greater amount of uncertainty due to severe lack of identification and/or greater estimation error.

This tradeoff does not exist if the baseline policy is stochastic and there is overlap between actions. In this case, the conditional expectation function is non-parametrically identifiable. While we can still account for statistical uncertainty by constructing the empirical model class $\widehat{\mathcal{M}}_n(\alpha)$, we stress that our approach is not appropriate when the baseline policy is stochastic. It only uses a model for the outcomes and so will rely on stronger assumptions about the outcome model and be inefficient relative to a doubly

[2] In Appendix A.2, we extend these results to consider the case where $1-\alpha = 0$ and we use point estimates rather than confidence bounds. We show that the bounds have additional terms due to estimation error of the model.

robust approach that incorporates the action probabilities as proposed by Athey and Wager (2021).

In practice, we do not know the oracle policy. To operationalize the bound in Theorem 2, we can further upper bound the optimality gap by finding the policy that leads to the *worst-case* width, were it the oracle policy: $\widehat{\mathcal{S}}(\mathcal{F}, \Pi; \tilde{\pi}) \equiv \sup_{\pi \in \Pi} \widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}(\pi(1-\tilde{\pi}))$. We refer to this quantity as the "size" of the empirical restricted model class because it measures the degree of uncertainty about the true model m^* in regions of the covariate space where a policy $\pi \in \Pi$ could deviate from the baseline.

We use this as a diagnostic measure in Section 5. Note that the policy class Π affects the size. Restricting to policies that can only disagree with the baseline in only a few cases will lead to a small size. Conversely, if we attempt to optimize over an expansive policy class, the size diagnostic can be large. However, the size term is a loose upper bound: even if the size is large, the optimality gap may still be small if it happens that the oracle policy π^* is similar to the baseline $\tilde{\pi}$. Therefore, a large size term is a warning that there may be insufficient information to learn an improved policy, but it does not rule it out entirely.

4.5. Learning from Experiments Evaluating a Deterministic Policy

In our empirical study, the existing PSA-DMF system was compared to not providing algorithmic recommendations. While a primary goal of this randomized controlled trial was to evaluate whether one should adopt the algorithmic policy, we can leverage the control group data to weaken the restrictions of the underlying model class $\mathcal{M}$ by placing assumptions on *treatment effects* rather than the expected potential outcomes.

We consider an expanded set of actions that includes all actions in $\mathcal{A}$ and a "null" action (i.e., do not provide an algorithmic recommendation). We denote the null action as $a = -1$, with potential outcome $Y(-1)$. Let $Z_i \in \{0, 1\}$ be a treatment assignment indicator where $Z_i = 0$ if no policy is enacted (i.e., the null policy), and $Z_i = 1$ if the policy follows the baseline policy $\tilde{\pi}$. Let $e(x) = P(Z = 1 \mid \boldsymbol{X} = \boldsymbol{x})$ be the probability of assigning the treatment condition for an individual with covariates $\boldsymbol{x}$. This is the propensity score *for the treatment assignment* and since this is an experiment, it is known and strictly between 0 and 1. While we consider general propensity scores when describing the method, in our experiment $e(x) = 0.5$ for all cases. This allows us to identify the conditional expectation function, $m^*(-1, \boldsymbol{x}) = \mathbb{E}[Y \mid \boldsymbol{X} = \boldsymbol{x}, Z = 0]$. Defining the true conditional average treatment effect (CATE) of the action a relative to the null action -1 as $\tau^*(a, \boldsymbol{x}) \equiv m^*(a, \boldsymbol{x}) - m^*(-1, \boldsymbol{x})$, we can also identify the CATE under the baseline policy $\tilde{\pi}(\boldsymbol{x})$, $\tilde{\tau}(\boldsymbol{x}) = \tau^*(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x})$.

We now write the value function in terms of the (partially-identified) CATE and the (point-identified) expected outcome under the null action. With a constant utility gain $u(a) = u$, we can write it as:[3]

$$V(\pi) = \mathbb{E}\left[\sum_{a \in \mathcal{A}} \pi(\boldsymbol{X}, a)\{u \cdot \tau^*(a, \boldsymbol{X}) + c(a)\}\right] + u \cdot \mathbb{E}[m^*(-1, \boldsymbol{X})].$$

[3] Proposition A.1 in the Appendix shows this result for the general utility case.

Because the baseline term $\mathbb{E}[m^*(-1, \boldsymbol{X})]$ does not depend on π and is point-identified, we can re-parameterize the model class to impose restrictions on the *treatment effects* $\mathcal{T} = \{f(a, \boldsymbol{x}) \equiv m^*(-1, \boldsymbol{x}) + h(a, \boldsymbol{x}) \mid h \in \mathcal{F}, h(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x}) = \tau(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x})\}$. The CATE function is sometimes assumed to be simpler (e.g., smoother, sparser, fewer interaction terms) than the conditional expected potential outcome (see, e.g., Künzel et al. 2019, who argue that the CATE should be easier to estimate). Therefore, we may consider a smaller model class for the treatment effects than for the baseline outcomes, leading to a smaller optimality gap in Theorem 2. We can also construct the empirical analog by creating a larger empirical model class $\widehat{\mathcal{T}}_n(\alpha)$ as in Section 4.3.

Finally, following Kitagawa and Tetenov (2018), to account for potential unequal assignment into treatment, we can solve the population and empirical robust optimization problems using the inverse probability weighted outcome $\Gamma(Z, \boldsymbol{X}, Y) \equiv Y\{Z(1 - 2e(\boldsymbol{X})) + e(\boldsymbol{X})\}/\{e(\boldsymbol{X})(1 - e(\boldsymbol{X}))\}$, which equals the conditional expected potential outcome in expectation, that is, $\mathbb{E}[\Gamma(Z, \boldsymbol{X}, Y) \mid Z = z, \boldsymbol{X} = \boldsymbol{x}] = z \cdot m^*(\tilde{\pi}(\boldsymbol{x}), \boldsymbol{x}) + (1 - z) \cdot m^*(-1, \boldsymbol{x})$.

5. Empirical Analysis of the Pre-Trial Risk Assessment

5.1. Implementation Details

For our empirical analysis, we will represent the empirical restricted model classes as the set of functions that are upper and lower bounded point-wise by two bounding functions, $\widehat{\mathcal{T}}_n(\alpha) = \{f : \mathcal{A} \times \boldsymbol{X} \to \mathbb{R} \mid \widehat{B}_{\alpha\ell}(a, \boldsymbol{x}) \le f(a, \boldsymbol{x}) \le \widehat{B}_{\alpha u}(a, \boldsymbol{x})\}$, where the upper and lower bounds are chosen to satisfy $P\left(\mathcal{T} \subseteq \widehat{\mathcal{T}}_n(\alpha)\right) \ge 1 - \alpha$. In Appendix C, we show how to compute these bounds using simultaneous confidence intervals when the underlying model class is the set of Lipschitz functions or linear models.

The point-wise bound allows us to solve for the worst-case empirical value $\hat{V}^{\inf}(\pi)$ by finding the minimal value for each action-covariate pair (see Pu and Zhang 2021). Finding the empirical safe policy by solving (6) is equivalent to solving an empirical welfare maximization problem using a quasi-outcome that imputes the counterfactual outcome with the lower bound when the action disagrees with the baseline policy:

$$\max_{\pi \in \Pi} \frac{1}{n} \sum_{i=1}^{n} \sum_{a \in \mathcal{A}} \pi(\boldsymbol{X}_i, a)\big\{u(a)\big[\tilde{\pi}(\boldsymbol{X}_i, a)\{\Gamma(1, \boldsymbol{X}_i, Y_i) - \Gamma(0, \boldsymbol{X}_i, Y_i)\} + \{1 - \tilde{\pi}(\boldsymbol{X}_i, a)\}\widehat{B}_{\alpha\ell}(a, \boldsymbol{X}_i)\big] + c(a)\big\}, \tag{7}$$

where we have omitted terms that do not depend on π. A similar implementation strategy is applicable to cases where we model potential outcomes rather than treatment effects.

5.2. Learning a New NVCA Flag Threshold

We begin our analysis by considering a small change to the existing system: learning a new threshold for the NVCA flag. Our goal here is to find the optimal NVCA threshold in the worst case, where our preferred outcome is no NVCA.

Choosing the policy class. We first formalize our choice of policy class. Let $x_{\text{nvca}} \in \{0, \ldots, 6\}$ be the total number of NVCA points for an arrestee, computed using the point system in Appendix Table G.1. Recall that the baseline NVCA algorithm

is to trigger the flag if the number of points is greater than or equal to 4, that is, $\tilde{\pi}(x_{\text{nvca}}) = \mathbb{1}\{x_{\text{nvca}} \geq 4\}$. Our policy learning problem is to choose a policy among the class of threshold policies, $\Pi_{\text{thresh}} = \{\pi(x) = \mathbb{1}\{x_{\text{nvca}} \geq \eta\} \mid \eta \in \{0, \ldots, 7\}\}$. We will keep the baseline weighting on arrestee risk factors and *only* change the threshold η. Since this policy class only has eight elements, we can compute the empirical maximin policy $\hat{\pi}$ by solving (7) via an exhaustive search.

Choosing the model class. We next choose a model class for the CATE on no NVCA occurring, $\tau^*(a, x_{\text{nvca}})$. There are many potential ways to characterize the complexity of functions of one variable such as $\tau^*(a, \cdot)$. Here, we characterize it via a Lipschitz constraint that $\left|\tau(a, x_{\text{nvca}}) - \tau(a, x'_{\text{nvca}})\right| \leq \lambda_a |x_{\text{nvca}} - x'_{\text{nvca}}|$ for any pair of NVCA points $x_{\text{nvca}}, x'_{\text{nvca}}$.

To construct the empirical restricted model class, we set the level to $1-\alpha = 0.8$, allowing some tolerance for statistical uncertainty and construct a simultaneous 80% confidence interval for the CATE via a Bonferroni correction for the seven unique values (see Appendix C.2 for details on computing the bounds). We also restrict the treatment effects to be bounded between -1 and 1, since the outcome is binary.[4]

For this model class, we need to specify the Lipschitz constants for the CATE when the flag is and is not triggered (λ_1 for $\tau(1, x_{\text{nvca}})$) and λ_0 for $\tau(0, x_{\text{nvca}})$, respectively). We adapt a suggestion from Imbens and Wager (2019) for model classes with a bounded second derivative to the Lipschitz case. We estimate the CATE function by taking the difference in NVCA rates with and without provision of the PSA at each level of x_{nvca}. Then, we measure the largest consecutive difference between CATE estimates (0.016 and 0.433 for $a = 0, 1$, respectively). Finally, we set the Lipschitz constants to be a constant multiple C of this difference yielding $\lambda_0 = C \times 0.016$ and $\lambda_1 = C \times 0.433$. Setting $C = 1$ gives the smallest Lipschitz constants supported by the data; increasing C will be more conservative.

Choosing the utility function. Recall that in our parameterization we must define the difference in utilities when there is and is not an NVCA, $u(a) = u(1, a) - u(0, a)$, for both actions $a \in \{0, 1\}$. This captures the benefits of avoiding an NVCA. Countering this benefit is the baseline cost of action a, $c(a) = u(0, a)$. The marginal monetary cost of triggering the NVCA flag is zero given the initial fixed cost of collecting the data for the PSA. However, to the extent that triggering the NVCA flag increases the likelihood of pre-trial detention, it will lead to an increase in fiscal costs—for example, housing, security, and transportation—for the jurisdiction. Furthermore, there are potential socioeconomic costs to the defendant and their community that balance against potential benefits from avoiding more criminal activity.

To represent these costs, we will place zero cost on not triggering the NVCA flag, $c(0) = 0$, and a cost of 1 on triggering the flag, $c(1) = -1$. We then assign an equal utility gain from avoiding an NVCA, $u(1) = u(0) = u$ (equivalently, the cost of an NVCA is $-u$). This yields a utility function of the form $u(y, a) = u \times y - a$, where u is the ratio of the cost of an NVCA to the cost of triggering the flag. Choosing a particular value of u is outside the scope of this article and indeed would be inappropriate for us to do: the choice depends on societal preferences and may need to be arrived at in a collaborative process between policy-makers in the criminal justice system and the communities impacted by it. Instead, we examine how adjusting the ratio u affects the policies we learn.[5]

Learning a maximin NVCA threshold. Figure 2(a) presents the empirical restricted model class with a particular multiplicative constant of $C = 3$ by showing point estimates and simultaneous 80% confidence intervals for the observable component of the CATE function $\tau^*(\tilde{\pi}(x_{\text{nvca}}), x_{\text{nvca}})$ and the partial identification set for the unobservable component. There is substantially more information when extrapolating the CATE for the case that the NVCA flag is triggered. This is because the point estimates do not vary much with the NVCA points, leading to a small Lipschitz constant. On the other hand, when extrapolating in the other direction, there is a large jump in the point estimates between $x_{\text{nvca}} = 5$ and $x_{\text{nvca}} = 6$, leading to a large Lipschitz constant. This means that the empirical restricted model class puts essentially no restrictions on $\tau^*(1, x_{\text{nvca}})$ for $x_{\text{nvca}} < 4$.

Estimating this policy requires choosing the Lipschitz multiplicative factor $C \geq 1$. We fit the empirical safe policy across a range of values to see if the results are stable. Figure 2(b) shows the learned thresholds as we vary both the relative cost u of an NVCA in the utility function and the multiplicative factor C. When the cost of an NVCA is $u \leq 7$, the data support increasing the threshold to at least 6 even in the worst case and even with $C = 10$, only triggering the flag for arrestees with the observed maximum of 6 total NVCA points. The results for larger costs are more sensitive to the choice of C, and the learned threshold collapses back to the baseline of 4 for intermediate choices of C.

Raising the threshold to $\eta = 6$ is a much more lenient policy than the status quo, reducing the number of arrestees flagged as at risk of an NVCA by 95%. We find evidence for such a large change because there is no meaningful effect of providing the PSA on the absence of an NVCA, except for those arrestees who have the maximum of 6 points (Figure 2(a)). One possible reason for these small effects is that the judge's behavior is not affected. This appears to be the case when $x_{\text{nvca}} \leq 4$: there is little effect on the judge's bail decision in these cases. However, for $x_{\text{nvca}} > 4$, providing the PSA increases cash bail decisions by over 30 pp (see Appendix G.1 for further discussion). This suggests that PSA provision is leading to additional bail decisions without a requisite decrease in NVCAs for $x_{\text{nvca}} = 5$.

Thus, even in the worst case, the threshold could be raised to $\eta = 6$ without an increase in the NVCA rate that outweighs costs from triggering the flag. As we increase the cost of an NVCA, however, at some point (e.g., $u \geq 13$ for $C = 3$), the cost becomes large enough, making the empirical safe policy revert to the status quo with the threshold at $\eta = 4$.

[4] This is not the tightest possible bound, since the restriction is that $0 \leq m(-1, x) + \tau(a, x) \leq 1$. To incorporate the uncertainty in estimating $m(-1, x)$ in finite samples we could use analogous techniques to those in Section 4.3; we leave this to future work.

[5] Note that mathematically one could use a negative cost of triggering the flag, but this would encourage triggering the flag even if it would not avoid an NVCA.

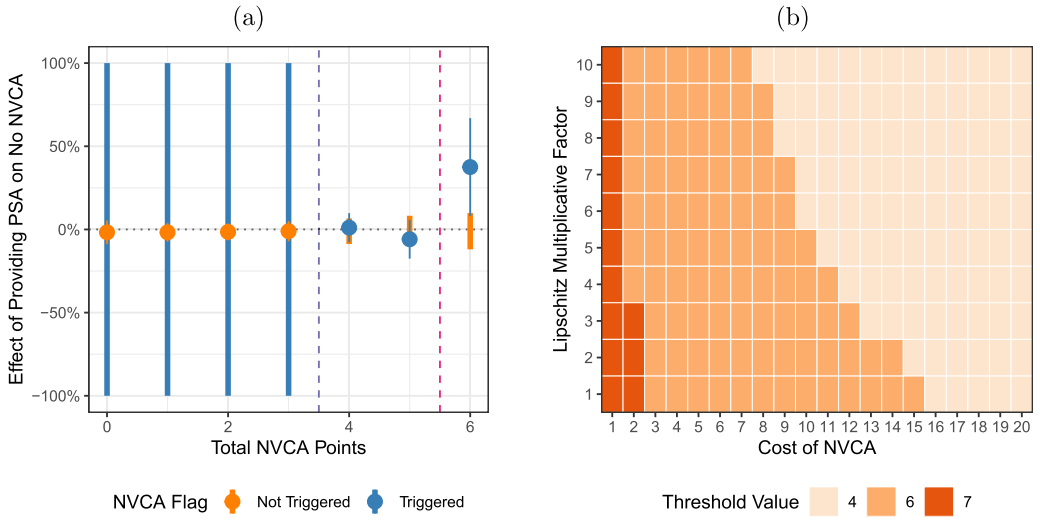

Figure 2. Learning a new NVCA flag threshold. (a) Empirical restricted model class and maximin threshold with a Lipschitz multiplicative factor of $C = 3$. The points and thin lines around them are point estimates and a simultaneous 80% confidence interval for the partial CATE function $\tau(\tilde{\pi}(x_{\text{nvca}}), x_{\text{nvca}})$ when the NVCA flag is not triggered ($\tilde{\pi}(x_{\text{nvca}}) = 0$, in orange) and is triggered ($\tilde{\pi}(x_{\text{nvca}}) = 1$, in blue). The thick solid lines represent the partial identification set for the unobservable components of the CATE, $\tau(1, x_{\text{nvca}})$ for $x_{\text{nvca}} < 4$ and $\tau(0, x_{\text{nvca}})$ for $x_{\text{nvca}} \geq 4$. The purple dashed line represents the baseline policy of triggering the flag when $x_{\text{nvca}} \geq 4$, and the pink dashed line is the empirical safe policy that only triggers the flag when $x_{\text{nvca}} \geq 6$. (b) Maximin threshold values solving (7) for the NVCA flag threshold rule with a level of $1 - \alpha = 80\%$ as the cost of an NVCA increases from 1 to 20 times of the cost of triggering the NVCA flag, and the multiplicative factor on the estimated Lipschitz constant varies from 1 to 10.

5.3. Learning New FTA, NCA, and NVCA Risk Scoring Rules

We next turn to constructing new, maximin optimal FTA, NCA, and NVCA risk scores. For each risk score, we focus on the *absence* of the corresponding negative outcome.

Choosing the policy classes. A key consideration is the form of the policy classes used for each risk score. One possibility is to allow the policies to be flexible functions of all the information available in the system. Although the oracle policy may have a high expected utility in this case, in finite samples a complex policy can over-fit and reduce the quality of the safety property in Theorem 1. In addition, the oracle policy may be substantially different from the baseline policy, leading to a large optimality gap in Theorem 2. Lastly, in real-world applications, policy makers might be reluctant to adapt the existing system to an entirely new policy. For these reasons, we use the same set of risk factors and focus on changing the weight applied to each risk factor (see Appendix Table G.1).

For each risk score, we formally describe the status quo rule as consisting of a vector of integer-valued weights $\tilde{\boldsymbol{\theta}}$ on the risk factors $\boldsymbol{x}$, and a mapping from the linear combination of the risk factors $\tilde{\boldsymbol{\theta}}^\top \boldsymbol{x}$ to the K risk levels via thresholds. We consider the policy class that consists of all possible vectors of integer-valued weights and all possible thresholds, restricting to inter-valued weights in order to mimic the structure of the existing risk scores. For example, recall that the NVCA flag has $K = 2$ risk levels (a flag for elevated NVCA risk), 7 binary risk factors, and the baseline policy is $\tilde{\pi}(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j=1}^{7} \tilde{\theta}_j x_j \geq 4\right\}$. We then write the corresponding NVCA flag policy class as

$$\Pi_{\text{nvca}} = \left\{ \pi(\boldsymbol{x}) = \mathbb{1}\left\{ \sum_{j=1}^{7} \theta_j x_j \geq \eta \right\} \;\middle|\; \theta_j \in \mathbb{Z}, \eta \geq 0 \right\}. \quad (8)$$

This policy class includes the original NVCA flag rule as a special case. We can construct the policy classes for the FTA and NCA rules similarly by including multiple thresholds (see Appendix G.2 for a formal definition). To simplify comparisons to the status quo and avoid identifiability issues, we will primarily constrain the thresholds η to be equal to the status quo values. This allows us to understand any differences from the status quo rule by comparing the learned weight vector to the baseline weight vector $\tilde{\boldsymbol{\theta}}$. With this policy class, the optimization problem is a mixed integer program; we solve this with the Gurobi solver.

Choosing the model class. In contrast to changing only the NVCA threshold above, here the CATE is a function of multiple binary variables. A natural way to characterize the complexity of such models is by the number and strength of interaction terms between the variables. We focus on the two simplest models: an additive effect model $\mathcal{T}_{\text{add}} \equiv \left\{\tau(a, \boldsymbol{x}) = \sum_j \tau_{aj} x_j\right\}$ and a second order effect model $\mathcal{T}_{\text{two}} \equiv \left\{\tau(a, \boldsymbol{x}) = \sum_j \sum_{k<j} \tau_{ajk} x_j x_k\right\}$. Because the covariates are discrete, we can write these using linear models. We again restrict the treatment effects to be bounded between -1 and 1. These two model classes lead to different restrictions and ultimately affect what policies we learn from the experiment (see Appendix C.3 for details). This is partly because even with infinite data the models may not be identifiable. But it is also because with finite data there is a different amount of uncertainty in each model class.

To diagnose the amount of information available in each model class, we use the size measure $\widehat{\mathcal{S}}(\widehat{\mathcal{T}}_n(\alpha), \Pi; \tilde{\pi})$. Figure 3 depicts this information by showing how the size of the model class (vertical axis), changes with the desired confidence level $1 - \alpha$ (horizontal axis) for each risk score and model class. We also show the difference in the size for the NVCA rule when fixing the threshold to the existing value versus including it as a decision variable.

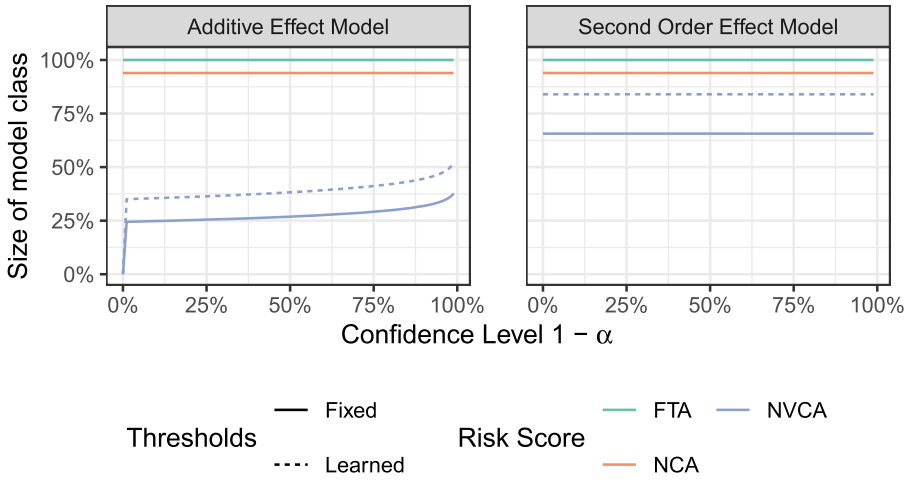

Figure 3. The size (as a percentage of its maximum value) of two different model classes with respect to the linear threshold policy class versus the confidence level $1-\alpha$ for the FTA (green), NCA (orange), and NVCA (purple) scoring rules. The dashed purple line shows the size for the NVCA model class when the threshold is included as a decision variable and learned in addition to the weights.

There is a stark difference in the amount of information between the different risk scores within the same model class. Under the additive model for the NVCA rule, the size is zero when the confidence level is zero, implying that this model is identifiable. This is due to the structure of the NVCA flag rule: for any given value of a covariate, it is possible to observe cases with the flag set to zero or one. When accounting for the statistical uncertainty, the size increases, but it is substantially smaller than the size for the FTA and NCA rules, both of which are near or at the maximum value of 2. Because these risk scores have six levels, we would need to observe cases with all six levels for any given value of a covariate in order to identify the additive model. Overall, there is little information available to support changing the FTA and NCA scoring rules.

Turning to the second order treatment effect model for the NVCA score, which makes weaker assumptions, we find that it is likely too large a class for us to learn a new NVCA rule, with roughly twice the size as for the additive effect model. This is because there are several pairs of variables that always trigger the NVCA flag (e.g., if both the current offense is violent and the arrestee has 3 or more prior violent convictions). Finally, we observe that increasing the flexibility of the policy class by including the threshold as a decision variable rather than keeping it fixed increases the size because it is a function of *both* the model class and the policy class.

These diagnostics point to focusing on the NVCA score with an additive effect model. There is likely not enough information to make any changes to the NCA and FTA scores under either model, and the second order effect model for the NVCA flag is not well enough identified. However, in Appendix G.1, we find some evidence for the existence of interactions for the NVCA score via classical model testing procedures. Therefore, we caution over-interpreting our results. For completeness, we show these results in Appendix G and indeed find that the optimal solution for the worst case is to not deviate from the status quo rules.

Choosing the utility function. We use the same utility parameterization as in Section 5.2. For this value function, the marginal decrease in the utility from triggering the flag is constant regardless of the proportion of arrestees that are classified as an NVCA risk. However, higher levels of pre-trial incarceration can have additional negative impacts on the community above and beyond the cost to the individual. In Appendix G.1, we include an additional penalty to triggering the NVCA flag that scales with the proportion of arrestees classified as being at risk.

Learning a maximin NVCA flag. Figure 4(a) presents the changes to the original rule made by the maximin policy that solves the optimization problem given in (7) under the additive treatment effect class $\mathcal{T}_{\text{add}}$. The changes are shown in terms of the proportion of arrestees flagged for an NVCA risk as we vary the cost of an NVCA $-u$ and the confidence level $1-\alpha$. Across every confidence level, the maximin policy differs less and less from the original rule as the cost of an NVCA increases, moving from never triggering the flag at a 1–1 cost to eventually collapsing back to the status quo if the cost crosses an α dependent threshold. For a given cost of an NVCA, policies at lower confidence levels are more aggressive in deviating from the original rule, prioritizing a potentially lower regret relative to the (infeasible) optimal policy at the expense of a higher chance that the new policy is worse than the original rule.[6]

Figure 4(b) shows the integer weights on the risk factors for the maximin policy at the $1-\alpha = 80\%$ level as the cost of an NVCA increases. In the limiting setting where an NVCA is given the same cost as triggering the NVCA flag, the maximin policy never triggers the flag because it cannot be worth the cost. Once the cost is at least 14 times the cost of triggering the flag, the learned policy reduces to the original rule. In light of the sizes shown in Figure 3, this behavior is primarily due to increased uncertainty in the effect of triggering the NVCA flag. If the policy maker were to set the cost of an NVCA above a certain point, any change in the policy would be too risky to act upon. For intermediate values, the learned policy places less weight on the number of prior violent convictions and whether the current

[6]Except for when the cost of an NVCA is greater than 12 and the confidence level is 0%, the maximin policies do not trigger the flag when the original rule does not.

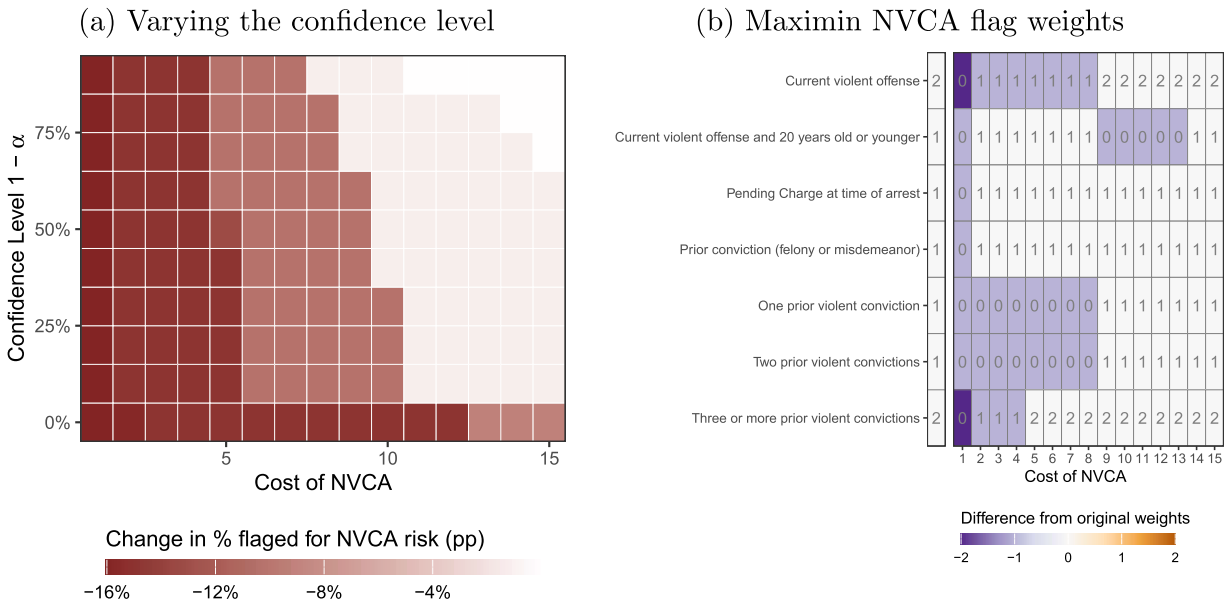

Figure 4. (a) The percentage point difference in the proportion of arrestees flagged for NVCA risk between the maximin policy and the original NVCA score as the cost of an NVCA increases from 1 to 15 times of the cost of triggering the NVCA flag and the confidence level varies between 0% and 100%. (b) Change in Maximin NVCA flag weights θ (in (8)) as the cost of an NVCA increases from 1 to 15 times the cost of triggering the NVCA flag, at a confidence level of $1 - \alpha = 80\%$.

offense is violent than the original rule. In Appendix G.1, we consider a more flexible policy class that includes additional risk factors.

5.4. Learning a New DMF Matrix for Bail Recommendation

Finally, we analyze the overall recommendation given by the DMF matrix (see Figure 1). This aggregates the FTA and NCA scores into an overall recommendation on the level of cash bail and pre-trial supervision and monitoring conditions. Below, we focus on using the absence of an NVCA as the primary outcome.

Choosing the policy class. We consider constructing a new DMF matrix based on the FTA and NCA scores, which we combine into a vector $(x_{\text{fta}}, x_{\text{nca}}) \in \{1, \ldots, 6\}^2$, restricting our analysis to the 1544 cases that used the DMF matrix rather than those for whom cash bail was automatically recommended. We will focus on a policy class that keeps the structure of the existing rule encoded by the DMF matrix. An important aspect of the rule is that it is *monotonic*; the risk level cannot decrease if either the FTA or NCA score increases. Formally, we can encode the monotonic policy class as, $\Pi_{\text{mono}} \equiv \{\pi((x_{\text{fta}}, x_{\text{nca}})) \leq \pi((x_{\text{fta}} + 1, x_{\text{nca}})) \text{ and } \pi((x_{\text{fta}}, x_{\text{nca}})) \leq \pi((x_{\text{fta}}, x_{\text{nca}} + 1))\}$. Again, this leads to an integer program, which we solve via the Gurobi solver. We will consider four variations of the policy: (i) the overall risk level from 1 to 7; (ii) the quaternary recommendation of a signature bond, modest cash bail, moderate cash bail, or (full) cash bail; (iii) the ternary recommendation that combines modest and moderate cash bail; and (iv) the binary recommendation that collapses together all cash bail recommendations.

Choosing the model class. We again focus on the class of additive treatment effect models $\tau_{\text{add}}(a, \boldsymbol{x}) = \tau_{\text{fta}}(a, x_{\text{fta}}) + \tau_{\text{nca}}(a, x_{\text{nca}})$. We only condition on the FTA and NCA scores since they are the two components of the DMF decision matrix. Because x_{fta} and x_{nca} are discrete with six values, we can further parameterize the additive terms as six-dimensional vectors. Importantly, this rules out interactions between the FTA and NCA scores in the effect. In Appendix G.3 we test for the presence of interactions and do not find evidence against the null of an additive model.[7]

Figure 5(a) presents the size of this model class relative to the monotone policy class for the four types of recommendations as we vary the confidence level for the three types of DMF recommendations. There is no information to learn reliably a new fine-grained overall risk score or quaternary bail recommendation. This is due to the structure of the DMF matrix: some risk levels are only possible for a single NCA score, and others (such as the moderate cash bail condition) only for a single combination of FTA and NCA scores.

Therefore, we focus here on the binary cash bail recommendation, where the size of the model class is large, but smaller than for the ternary bail recommendation. This is because we can never observe a case where the DMF recommends a signature bond with *either* an FTA score or NCA score above 4, nor can we observe a case where the DMF recommends cash bail with either an FTA score below 2 or an NCA score below 3. In the middle is an intermediate area with FTA scores between 2 and 4 and NCA scores between 3 and 4 where we can fully identify the effect of assigning cash bail under the additive model. For this intermediate area, there is a significant amount of uncertainty due to small sample sizes: there are only three cases where cash bail is recommended that have an NCA score of 3. Appendix Figure G.14 visualizes this uncertainty.

[7] Note that we could also use a Lipschitz restriction as in Section 5.2. This alternative assumption may be significantly weaker, though it would require choosing the Lipschitz constant.

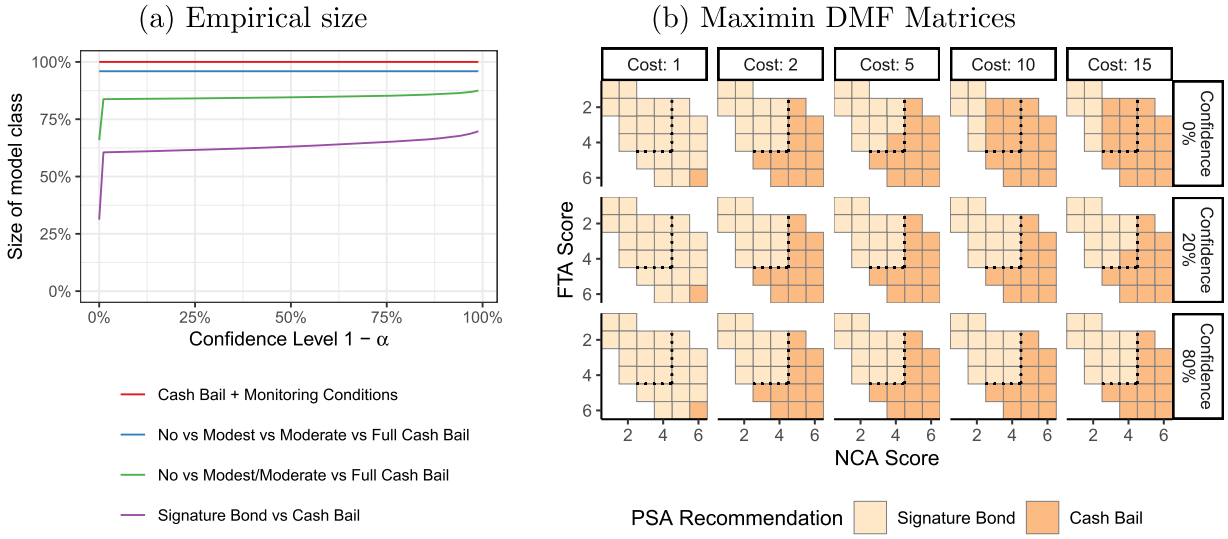

Figure 5. (a) The size (as a percentage of the maximum value) of the additive model class with respect to the monotone policy class as the confidence level varies for cash bail recommendation policies, collapsing together successively more gradations on bail. The coarsest policy—Signature Bond versus any Cash Bail—has the most information available. (b) Maximin monotone cash bail recommendations under an additive model for the treatment effects, as the cost of an NVCA and the confidence level vary. The dashed black line indicates the original decision boundary between a signature bond (above and to the left) and cash bail (below and to the right). The original decision boundary is modified only when the cost and confidence are low.

Choosing the utility function. We follow Sections 5.2 and 5.3, and parameterize the utility as a fixed cost of 1 for recommending cash bail and varying the cost of an NVCA.

Learning a maximin DMF Matrix. Figure 5(b) shows the learned policies when varying the cost of an NVCA and different confidence levels. In the limiting case where the cost of an NVCA is equal to recommending cash bail, the safe policy recommends cash bail only for the most extreme cases. In the other limiting case, where we set the confidence level to 0 and rely on the point estimates directly rather than accounting for the statistical uncertainty, increasing the cost of an NVCA leads to more of the intermediate area with FTA scores between 2 and 4 and NCA scores between 3 and 4 being assigned cash bail until the cost is high enough that the entire identified area is assigned cash bail. However, this does not hold up to even the slightest degree of statistical uncertainty due to the uncertainty in the treatment effects. Because the effects of assigning cash bail are both small and uncertain, the learned policy reduces to the existing DMF matrix.

6. Discussion

Data-driven algorithmic policies and recommendations have become an integral part of our society. An important challenge when learning a new policy is to ensure that it does not perform worse than the existing one. In settings like ours where decisions are highly consequential, policy makers should be able to limit the probability that a new algorithmic recommendation system achieves a worse outcome than the existing system. This is particularly essential when it is impossible to randomize the algorithm output for ethical and logistical reasons. The lack of identification necessitates extrapolation, making it impossible to learn a new policy using standard statistical methods.

We address these challenges by partially identifying the value of potential policies. Since this leads to a decision-making problem under ambiguity, we use the maximin criterion that selects the best policy in the worst case. Our methodology has a statistical safety property: if we make correct structural assumptions about the true model, the resulting policy will not be worse than the status quo policy with some probability, up to sampling uncertainty.

Our goal is to understand what changes to the PSA-DMF recommendation system should be made, if any. We do not find strong support to change the existing FTA and NCA scores, nor the overall risk score and bail recommendation. This is due to a confluence of factors. Foremost is the conservative nature of the maximin criterion that yields a strong bias toward the status quo. We emphasize that failing to find strong evidence to change the status quo policy does not necessarily imply that the status quo is desirable.

With the conservative criterion, our analysis is not informative about the FTA and NCA scores and the overall risk level due to the design of these algorithms. They have many fine gradations and in some cases only a single unique combination of inputs can lead to a particular output. This means that there is little to extrapolate from and the bounds are uninformative, even with strong structural assumptions. In contrast, our analysis is not informative about the binary bail recommendation due to a combination of identifiability issues and limited sample sizes. With an additive model, we can only identify impacts for cases with intermediate FTA and NCA scores, but the sample sizes in this intermediate area are too small to make strong conclusions.

However, the data do support altering the NVCA flag, even with this conservative criterion, either by raising the threshold or by putting less weight on violent convictions and offenses. Both of these would lead to a more lenient rule that flags fewer arrestees, and the data support these changes even when the cost of an NVCA is 8–13 times the cost of triggering the flag. Stevenson and Mayson (2022) present survey evidence showing

that 50% of individuals rate being a victim of an assault as bad as between 5 days and 6 months of detention; implying a cost ratio for one month of detention between $\frac{1}{6}$ and 6. Choosing any ratio within this range would lead to a change to the system, though since our focus is on triggering the flag, rather than detention, a larger benchmark may be more appropriate.

Our analysis serves as an initial proof of concept, probing various elements of the existing risk assessment system. As such, there are several limitations and various ways that the analysis has been simplified. In particular, we place costs on the algorithmic outputs (the PSA recommendations) rather than on the resulting human decisions (the judge's bail decision). In Appendix D, we directly incorporate the costs of the judges' decisions, but find that this adds too much statistical uncertainty to improve reliably upon the existing rule.

Another limitation of our analysis in Section 5.3 is that we separately consider each outcome and its risk score. Since each risk score can affect each outcome, all pairs of risk scores and outcomes could be considered. This issue is also present in our analysis of the DMF matrix (Section 5.4), where we focus on NVCAs but the bail recommendations can impact all three outcomes. A fuller analysis may consider all three risk scores and the bail recommendation simultaneously for all three outcomes, using a utility function that incorporates all of the outcomes and includes measures of costs such as economic and social outcomes. However, such an analysis may not be informative, given the limitations in the design and the data discussed above.

An important limitation of our methodology is that learning algorithmic policies requires making many nontrivial choices. For example, focusing on the simple setting of changing the NVCA flag threshold in Section 5.2 requires (a) specifying a model class; (b) specifying a significance level; and (c) choosing a utility function, among other things. The analyses in Sections 5.3 and 5.4 include even more involved analytical and implementation decisions. Therefore, it is important to examine the sensitivity of empirical results to these choices.

Choosing the model class can be difficult. With randomized evaluations of the status quo policy, simple treatment effect structures may be plausible because treatment effects are often far less heterogeneous than baseline outcomes. We inspected the sensitivity and stability of the maximin policy to modeling choices and hyper-parameters, such as the choice of Lipschitz constant. Formalizing these heuristics is an important direction for future work.

Another key choice is the policy class to optimize over. We recommend choosing a policy class that can lead to limited adjustments to the baseline policy rather than wholesale changes. While more flexible policy classes could yield better results, we are unlikely to achieve them, and large changes to existing systems may not be practically feasible.

Finally, our methodological approach has a wide range of potential applications. For transparency and interpretability, many data-driven algorithms in public policy and medicine are based on known, deterministic rules rather than randomized rules. Examples include the SNAP eligibility rule, the MELD score for liver transplantation, and other risk assessment instruments used across public policy contexts (see Coston et al. 2020, and references therein). These instances will all have identifiability issues due to lack of overlap, and our methodology addresses this challenge by learning a new, safe policy that improves upon the status quo.

If the algorithm is designed in such a way that there is little to extrapolate from—as was the case for the FTA and NCA scores—our approach is unlikely to be informative. Our methodology may be more effective when the baseline policy includes multiple inputs, each with a large region where multiple actions are possible. This can be true when there are group-specific thresholds for a common risk score or decision variable, for example as with school enrollment and loan access, and income limits for social programs (Zhang, Ben-Michael, and Imai 2023). However, different studies may require other implementation details. For instance, our study only includes discrete covariates; incorporating continuous covariates will require additional implementation work. In addition, analyzing continuous outcomes with nonlinear utility functions, incorporating other criteria such as fairness measures, or changing the optimality criterion to minimax regret, would require additional implementation and analysis.

Supplementary Materials

The supplementary materials include additional empirical and theoretical results, a simulation study, and proofs of all theoretical results.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

We acknowledge the partial support from Cisco Systems, Inc. (CG# 2370386), National Science Foundation (SES–2051196), Sloan Foundation (Economics Program; 2020–13946), National Natural Science Foundation of China (Grant No. 12371285, 12292984), and Arnold Ventures. We thank Benedikt Koch and anonymous reviewers of the IQSS's Alexander and Diviya Magaro Peer Pre-Review Program for useful feedback.

ORCID

Eli Ben-Michael http://orcid.org/0000-0002-1175-4129
Kosuke Imai http://orcid.org/0000-0002-2748-1022
Zhichao Jiang http://orcid.org/0000-0002-8571-0217

References

Athey, S., and Wager, S. (2021), "Policy Learning with Observational Data," *Econometrica*, 89, 133–161. [1389,1392]

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011), "Theory and Applications of Robust Optimization," *SIAM Review*, 53, 464–501. [1390]

Coston, A., Mishler, A., Kennedy, E. H., and Chouldechova, A. (2020), "Counterfactual Risk Assessments, Evaluation, and Fairness," in *FAT* 2020*, pp. 582–593. [1398]

Cui, Y. (2021), "Individualized Decision Making Under Partial Identification: Three Perspectives, Two Optimality Results, and One Paradox," *Harvard Data Science Review*, 3. [1390]

Duchi, J. C., and Namkoong, H. (2021), "Learning Models with Uniform Performance via Distributionally Robust Optimization," *Annals of Statistics*, 49, 1378–1406. [1390]

Dudik, M., Langford, J., and Li, L. (2011), "Doubly Robust Policy Evaluation and Learning," in *Proceedings of the 28th International Conference on Machine Learning*. [1389]

Gilboa, I., and Schmeidler, D. (1989), "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153. [1390]

Greiner, D. J., Halen, R., Stubenberg, M., Chistopher, J., and Griffen, L. (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. [1386,1387]

Imai, K., Jiang, Z., Greiner, D. J., Halen, R., and Shin, S. (2023), "Experimental Evaluation of Computer-Assisted Human Decision-Making: Application to Pretrial Risk Assessment Instrument," (with Discussion), *Journal of the Royal Statistical Society*, Series A, 186, 167–189. [1386,1387]

Imbens, G., and Wager, S. (2019), "Optimized Regression Discontinuity Designs," *The Review of Economics and Statistics*, 101, 264–278. [1393]

Jia, Z., Ben-Michael, E., and Imai, K. (2023), "Bayesian Safe Policy Learning with Chance Constrained Optimization: Application to Military Security Assessment During the Vietnam War." [1390]

Kallus, N., and Zhou, A. (2021), "Minimax-Optimal Policy Learning Under Unobserved Confounding," *Management Science*, 67, 2870–2890. [1390]

Kitagawa, T., and Tetenov, A. (2018), "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616. [1389,1392]

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, 4156–4165. [1392]

Manski, C. F. (2005), *Social Choice with Partial Knowledge of Treatment Response*, Princeton„ NJ: Princeton University Press. [1389,1390]

——— (2007), "Minimax-Regret Treatment Choice with Missing Outcome Data," *Journal of Econometrics*, 139, 105–115. [1390]

Neyman, J. (1990 [1923]), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science*, 5, 465–472. [1388]

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, 83, 318–345. [1390,1392]

Qian, M., and Murphy, S. A. (2011), "Performance Guarantees for Individualized Treatment Rules," *The Annals of Statistics*, 39, 1180–1210. [1389]

Rubin, D. B. (1980), "Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization Test," *Journal of the American Statistical Association*, 75, 591–593. [1388]

Song, K. (2014), "Point Decisions for Interval-Defined Parameters," *Econometric Theory*, 96, 334–356. [1390]

Stevenson, M. T., and Mayson, S. G. (2022), "Pretrial Detention and the Value of Liberty," *Virginia Law Review*, 108, 709–782. [1397]

Stoye, J. (2012), "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments," *Journal of Econometrics*, 166, 138–156. [1390]

Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press. [1391]

Zhang, Y., Ben-Michael, E., and Imai, K. (2023), "Safe Policy Learning under Regression Discontinuity Designs with Multiple Cutoffs," arXiv: 2208.13323. [1398]

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, 1106–1118. [1389]