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

A Additional theoretical results

A.1 Population optimality gap

We define the population width of function class $\mathcal{F}$ as

$$\mathcal{W}_{\mathcal{F}}(g) \;=\; \sup_{f\in\mathcal{F}} \mathbb{E}\left[\sum_{a\in\mathcal{A}} f(a,X)g(a,X)\right] - \inf_{f\in\mathcal{F}} \mathbb{E}\left[\sum_{a\in\mathcal{A}} f(a,X)g(a,X)\right].$$

Given this definition, the following theorem shows that the population optimality gap is bounded by the width of the function class.

Theorem A.1 (Population optimality gap)**.** Let $u(a) = u > 0$ for all actions $a \in \mathcal{A}$, and $\pi^{\inf}$ be a solution to Eqn (4). If $m^* \in \mathcal{M}$, the regret of $\pi^{\inf}$ relative to the optimal policy $\pi^* \in \operatorname{argmax}_{\pi\in\Pi} V(\pi)$ is

$$\frac{V(\pi^*) - V(\pi^{\inf})}{u} \leq \mathcal{W}_{\mathcal{M}}(\pi^*(1-\tilde{\pi})).$$

A.2 Extension of Theorems 1 and 2 to the case where $\alpha = 1$

In this section we extend Theorems 1 and 2 to include results for the case where we set $\alpha = 1$. We do so by providing bounds that hold regardless of the level α. In addition, we also provide a tighter bound on the optimality gap involving the difference between the true optimal policy π^* and the baseline policy $\tilde{\pi}$. For clarity, we present them with the bounds in Theorems 1 and 2 restated.

For a model class define the empirical support function as

$$h_{\mathcal{F}}(z) \equiv \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} z_{ia} f(X_i, a),$$

where $z = (z_{10}, \ldots, z_{1K-1}, \ldots, z_{n0}, \ldots, z_{nK-1})$ is a length $n(K-1)$ vector.

Theorem A.2 (Statistical safety (with $\alpha = 1$))**.** If the baseline policy $\tilde{\pi} \in \Pi$ and the true conditional expectation $m^*(a, x) \in \mathcal{M}$, for any $0 < \delta \leq e^{-1}$, the value of $\hat{\pi}^{\text{inf}}$ relative to the baseline $\tilde{\pi}$ is,

$$V(\tilde{\pi}) \; - V(\hat{\pi}) \leq \; 6C(K-1)\left[\max_a \mathcal{R}_n(\Pi_a) + 2\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}}\right] + \sup_{\pi\in\Pi} |h_{\widehat{\mathcal{M}}_n(\alpha)}\left(-\pi(1-\tilde{\pi})u(\cdot)\right) - h_{\mathcal{M}}\left(-\pi(1-\tilde{\pi})u(\cdot)\right)|,$$

with probability at least $1-\delta$, where $C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|$.

For the point-wise bounded model class that we consider, the extra term simplifies to be the worst-case difference between the true lower bound and the estimated lower bound.

Corollary A.1 (Statistical safety (with $\alpha = 1$))**.** Under the setting of Theorem A.2, let the restricted model class $\mathcal{M}$ and the empirical restricted model class $\widehat{\mathcal{M}}_n(\alpha)$ consist of point-wise bounded functions, $\mathcal{M} = \{f : \mathcal{A}\times\mathcal{X} \to \mathbb{R} \;\mid\; B_\ell(a,x) \leq f(a,x) \leq \mathbb{B}_u(a,x)\}$ and $\widehat{\mathcal{M}}_n(\alpha) = \{f : \mathcal{A}\times\mathcal{X} \to \mathbb{R} \;\mid\; \widehat{B}_{\alpha\ell}(a,x) \leq f(a,x) \leq \widehat{B}_{\alpha u}(a,x)\}$. Then the value of $\hat{\pi}^{\text{inf}}$ relative to the baseline $\tilde{\pi}$ is,

$$V(\tilde{\pi}) \; - V(\hat{\pi}) \leq \; 6C(K-1)\left[\max_a \mathcal{R}_n(\Pi_a) + 2\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}}\right] + 2C\sup_{a,x} |\widehat{B}_{\alpha\ell}(a,x) - B_\ell(a,x)|,$$

with probability at least $1-\delta$, where $C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|$.

Theorem A.3 (Optimality gap (with $\alpha = 1$))**.** 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

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

with probability at least $1-\delta$, where $C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|$.

The statement similarly simplifies under the point-wise bounded setting.

Corollary A.2 (Optimality gap (with $\alpha = 1$))**.** Under the setting of Theorem A.3, let the restricted model class $\mathcal{M}$ and the empirical restricted model class $\widehat{\mathcal{M}}_n(\alpha)$ consist of point-wise bounded functions, $\mathcal{M} = \{f : \mathcal{A}\times\mathcal{X} \to \mathbb{R} \;\mid\; B_\ell(a,x) \leq f(a,x) \leq \mathbb{B}_u(a,x)\}$ and $\widehat{\mathcal{M}}_n(\alpha) = \{f : \mathcal{A}\times\mathcal{X} \to \mathbb{R} \;\mid\; \widehat{B}_{\alpha\ell}(a,x) \leq f(a,x) \leq \widehat{B}_{\alpha u}(a,x)\}$. Then for any $0 < \delta \leq e^{-1}$, the optimality gap is

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

$$+ 2C \sup_{a,x} |\widehat{B}_{\alpha\ell}(a,x) - B_\ell(a,x)|,$$

with probability at least $1-\delta$, where $C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|$.

A.3 Learning from experiments evaluating a deterministic policy: a generic form of value function

Below we state a generic form of the value function with access to experimental data as in Section 4.5. The first line shows how to write the value of π in terms of the true CATE τ^* and the conditional expected outcome under the null action $m^*(-1,x)$. The second line further shows how to identify this expression with observable data via inverse probability weighting.

Proposition A.1. If $Z \perp\!\!\!\perp Y(a) \mid X$ and $0 < e(x) < 1$, then the expected utility can be written as

$$\begin{aligned} V(\pi) &= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)\{u(a)\tau^*(a,X) + c(a) + u(a)m^*(-1,X)\}\right] \\ &= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)u(a)\left[\tilde{\pi}(X,a)\left(\Gamma(1,X,Y) - \Gamma(0,X,Y)\right) + c(a) + u(a)\Gamma(0,X,Y)\right]\right] \\ &\quad + \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)u(a)\left\{1-\tilde{\pi}(X,a)\right\}\tau^*(a,X)\right], \end{aligned}$$

where $\Gamma(Z,X,Y) \equiv Y\{Z(1-2e(X)) + e(X)\}/\{e(X)(1-e(X))\}$ is the inverse probability weighted outcome.

Note that when the utility gain is constant, ($u(a) = u$ for all $a \in \mathcal{A}$), we have that

$$\mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)u(a)m(-1,X)\right] = u\mathbb{E}[m^*(-1,X)],$$

and does not depend on the policy, because $\sum_{a\in\mathcal{A}} \pi(X,a) = 1$.

B Proofs

Proof of Proposition 1.

$$V(\tilde{\pi}) = V^{\inf}(\tilde{\pi}) \leq V^{\inf}(\pi^{\inf}) \leq V(\pi^{\inf}).$$

□

Proof of Theorem A.1. Since $V^{\inf}(\pi) \leq V(\pi)$ for all policies π, the regret is bounded by

$$\begin{aligned} V(\pi^*) - V(\pi^{\inf}) &\leq V(\pi^*) - V^{\inf}(\pi^{\inf}) \\ &= V(\pi^*) - V^{\inf}(\pi^*) + V^{\inf}(\pi^*) - V^{\inf}(\pi^{\inf}). \end{aligned}$$

Now since $\pi^{\inf}$ is a maximizer of $V^{\inf}(\pi)$, $V^{\inf}(\pi^*) - V^{\inf}(\pi^{\inf}) \leq 0$. Now note that for any π,

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

This yields that

$$\begin{aligned} V(\pi^*) - V(\pi^{\inf}) &\leq \sum_{a\in\mathcal{A}} u(a)\mathbb{E}\left[\pi^*(X,a)\{1-\tilde{\pi}(X,a)\}m^*(a,X)\right] \\ &\qquad - \inf_{f\in\mathcal{M}} \sum_{a\in\mathcal{A}} u(a)\mathbb{E}\left[\pi^*(X,a)\{1-\tilde{\pi}(X,a)\}f(a,X)\right] \\ &\leq \sup_{f\in\mathcal{M}} \left\{\sum_{a\in\mathcal{A}} u(a)\mathbb{E}\left[\pi^*(X,a)\{1-\tilde{\pi}(X,a)\}f(a,X)\right]\right\} \\ &\qquad - \inf_{f\in\mathcal{M}} \left\{\sum_{a\in\mathcal{A}} u(a)\mathbb{E}\left[\pi^*(X,a)\{1-\tilde{\pi}(X,a)\}f(a,X)\right]\right\} \\ &= |u|\mathcal{W}_{\mathcal{M}}\left(\pi^*(1-\tilde{\pi})\right). \end{aligned}$$

□

Lemma B.1. Define $\hat{V}(\pi) \equiv \hat{V}(\pi, m^*)$ Then for any $0 < \delta < e^{-1}$,

$$\sup_{\pi\in\Pi} |\hat{V}(\pi) - V(\pi)| \leq 3C(K-1)\max_a \mathcal{R}_n(\Pi_a) + 5C(K-1)\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}}$$

with probability at least $1-\delta$, where $C = \max_{y\in\{0,1\},a\in\{0,1\}} |u(y,a)|$.

Proof of Lemma B.1. First, since $\sum_{a\in\mathcal{A}} \pi(x,a) = 1$ for all x, we can write the empirical value as

$$\begin{aligned} \hat{V}(\pi) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{K-1} \pi(X_i,a)\{u(a)[(\tilde{\pi}(X_i,a) - \tilde{\pi}(X_i,0))Y_i \\ + \{1-\tilde{\pi}(X_i,a)\}m^*(a,X_i) - (1-\tilde{\pi}(X_i,0)m^*(0,X_i))] + c(a) - c(0)\} \\ + u(0)[\tilde{\pi}(X_i,0)Y_i + (1-\tilde{\pi}(X_i,0))m^*(0,X_i)] + c(0). \end{aligned}$$

Now, define the function class with functions $f_a(x, y)$ as

$$\mathcal{F}_a = \{\pi(X_i, a)\{u(a)\left[(\tilde{\pi}(X_i, a) - \tilde{\pi}(X_i, 0))\, Y_i + \{1 - \tilde{\pi}(X_i, a)\}m^*(a, X_i) - (1 - \tilde{\pi}(X_i, 0)m^*(0, X_i))\right] + c(a) - c(0)\} \mid \pi(X_i, a) \in \Pi_a\},$$

where $\Pi_a = \{\mathbb{1}\{\pi(\cdot) = a\} \mid \pi \in \Pi\}$ is the set of all potential policy assignments to action a. Now notice that

$$\sup_{\pi \in \Pi} |\hat{V}(\pi) - V(\pi)| \leq \mathbb{E}_{X,Y,\varepsilon}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(u(0)\left[\tilde{\pi}(X_i, 0)Y_i + (1 - \tilde{\pi}(X_i, 0))m^*(0, X_i)\right] + c(0))\varepsilon_i\right|\right] + \sum_{a=1}^{K-1}\sup_{f_a \in \mathcal{F}_a}\left|\frac{1}{n}\sum_{i=1}^{n} f_a(X_i, Y_i) - \mathbb{E}\left[f_a(X, Y)\right]\right|.$$

The class $\mathcal{F}_a$ is uniformly bounded by twice the maximum absolute utility $C = \max_{y \in \{0,1\}, a \in \{0,1\}} |u(y, a)|$, so by Theorem 4.5 in Wainwright (2019)

$$\sup_{f_a \in \mathcal{F}_a}\left|\frac{1}{n}\sum_{i=1}^{n} f_a(X_i, Y_i) - \mathbb{E}\left[f_a(X, Y)\right]\right| \leq 2\mathcal{R}_n(\mathcal{F}_a) + t,$$

with probability at least $1 - \exp\left(-\frac{nt^2}{8C^2}\right)$. Now because

$$u(0)\left[\tilde{\pi}(X_i, 0)Y_i + (1 - \tilde{\pi}(X_i, 0))m^*(0, X_i)\right] \leq u(0),$$

and by independence of the data points and ε, we can get the bound

$$\begin{aligned}
&\mathbb{E}_{X,Y,\varepsilon}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(u(0)\left[\tilde{\pi}(X_i, 0)Y_i + (1 - \tilde{\pi}(X_i, 0))m^*(0, X_i)\right] + c(0))\varepsilon_i\right|\right] \\
\leq\ & \frac{1}{n}\left(\mathbb{E}_{\varepsilon}\left[\left(\sum_{i=1}^{n}(u(0)\left[\tilde{\pi}(X_i, 0)Y_i + (1 - \tilde{\pi}(X_i, 0))m^*(0, X_i)\right] + c(0))\varepsilon_i\right)^2\right]\right)^{\frac{1}{2}} \\
&= \frac{1}{n}\left(\mathbb{E}\left[\sum_{i=1}^{n}(u(0)\left[\tilde{\pi}(X_i, 0)Y_i + (1 - \tilde{\pi}(X_i, 0))m^*(0, X_i)\right] + c(0))^2\varepsilon_i^2\right]\right)^{\frac{1}{2}} \\
&\leq \frac{1}{n}\left(\sum_{i=1}^{n}(u(0) + c(0))^2\right)^{\frac{1}{2}} \\
&= \frac{|u(0) + c(0)|}{\sqrt{n}} \leq \frac{2C}{\sqrt{n}}.
\end{aligned}$$

Furthermore, since $\mathcal{F}_a$ consists of compositions of functions $g \in \Pi_a$ with linear functions

with a bounded slope,

$$|u(a)\left[(\tilde{\pi}(X_i,a)-\tilde{\pi}(X_i,0))\,Y_i+\{1-\tilde{\pi}(X_i,a)\}m^*(a,X_i)-(1-\tilde{\pi}(X_i,0)m^*(0,X_i))\right]+c(a)-c(0)|\leq 3C,$$

we can use the Talagrand contraction principle (Theorem 4.12 Ledoux and Talagrand, 1991) to bound the Rademacher complexity for $\mathcal{F}_a$ by

$$\mathcal{R}_n(\mathcal{F}_a)\leq 3C\mathcal{R}_n(\Pi_a).$$

Doing this for each $a=1,\ldots,K-1$ and using the union bound gives that

$$\sup_{\pi\in\Pi}|\hat{V}(\pi)-V(\pi)|\leq\frac{2C(K-1)}{\sqrt{n}}+3C(K-1)\max_{a\in\mathcal{A}}\mathcal{R}_n(\Pi_a)+t$$

with probability at least $1-(K-1)\exp\left(-\frac{nt^2}{8C^2}\right)$. Choosing $t=C\sqrt{\frac{8}{n}\log\frac{K-1}{\delta}}$ and noting that $2(K-1)+\sqrt{8\log\frac{K-1}{\delta}}\leq(2(K-1)+\sqrt{8})\sqrt{\log\frac{K-1}{\delta}}\leq(2+\sqrt{8})(K-1)\sqrt{\log\frac{K-1}{\delta}}\leq 5(K-1)\sqrt{\log\frac{K-1}{\delta}}$ gives the result.

□

Lemma B.2. For the empirical restricted model class $\widehat{\mathcal{M}}_n(\alpha)$ and for a policy $\pi\in\Pi$

$$\hat{V}^{\inf}(\hat{\pi})-\hat{V}(\hat{\pi})\leq\sup_{\pi\in\Pi}|h_{\widehat{\mathcal{M}}_n(\alpha)}\left(-\pi(1-\tilde{\pi})u(\cdot)\right)-h_{\mathcal{M}}\left(-\pi(1-\tilde{\pi})u(\cdot)\right)|,$$

where the function $(\pi(1-\tilde{\pi})u(\cdot))(x,a)=\pi(x,a)*(1-\tilde{\pi}(x,a))u(a)$. Furthermore, if $\mathcal{M}\subseteq\widehat{\mathcal{M}}_n(\alpha)$, then

$$\hat{V}^{\inf}(\hat{\pi})-\hat{V}(\hat{\pi})\leq 0.$$

Proof of Lemma B.2. For a model class $\mathcal{F}$, define $\tilde{V}(\pi,\mathcal{F})=\inf_{f\in\mathcal{F}}\hat{V}(\pi,f)$, and define $\hat{V}^{\inf}(\pi)=\min_{m\in\widehat{\mathcal{M}}_n(\alpha)}\hat{V}(\pi,m)$, so that $\hat{V}^{\inf}(\pi)=\tilde{V}(\pi,\widehat{M}_n(\alpha))$. This implies that

$$\begin{aligned}\hat{V}^{\inf}(\hat{\pi})-\hat{V}(\hat{\pi})&\leq\hat{V}^{\inf}(\hat{\pi})-\tilde{V}(\hat{\pi},\mathcal{M})\\&=\tilde{V}(\hat{\pi},\widehat{\mathcal{M}}_n(\alpha))-\tilde{V}(\hat{\pi},\mathcal{M}).\end{aligned}$$

Now note that if $\mathcal{M}\subseteq\widehat{\mathcal{M}}_n(\alpha)$, then $\tilde{V}(\hat{\pi},\widehat{\mathcal{M}}_n(\alpha))-\tilde{V}(\hat{\pi},\mathcal{M})\leq 0$. Otherwise we can write this difference as

$$\begin{aligned}\tilde{V}(\hat{\pi},\widehat{\mathcal{M}}_n(\alpha))-\tilde{V}(\hat{\pi},\mathcal{M})=&\inf_{m\in\widehat{\mathcal{M}}_n(\alpha)}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}}\pi(X_i,a)(1-\tilde{\pi}(X_i,a))u(a)m(a,X_i)\right\}\\&-\inf_{m\in\mathcal{M}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}}\pi(X_i,a)(1-\tilde{\pi}(X_i,a))u(a)m(a,X_i)\right\}\end{aligned}$$

$$
\begin{aligned}
&= - \sup_{m \in \widehat{\mathcal{M}}_n(\alpha)} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \sum_{a \in \mathcal{A}} \pi(X_i, a)(1 - \tilde{\pi}(X_i, a)) u(a) m(a, X_i) \right\} \\
&\quad + \sup_{m \in \mathcal{M}} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \sum_{a \in \mathcal{A}} \pi(X_i, a)(1 - \tilde{\pi}(X_i, a)) u(a) m(a, X_i) \right\} \\
&= -h_{\widehat{\mathcal{M}}_n(\alpha)}(-\pi(1 - \tilde{\pi})u(\cdot)) + h_{\mathcal{M}}(-\pi(1 - \tilde{\pi})u(\cdot))
\end{aligned}
$$

Taking the supremeum over all possible policies $\hat{\pi} \in \Pi$ gives the result. □

Proof of Theorems 1 and A.2. The difference in values between $\tilde{\pi}$ and $\hat{\pi}$ is

$$
\begin{aligned}
V(\tilde{\pi}) - V(\hat{\pi}) &= V(\tilde{\pi}) - \hat{V}(\tilde{\pi}) + \hat{V}(\tilde{\pi}) - \hat{V}(\hat{\pi}) + \hat{V}(\hat{\pi}) - V(\hat{\pi}) \\
&\leq 2 \sup_{\pi \in \Pi} |\hat{V}(\pi) - V(\pi)| + \hat{V}(\tilde{\pi}) - \hat{V}(\hat{\pi})
\end{aligned}
$$

We have bounded the first term in Lemma B.1. To bound the second term, notice that

$$
\begin{aligned}
\hat{V}(\tilde{\pi}) - \hat{V}(\hat{\pi}) &= \hat{V}^{\inf}(\tilde{\pi}) - \hat{V}(\hat{\pi}) \\
&= \underbrace{\hat{V}^{\inf}(\tilde{\pi}) - \hat{V}^{\inf}(\hat{\pi})}_{\leq 0} + \hat{V}^{\inf}(\hat{\pi}) - \hat{V}(\hat{\pi}) \\
&\leq \hat{V}^{\inf}(\hat{\pi}) - \hat{V}(\hat{\pi}),
\end{aligned}
$$

where we have used that $\hat{\pi}$ maximizes $\hat{V}^{\inf}(\pi)$.

In Lemma B.2 we have bounded this difference. Combining the two bounds we have that

$$
\begin{aligned}
V(\tilde{\pi}) - V(\hat{\pi}) \leq 6C(K-1) \max_a \mathcal{R}_n(\Pi_a) + 10C(K-1)\sqrt{\frac{1}{n} \log \frac{K-1}{\delta}} \\
+ \sup_{\pi \in \Pi} |h_{\widehat{\mathcal{M}}_n(\alpha)}(-\pi(1 - \tilde{\pi})u(\cdot)) - h_{\mathcal{M}}(-\pi(1 - \tilde{\pi})u(\cdot))|,
\end{aligned}
$$

with probability at least $1 - \delta$. And if $\mathcal{M} \subseteq \widehat{\mathcal{M}}_n(\alpha)$, then we have the further bound

$$
V(\tilde{\pi}) - V(\hat{\pi}) \leq 6C(K-1) \max_a \mathcal{R}_n(\Pi_a) + 10C(K-1)\sqrt{\frac{1}{n} \log \frac{K-1}{\delta}},
$$

with probability at least $1 - \delta$. Noting that $P(\mathcal{M} \subseteq \widehat{\mathcal{M}}_n(\alpha)) \geq 1 - \alpha$, and taking the union bound gives that this second bound holds with probability at least $1 - \delta - \alpha$.

□

Proof of Theorems 2 and A.3. The regret of $\hat{\pi}$ relative to π^* is

$$
\begin{aligned}
V(\pi^*) - V(\hat{\pi}) &= V(\pi^*) - \hat{V}(\pi^*) + \hat{V}(\pi^*) - \hat{V}(\hat{\pi}) + \hat{V}(\hat{\pi}) - V(\hat{\pi}) \\
&\leq \sup_{\pi \in \Pi} 2|\hat{V}(\pi) - V(\pi)| + \hat{V}(\pi^*) - \hat{V}(\hat{\pi}).
\end{aligned}
$$

We have bounded the first term in Lemma B.1, and we now turn to the second term.

$$\begin{aligned}\hat{V}(\pi^*) - \hat{V}(\hat{\pi}) &= \hat{V}(\pi^*) - \hat{V}^{\inf}(\hat{\pi}) + \hat{V}^{\inf}(\hat{\pi}) - \hat{V}(\hat{\pi}) \\ &= \hat{V}(\pi^*) - \hat{V}^{\inf}(\pi^*) + \underbrace{\hat{V}^{\inf}(\pi^*) - \hat{V}^{\inf}(\hat{\pi})}_{\leq 0} + \hat{V}^{\inf}(\hat{\pi}) - \hat{V}(\hat{\pi}) \\ &\leq \hat{V}(\pi^*) - \hat{V}^{\inf}(\pi^*) + \hat{V}^{\inf}(\hat{\pi}) - \hat{V}(\hat{\pi}),\end{aligned}$$

where we have again used that $\hat{\pi}$ maximizes $\hat{V}^{\inf}(\pi)$. We have bounded the second term in Lemma B.2, now we turn to the first term:

$$\begin{aligned}\hat{V}(\pi^*) - \hat{V}^{\inf}(\pi^*) &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} u(a)\pi^*(X_i,a)\{1-\tilde{\pi}(X_i,a)\}m^*(a,X_i) \\ &\quad - \inf_{f\in\widehat{\mathcal{M}}_n(\alpha)} \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} u(a)\pi^*(X_i,a))\{1-\tilde{\pi}(X_i,a)\}f(a,X_i) \\ &\leq \sup_{f\in\widehat{\mathcal{M}}_n(\alpha)} \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} u(a)\pi^*(X_i,a))\{1-\tilde{\pi}(X_i,a)\}f(a,X_i) \\ &\quad - \inf_{f\in\widehat{\mathcal{M}}_n(\alpha)} \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} u(a)\pi^*(X_i,a))\{1-\tilde{\pi}(X_i,a)\}f(a,X_i) \\ &= |u|\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}\left(\pi^*(1-\tilde{\pi})\right).\end{aligned}$$

Combined with Lemmas B.1 and B.2 and the union bound this gives that

$$\begin{aligned}V(\pi^*) - V(\hat{\pi}) = |u|\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}\left(\pi^*(1-\tilde{\pi})\right) + 6C(K-1)\max_a \mathcal{R}_n(\Pi_a) + 10C(K-1)\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}} \\ + |u|\sup_{\pi\in\Pi}\left|h_{\widehat{\mathcal{M}}_n(\alpha)}\left(-\pi(1-\tilde{\pi})\right) - h_{\mathcal{M}}\left(-\pi(1-\tilde{\pi})\right)\right|,\end{aligned}$$

with probability at least $1-\delta$ and

$$V(\pi^*) - V(\hat{\pi}) = |u|\widehat{\mathcal{W}}_{\widehat{\mathcal{M}}_n(\alpha)}\left(\pi^*(1-\tilde{\pi})\right) + 6C(K-1)\max_a \mathcal{R}_n(\Pi_a) + 10C(K-1)\sqrt{\frac{1}{n}\log\frac{K-1}{\delta}},$$

with probability at least $1-\delta-\alpha$. Noting that $|u| \leq 2C$ gives the result.

□

Proof of Corollary A.1 and A.2. These Corollaries follow from noting that

$$\begin{aligned}&h_{\widehat{\mathcal{M}}_n(\alpha)}\left(-\pi(1-\tilde{\pi})u(\cdot)\right) - h_{\mathcal{M}}\left(-\pi(1-\tilde{\pi})u(\cdot)\right) \\ =&\frac{1}{n}\sum_{i=1}^{n}\pi(X_i,a)(1-\tilde{\pi}(X_i,a))u(a)\widehat{B}_{\alpha\ell}(X_i,a) - \frac{1}{n}\sum_{i=1}^{n}\pi(X_i,a)(1-\tilde{\pi}(X_i,a))u(a)B_{\ell}(X_i,a) \\ =&\frac{1}{n}\sum_{i=1}^{n}\pi(X_i,a)(1-\tilde{\pi}(X_i,a))u(a)(\widehat{B}_{\alpha\ell}(X_i,a) - B_{\ell}(X_i,a))\end{aligned}$$

$$\leq C \sup_{x,a} |\widehat{B}_{\alpha\ell}(x,a) - B_\ell(x,a)|.$$

□

Proof of Proposition A.1. For the first equality, note that $m^*(a,X) = \tau^*(a,X) + m^*(-1,X)$. So the first equality follows by plugging in this equality to Equation (1). Next, note that we can decompose this expression as in Equation (2) to get that

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

Now using that $\mathbb{E}[\Gamma(Z,X,Y) \mid Z=z, X=x] = z\cdot m^*(\tilde{\pi}(x),x) + (1-z)\cdot m^*(-1,x)$, and noting that $\tilde{\tau}(x) = \mathbb{E}[\Gamma(1,X,Y) - \Gamma(1,X,Y) \mid X=x]$, gives the second expression. □

C Computation for restricted model classes

In this section, we show, in detail, how to compute the population and empirical model classes in a variety of cases: no restrictions, Lipschitz functions, linear models, and additive models.

First, for point-wise bounded model classes, we can compute the size term in Theorem 2 by looking for the policy $\pi \in \Pi$ that disagrees with the baseline policy $\tilde{\pi}$ when the upper and lower bounds are farthest apart:

$$\widehat{\mathcal{S}}(\widehat{\mathcal{T}}_n(\alpha), \Pi; \tilde{\pi}) = \sup_{\pi\in\Pi} \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} \pi(X_i,a)(1-\tilde{\pi}(X_i,a))\left(\widehat{B}_{\alpha u}(a,X_i) - \widehat{B}_{\alpha\ell}(a,X_i)\right). \quad \text{(C.1)}$$

C.1 No restrictions

Suppose that the conditional expectation has no restrictions, other than that it must lie between L and U, i.e., $\mathcal{F} = \{f \mid L \leq f(a,x) \leq U \;\; \forall a \in \mathcal{A},\; x \in \mathcal{X}\}$. The restricted model class $\mathcal{M} = \{f \in \mathcal{F} \mid f(a,x) = \tilde{m}(x) \;\text{ for } a \text{ with } \tilde{\pi}(x) = a\}$ provides no additional information when the policy π disagrees with the baseline policy $\tilde{\pi}$. The upper and lower bounds are given by $B_u(a,x) = \tilde{\pi}(x,a)\tilde{m}(x) + \{1-\tilde{\pi}(x,a)\}U$ and $B_\ell(a,x) = \tilde{\pi}(x,a)\tilde{m}(x) + \{1-\tilde{\pi}(x,a)\}L$, respectively.

To construct the empirical model class $\widehat{\mathcal{M}}_n(\alpha)$, we begin with a simultaneous $1-\alpha$ confidence interval for the conditional expectation function $\tilde{m}(x)$, with lower and upper bounds $\widehat{C}_\alpha(x) = [\widehat{C}_{\alpha\ell}(x), \widehat{C}_{\alpha u}(x)]$ such that

$$P\left(\tilde{m}(x) \in \widehat{C}_\alpha(x) \;\; \forall\, x \in \mathcal{X}\right) \geq 1-\alpha. \quad \text{(C.2)}$$

See Srinivas et al. (2010); Chowdhury and Gopalan (2017); Fiedler et al. (2021) for examples

on constructing such simultaneous bounds via kernel methods in statistical control settings. With this confidence band, we can use the upper and lower bounds of the confidence band in place of the true conditional expectation $\tilde{m}(x)$, i.e. $\widehat{B}_{\alpha u}(a,x) = \tilde{\pi}(X,a)\widehat{C}_{\alpha u}(x) + \{1 - \tilde{\pi}(X,a)\}U$ and $\widehat{B}_{\ell}(a,x) = \tilde{\pi}(X,a)\widehat{C}_{\alpha \ell}(x)\{1 - \tilde{\pi}(x,a)\}L$.

C.2 Lipschitz functions

We next consider the case where the covariate space $\mathcal{X}$ has a norm $\|\cdot\|$, and that $m(a,\cdot)$ is a λ_a-Lipschitz function,

$$\mathcal{F} = \{f : \mathcal{A} \times \mathcal{X} \to \mathbb{R} \quad | \quad |f(a,x) - f(a,x')| \leq \lambda_a \|x - x'\|\}.$$

Taking the greatest lower bound and least upper bound implied by this model class leads to lower and upper bounds, $B_\ell(a,x) = \sup_{x' \in \tilde{\mathcal{X}}_a} \{\tilde{m}(x') - \lambda_a \|x - x'\|\}$, and $B_u(a,x) = \inf_{x' \in \tilde{\mathcal{X}}_a} \{\tilde{m}(x') + \lambda_a \|x - x'\|\}$, where $\tilde{\mathcal{X}}_a = \{x \in \mathcal{X} \mid \tilde{\pi}(x) = a\}$ is the set of covariates with the baseline policy giving action a. The further we extrapolate from the area where the baseline action $\tilde{\pi}(x) = a$, the larger the value of $\|x - x'\|$ will be and so there will be more ignorance about the values of the function.

The size of $\mathcal{M}$ will depend on the expected distance to the boundary between baseline actions and the value of the Lipschitz constant. If most individuals are close to the boundary, or the Lipschitz constant is small, $\mathcal{M}$ will be small and the safe policy will be close to optimal. Conversely, a large number of individuals far away from the boundary or a large Lipschitz constant will increase the potential for suboptimality.

To construct the empirical version, we again use a simultaneous confidence band $\widehat{C}_\alpha(x)$ satisfying Equation (C.2). Then, the lower and upper bounds use the lower and upper confidence limits in place of the function values, $\widehat{B}_{\alpha\ell}(a,X) = \sup_{x' \in \tilde{\mathcal{X}}_a} \left\{\widehat{C}_{\alpha\ell}(x') - \lambda_a \|X - x'\|\right\}$ and $\widehat{B}_{\alpha u}(a,X) = \inf_{x' \in \tilde{\mathcal{X}}_a} \left\{\widehat{C}_{\alpha u}(x') - \lambda_a \|X - x'\|\right\}$. In our analysis of the NVCA flag threshold in Section 5.2, the covariate space $\mathcal{X}$ is discrete, so we construct a simultaneous confidence interval via the a Bonferroni correction on the 7 unique values.

Note that it is also possible to construct bounds using a finite set of evaluation points. For example, if $\check{\mathcal{X}}_a$ is a finite set of points such that the baseline policy satisfies $\tilde{\pi}(x) = a$, an alternative procedure to construct a lower bound is to take the greatest lower bound over the finite set $\check{\mathcal{X}}_a$, i.e.

$$\check{B}_\ell(a,x) = \max_{x' \in \check{\mathcal{X}}_a} \tilde{m}(x') - \lambda_a \|x - x'\|.$$

Because the finite set $\check{\mathcal{X}}_a \subseteq \tilde{\mathcal{X}}_a$, the greatest lower bound over $\check{\mathcal{X}}_a$ will be less than or equal to the greatest lowest bound over the entire set $\tilde{\mathcal{X}}_a$, i.e. $\check{B}_\ell(a,x) \geq B_\ell(a,x)$. With this finite set, we can create the empirical version using a simultaneous confidence band $\check{C}_\alpha(x)$ over

only $\check{\mathcal{X}}_a$ that satisfies

$$P\left(\tilde{m}(x) \in \check{C}_\alpha(x) \ \ \forall\, x \in \check{X}_a\right) \geq 1 - \alpha.$$

Such a bound can be constructed with a simple Bonferroni correction, or via a more tailored approach. Then the empirical lower bound would be $\check{B}_{\alpha\ell}(a, X) = \max_{x' \in \check{\mathcal{X}}_a} \left\{\check{C}_{\alpha\ell}(x') - \lambda_a \|X - x'\|\right\}$. Unlike in the population case, the empirical lower bound using the finite set, $\check{B}_{\alpha\ell}(a, x)$, may be greater than the empirical lower bound using the simultaneous confidence band $\widehat{B}_{\alpha\ell}(a, x)$ if the simultaneous confidence band over the entire set $\mathcal{X}$ is wider than that over the smaller, finite set $\check{X}_a$.

C.3 Linear models

We next consider, as a model class, a linear model in a set of basis functions $\phi : \mathcal{A} \times \mathcal{X} \to \mathbb{R}^d$, $\mathcal{F} = \{f(a, x) = h^{-1}(b^\top \phi(a, x))\}$, where we still enforce the upper and lower bounds of U and L. The restricted model class is the set of coefficients b that satisfy $\tilde{m}(x) = b^\top \phi(a, x)$ for all x and a such that $\tilde{\pi}(x) = a$. With discrete covariates, this is a linear system of equations. Slightly abusing notation, define $\phi(A, X) \in \mathbb{R}^{p \times dK}$ as the matrix of values $\phi(a, x)$ for the p unique combinations observable in the data, and $\tilde{m} \in \mathbb{R}^p$ as the corresponding values of $\tilde{m}(x)$. If the model class is not point identified (e.g. if $p < dK$), then there will be infinitely many solutions to the equation $\tilde{m} = \phi(A, X)b$. To characterize these, define β^* as the *minimum norm solution*:

$$\begin{aligned} &\min_{b \in \mathbb{R}^d} \ \|b\|_2 \\ &\text{subject to} \ \ \tilde{m} = \phi(A, X)b. \end{aligned}$$

There will also be an unidentified component arising from the *null space* of the system of linear equations, $\mathcal{N} = \{b \in \mathbb{R}^d \mid \phi(A, X)b = 0\}$. Let $D \in \mathbb{R}^{d \times d^\perp}$ be an orthonormal basis for this null space. Then, any solution to the linear equations $\tilde{m} = \phi(A, X)b$ can be written as the minimum norm solution β^* plus a vector in the null space, which we can write as $Db_{\mathcal{N}}$, where $b_{\mathcal{N}}$ are free parameters. Therefore, we can re-write the restricted model in terms of these free parameters:

$$\mathcal{M} \ = \ \{f(a, x) = (\beta^* + Db_{\mathcal{N}})^\top \phi(a, x) \ \ \mid \ \ b_{\mathcal{N}} \in \mathbb{R}^{d^\perp}\}.$$

Finding the worst-case value will involve a non-linear optimization over $b_{\mathcal{N}}$. Rather than taking such an approach, we will consider a larger class $\overline{\mathcal{M}} \equiv \{f \ \ \mid \ \ B_\ell(a, x) \leq f(a, x) \leq B_u(a, x)\}$ that contains the restricted model class $\mathcal{M}$. To construct it, we choose the upper and lower bounds

$$B_\ell(a, x) = {\beta^*}^\top \phi(a, x) \mathbb{1}\{D^\top \phi(a, x) = 0\} + \mathbb{1}\{D^\top \phi(a, x) \neq 0\} L,$$

$$B_u(a,x) = {\beta^*}^\top \phi(a,x) \mathbb{1}\{D^\top \phi(a,x) = 0\} + \mathbb{1}\{D^\top \phi(a,x) \neq 0\} U.$$

For a given action a and covariate vector x, we first check whether $\phi(x,a)$ is in the null space $\mathcal{N}$ by checking whether $D^\top \phi(a,x) = 0$. If it is not in the null space (i.e. $D^\top \phi(a,x) = 0$), then the lower and upper bounds are equal, $B_\ell(a,x) = B_u(a,x) = h^{-1}({\beta^*}^\top \phi(a,x))$ because for any choice of free parameter $b_\mathcal{N}$, $b_\mathcal{N}^\top D^\top \phi(a,x) = 0$. In contrast, if $\phi(a,x)$ is in the null space (i.e. $D^\top \phi(a,x) \neq 0$), then the free parameter is unrestrained and $(\beta^* + D b_\mathcal{N})^\top \phi(a,x)$ can take on any value between L and U.

To construct the empirical model class we again begin with a simultaneous confidence band, this time for the minimum norm prediction, $\beta^* \cdot \phi(a,x) \in [\widehat{C}_{\alpha\ell}(a,x), \widehat{C}_{\alpha u}(a,x)]$ where we apply a Bonferroni correction for the p unique observed values

$$\beta^* \cdot \phi(a,x) \in \hat{\beta}^* \cdot \phi(a,x) \pm \hat{\sigma} t_{n-p-1,1-\frac{\alpha}{2p}} \sqrt{\phi(a,x)^\top (\Phi^\top \Phi)^\dagger \phi(a,x)},$$

where $\hat{\beta}^*$ is the least squares estimate of the minimum norm solution, $\hat{\sigma}^2$ is the estimate of the variance from the MSE, $\Phi = [\phi(\tilde{\pi}(x_i), x_i)]_{i=1}^n \in \mathbb{R}^{n \times d}$ is the design matrix,$t_{n-p-1,1-\frac{\alpha}{2p}}$ is the is the $1-\alpha/p$ quantile of an t distribution $n-p-1$ degrees of freedom, and $A^\dagger$ denotes the pseudo-inverse of a matrix A. This gives lower and upper bounds,

$$\begin{aligned}
\widehat{B}_{\alpha\ell}(a,x) &= \max\{L, \widehat{C}_{\alpha\ell}(a,x)\} \mathbb{1}\{D^\top \phi(a,x) = 0\} + L \mathbb{1}\{D^\top \phi(a,x) \neq 0\}, \\
\widehat{B}_{\alpha u}(a,x) &= \min\{U, \widehat{C}_{\alpha u}(a,x)\} \mathbb{1}\{D^\top \phi(a,x) = 0\} + U \mathbb{1}\{D^\top \phi(a,x) \neq 0\},
\end{aligned}$$

where we enforce the constraint that the predictions must be between L and U post-hoc.

C.4 Additive models

If the model class for action a consists of additive models, we have

$$\mathcal{F} = \left\{ f(a,x) = \sum_{j=1}^d f_j(a, x_j) + \sum_{j<k} f_{jk}(a, (x_j, x_k)) + \dots \;\middle|\; f_j(a,\cdot), f_{jk}(a,\cdot), \dots, \lambda_a - \text{Lipschitz} \right\},$$

where the component functions $f_j(a,\cdot), f_{jk}(a,\cdot), \dots$ can be subject to additional restrictions so that the decomposition is unique. This additive decomposition formulation amounts to an assumption that no interactions exist above a certain order.

By using the same additive decomposition for $\tilde{m}(x)$ into $\tilde{m}(x) = \sum_j \tilde{m}_j(X_j) + \sum_{j<k} \tilde{m}_{jk}(X_j, X_k) + \dots$, we can follow the same bounding approach as in Appendix C.2 for each of the component functions. For example, for the additive term for covariate j, $m_j(a, x_j)$, the Lipschitz property implies that,

$$\tilde{m}_j(x_j') - \lambda_a |x_j - x_j'| \leq m_j(a, x_j) \leq \tilde{m}(x_j') + \lambda_a |x_j - x_j'| \quad \forall\, x' \in \tilde{\mathcal{X}}_a.$$

Taking the greatest lower bound and least upper bound for each component function, the

overall lower and upper bounds are,

$$\begin{aligned} B_\ell(a,X) &= \sum_j \sup_{x'\in\tilde{\mathcal{X}}_a} \left\{m_j(x'_j) - \lambda_a|X_j - x'_j|\right\} + \sum_{j<k} \sup_{x'\in\tilde{\mathcal{X}}_a} \left\{m_{jk}(x'_j,x'_k) - \lambda_a\|X_{(j,k)} - x'_{(j,k)}\|\right\} + \cdots \\ B_u(a,X) &= \sum_j \inf_{x'\in\tilde{\mathcal{X}}_a} \left\{m_j(x'_j) + \lambda_a|X_j - x'_j|\right\} + \sum_{j<k} \inf_{x'\in\tilde{\mathcal{X}}_a} \left\{m_{jk}(x'_j,x'_k) + \lambda_a\|X_{(j,k)} - x'_{(j,k)}\|\right\} + \cdots, \end{aligned} \tag{C.3}$$

where $x_{(j,k)}$ is the subvector of components j and k of x. Unlike in Appendix C.2, this extrapolates covariate by covariate, finding the tightest bounds for each component. For instance, for a first-order additive model, the level of extrapolation depends on the distance in each covariate $|x_j - x'_j|$ separately.

To construct the empirical model class for the class of additive models, we use a $1-\alpha$ confidence interval that holds simultaneously over all values of x and for all components, i.e.,

$$\tilde{m}_j(x_j) \in \widehat{C}^{(j)}_\alpha(x_j), \quad m_{jk}(x_j,x_k) \in \widehat{C}^{(j,k)}_\alpha(x_j,x_k), \ldots, \quad \forall\, j = 1,\ldots,d, \quad k<j, \quad \ldots,$$

with probability at least $1-\alpha$. Analogous to the Lipschitz case in Appendix C.2 above, we can then construct the lower and upper bounds using the lower and upper bounds of the confidence intervals,

$$\begin{aligned} \widehat{B}_{\alpha\ell}(a,X) &= \sum_j \sup_{x'\in\tilde{\mathcal{X}}_a} \left\{\widehat{C}^{(j)}_{\alpha\ell}(x'_j) - \lambda|X_j - x'_j|\right\} + \sum_{j<k} \sup_{x'\in\tilde{\mathcal{X}}_a} \left\{\widehat{C}^{(j,k)}_{\alpha\ell}(x'_j,x'_k) - \lambda\|X_{(j,k)} - x'_{(j,k)}\|\right\} + \ldots \\ B_{\alpha u}(a,X) &= \sum_j \inf_{x'\in\tilde{\mathcal{X}}_a} \left\{\widehat{C}^{(j)}_{\alpha u}(x'_j) + \lambda|X_j - x'_j|\right\} + \sum_{j<k} \inf_{x'\in\tilde{\mathcal{X}}_a} \left\{\widehat{C}^{(j,k)}_{\alpha u}(x'_j,x'_k) + \lambda\|X_{(j,k)} - x'_{(j,k)}\|\right\} + \ldots. \end{aligned}$$

D Incorporating human decision-making

The PSA-DMF system we study is an example of a "human-in-the-loop" framework: rather than an algorithmic policy being the final arbiter of decisions, the policy merely provides recommendations to a human that makes an ultimate decision (Imai et al., 2023; Ben-Michael et al., 2024). In this section, we formalize and extend the potential outcomes framework to incorporate human decisions, and then briefly explore how our framework can be extended to explicitly model human decisions and apply it to learn a new NVCA system.

D.1 Potential human decisions and potential outcomes

We first show how to extend our framework to incorporate human decisions. Let $D_i(a) \in \{0,1\}$ be the potential (binary) decision for individual i under action $a \in \mathcal{A}$ (an algorithmic recommendation in our application), and $Y_i(d,a) \in \{0,1\}$ be the potential (binary) outcome for individual i under human decision $d \in \{0,1\}$ and algorithmic action $a \in \mathcal{A}$. This setup nests our main framework. To see this, note that we can re-define the the potential outcome under algorithmic action a as the potential outcome when the algorithmic action is set to a

and the human decision is the natural value under algorithmic action a:

$$Y_i(a) \equiv Y_i(D_i(a), a) = Y_i(0, a)(1 - D(a)) + Y_i(1, a)D(a).$$

If the human decision under algorithmic action a is $D(a) = 0$, then the potential outcome under algorithmic action a is $Y_i(a) = Y_i(0, a)$. Conversely, if the human decision under algorithmic action a is $D(a) = 1$, the potential outcome under algorithmic action a is $Y_i(a) = Y_i(1, a)$. Then, the observed decision is given by $D_i = D(\tilde{\pi}(X_i))$ whereas the observed outcome is $Y_i = Y_i(\tilde{\pi}(X_i)) = Y_i(D_i(\tilde{\pi}(X_i)), \tilde{\pi}(X_i))$.

Finally, we denote the expected potential human decision under algorithmic action a, conditional on covariates x, as $d(a, x) = \mathbb{E}[D(a) \mid X = x]$ and represent the conditional expectation of the potential outcome under algorithmic action a, conditional on covariates x, as $m(a, x) = \mathbb{E}[Y(a) = 1 \mid X = x]$.

D.2 Incorporating human decisions into the utility function

To incorporate human decisions into the utility function, we write the utility for outcome y under human decision d as $u(y, d)$. With this setup, the value for a policy π is:

$$V(\pi) \;=\; \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X, a) \sum_{d=0}^{1} \left[u(1, d)Y(d, a) + u(0, d)(1 - Y(d, a))\right] \mathbb{1}\{D(a) = d\}\right].$$

If we make the simplifying assumption that the utility gain is constant across decisions, i.e., $u(1, d) - u(0, d) = u$ for $d \in \{0, 1\}$, we can index the utility for $y = 0$ and $d = 0$ as $u(0, 0) = 0$, and denote the added cost of taking decision 1 as $c = u(0, 1) - u(0, 0)$. This allows us to write the value by marginalizing over the potential decisions, yielding,

$$V(\pi) \;=\; \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X, a)\left(uY(a) + cD(a)\right)\right]. \tag{D.1}$$

Comparing Equation (D.1) to the value in Equation (2) when actions are taken directly, we see that the key difference is the inclusion of the potential decision $D(a)$ in determining the cost of an action. Rather than directly assigning a cost to an action a, there is an indirect cost associated with the eventual decision $D(a)$ that action a induces in the decision maker. Therefore, the unidentifiability of the expected potential decision under an action given the covariates, $d(a, x)$, also must enter the robustness procedure.

We can treat the unidentifiability of the potential decisions in a manner parallel to the outcomes. Denoting the conditional expected observed decision as $d(\tilde{\pi}(x), x) = \mathbb{E}[D \mid X = x]$, we can posit a model class for the decisions $\mathcal{F}'$ and create the restricted model class $\mathcal{D} = \{f \in \mathcal{F}' \mid f(\tilde{\pi}(x), x) = d(\tilde{\pi}(x), x)\}$.[1] We can now construct a population safe policy by

[1] These restrictions being on the *decisions* gives more opportunities for structural restrictions on the

maximizing the worst case value across the model classes for both the outcomes $\mathcal{M}$ and the decisions $\mathcal{D}$,

$$\max_{\pi\in\Pi}\left\{\mathbb{E}\left[\sum_{a\in\mathcal{A}}\pi(X,a)\tilde{\pi}(X,a)uY\right]+\min_{f\in\mathcal{M}}\mathbb{E}\left[\sum_{a\in\mathcal{A}}\pi(X,a)\{1-\tilde{\pi}(X,a)\}uf(a,X)\right]\right.$$
$$\left.+\mathbb{E}\left[\sum_{a\in\mathcal{A}}\pi(X,a)\tilde{\pi}(X,a)cD\right]+\min_{g\in\mathcal{D}}\mathbb{E}\left[\sum_{a\in\mathcal{A}}\pi(X,a)\{1-\tilde{\pi}(X,a)\}cg(a,X)\right]\right\}. \tag{D.2}$$

By allowing for actions to affect decisions through the decision maker rather than directly, the costs of actions are not fully identified. Therefore, we now find the worst-case expected outcome *and decision* when determining the worst case value in Equation (D.2). In essence, we solve the inner optimization twice: once over outcomes for the restricted outcome model class $\mathcal{M}$ and once over decisions for the restricted decision model class $\mathcal{D}$.

From here, we can follow the development in the previous sections. We create empirical restricted model classes for the outcome and decision functions, $\widehat{\mathcal{M}}_n(\alpha/2)$ and $\widehat{D}_n(\alpha/2)$ using a Bonferonni correction so that $P(\mathcal{M}\in\widehat{\mathcal{M}}_n(\alpha/2),\mathcal{D}\in\widehat{D}_n(\alpha/2))\geq 1-\alpha$. Then, we solve the empirical analog to Equation (D.2). Finally, we can incorporate experimental evidence as above. In this case, the conditional expected potential decision $d(a,x)$ and outcome $m(a,x)$ — and their model classes — are replaced with the conditional average treatment effect on the decision $\mathbb{E}[D(a)-D(-1)\mid X=x]$ and on the outcome $\tau(a,x)$.

D.3 Learning a new NVCA point system

In Section 5, we only considered the outcomes of triggering the NVCA flag and have assigned costs directly to the flag. However, the PSA serves as a recommendation to the presiding judge who is the ultimate decision maker. Following the discussion above, we can incorporate this into the construction of the robust policy. Rather than place a cost on triggering the NVCA flag, we use the judge's binary decision of whether to assign a signature bond or cash bail and place a cost of -1 to assigning cash bail. Unlike the cost directly placed on the NVCA flag, this allows us to address the cost of cash bail decision. As discussed in Section 5, the cost of the judge's decision to assign cash bail includes the fiscal and socioeconomic costs, indexed to be -1.

Following the same analysis as in Section 5, we find maximin policies that take the decisions into account for increasing costs of an NVCA relative to assigning cash bail, at various confidence levels. For the additive and second order effect models, however, we find policies that differ from the original rule only when we do not take the statistical uncertainty into account — with confidence level $1-\alpha=0$ — and have no finite sample guarantee that the new policy is not worse than the existing rule. In this case, the policy is extremely

model. For example, we could make a monotonicity assumption that $d(a,x)\leq d(a',x)$ for $a\leq a'$.

aggressive, responding to noise in the treatment effects. Otherwise, we cannot find a new policy that safely improves on the original rule. This is primarily because the overall effects of the PSA on both the judge's decisions and defendant's behavior are small (Imai et al., 2023). Therefore, there is too much uncertainty to ensure that a new policy would reliably improve upon the existing rule.

E Imputation, IPW, and double robust methods

Here we briefly discuss how standard approaches to policy learning are not applicable in our setting. First, as discussed in Section 3.2, the key identification issue is that we can cannot point-identify the conditional expectation of the potential outcome $m^*(a,x) = \mathbb{E}[Y(a) \mid X = x]$ for all pairs of actions a and covariates x. In settings with overlap ($P(A = a \mid X = x) > 0$ for all $a \in \mathcal{A}$ and $x \in \mathcal{X}$), and unconfounded action assignment ($A \perp\!\!\!\perp \{Y(0), Y(1), \ldots, Y(K-1)\} \mid X$), we can identify $m^*(a,x)$ via the conditional expectation of the observed outcome given the action and the covariate $\tilde{m}(a,x) \equiv \mathbb{E}[Y \mid A = a, X = x]$. In such settings, we could then identify the value $V(\pi)$ using model-based imputation, IPW, or augmented IPW:

$$
\begin{aligned}
V(\pi, m^*) &= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)\{u(a)\tilde{m}(a,X) + c(a)\}\right] && \text{(Imputation)} \\
&= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)\left\{u(a)\frac{\mathbb{1}\{A=a\}}{P(A=a \mid X)}Y + c(a)\right\}\right] && \text{(IPW)} \\
&= \mathbb{E}\left[\sum_{a\in\mathcal{A}} \pi(X,a)\left\{u(a)\left(\tilde{m}(a,X) + \frac{\mathbb{1}\{A=a\}}{P(A=a \mid X)}(Y - m(a,X))\right) + c(a)\right\}\right] && \text{(AIPW)}
\end{aligned}
$$

In our setting, where the observed actions are the actions under the deterministic baseline policy $A_i = \tilde{\pi}(X_i)$, the actions are unconfounded given the covariates X (indeed, we know exactly how the actions are assigned), but there is no overlap because $P(A = a \mid X) = P(\tilde{\pi}(X) = a \mid X)$ is either 0 or 1. The implication is that the outcome model $m^*(a,x)$ is not point identifiable. It is impossible to estimate the conditional expectation of the observed outcome given $A = a$ and $X = x$, $\tilde{m}(a,x)$, if $a \neq \tilde{\pi}(x)$ because it is an event of measure zero (i.e. $P(A \neq \tilde{\pi}(X)) = 0$).

Nonetheless, we may try to use the imputation approach by estimating a model $\hat{m}(a,x)$ and relying on it for extrapolation. We would then solve

$$
\hat{\pi}^{\text{impute}} \in \max_{\pi\in\Pi} \frac{1}{n}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}} \pi(X_i,a)\left\{u(a)\left(\tilde{\pi}(X_i,a)Y + (1-\tilde{\pi}(X_i,a))\hat{m}(a,X_i)\right) + c(a)\right\}. \quad \text{(E.1)}
$$

This imputation-based policy will be highly sensitive to how the estimated model $\hat{m}(a,X_i)$ extrapolates to combinations of a and x that are not possible under the baseline policy, as

we show via simulation in Section F.

The identification problem is more transparent for the IPW and AIPW-based approaches. Note that the inverse probability term with a deterministic baseline policy is $\mathbb{1}\{A = a\}/\tilde{\pi}(a, X_i)$, which is equal to $\tilde{\pi}(a, X_i)/\tilde{\pi}(a, X_i)$. If $\tilde{\pi}(a, X_i) = 1$, then this term is equal to 1, but if $\tilde{\pi}(a, X_i) = 0$, it is $0/0$, which is undefined. Again, we may nonetheless try to use IPW by setting $0/0 = 0$. This would give:

$$\hat{\pi}^{\text{ipw}} \in \max_{\pi \in \Pi} \frac{1}{n} \sum_{i=1}^{n} \sum_{a \in \mathcal{A}} \pi(X_i, a) \left\{ u(a) \tilde{\pi}(X_i, a) Y + c(a) \right\}.$$

As long as $u(a) > c(a)$, then defining the IPW-based policy in this way will give that $\hat{\pi}^{\text{ipw}} = \tilde{\pi}$, and so we will always keep the baseline policy.

Finally, we might try to consider the AIPW estimator, again setting $0/0{=}0$, but note that

$$\hat{m}(a, X_i) + \tilde{\pi}(X_i, a)(Y_i - \hat{m}(a, X_i)) = \tilde{\pi}(X_i, a) Y_i - (1 - \tilde{\pi}(X_i, a)) \hat{m}(a, X_i),$$

and so the AIPW approach would recover the model-based imputation approach.

F Simulation study

We have a single discrete covariate with 10 levels, $x \in \{0, \ldots, 9\}$, and a binary action so that the action set is $\mathcal{A} = \{0, 1\}$. We choose a baseline policy $\tilde{\pi} = \mathbb{1}\{x \geq 5\}$, and set the utility gain to be $u(0) = u(1) = 10$ and the costs to be $c(0) = 0, c(1) = -1$, so that action 0 is costless and action 1 costs one tenth of the potential utility gain. For each simulation we draw n i.i.d. samples $X_1, \ldots, X_n$ uniformly on $\{0, \ldots, 9\}$. Then we draw a smooth model for the expected control potential outcome $m(0, x) \equiv \mathbb{E}[Y(0) \mid X = x]$ via random Fourier features. We draw three random vectors: $\omega \in \mathbb{R}^{100}$ with i.i.d. standard normal elements; $b \in \mathbb{R}^{100}$ with i.i.d. components drawn uniformly on $[0, 2\pi]$; and $\beta \in \mathbb{R}^{100}$ with i.i.d. standard normal elements. Then we set

$$m(0, x) = \text{logit}^{-1}\left(\sqrt{\frac{2}{100}} \beta \cdot \cos\left(\omega \frac{x}{9} + b \right) \right),$$

where the cosine operates element-wise. See Rahimi and Recht (2008) for more discussion on random features. For the potential outcome under treatment, $m(1, x) = \mathbb{E}[Y(1) \mid X = x]$, we add a linear treatment effect on the logit scale:

$$m(1, x) = \text{logit}^{-1}\left(\text{logit}\left(m(0, x) \right) + \frac{1}{2}\left(x - \frac{9}{2} \right) - \frac{8}{10} \right).$$

We then generate the potential outcomes $Y_i(0), Y_i(1)$ as independent Bernoulli draws with probabilities $m(0, X_i)$ and $m(1, X_i)$, respectively.

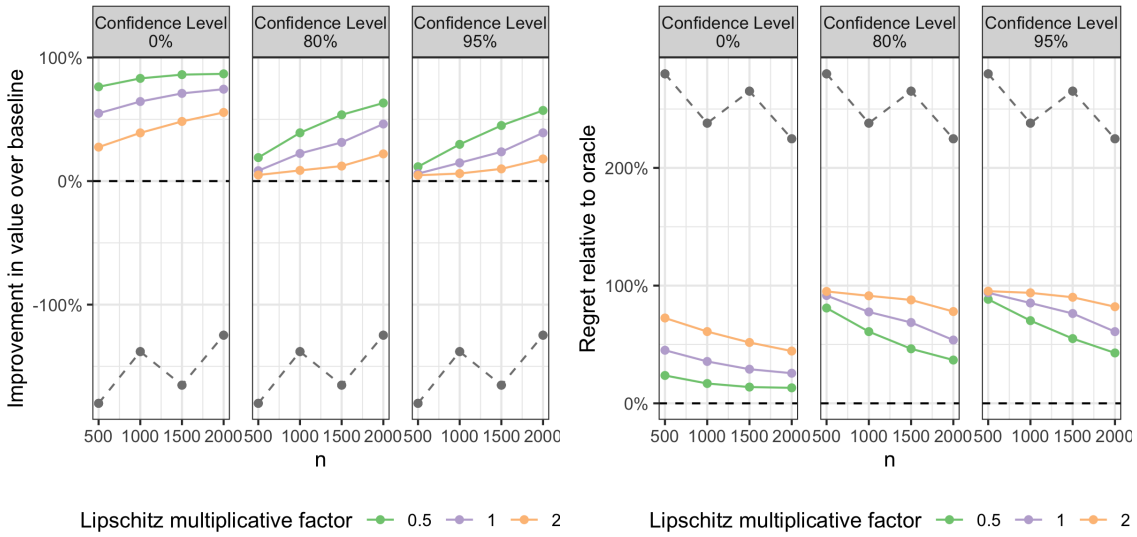


Figure F.1: Monte Carlo simulation results as the sample size n increases, varying the multiplicative factor on the empirical Lipschitz constant and the significance level $1-\alpha$. The left panel shows the difference in the expected utility between the empirical safe policy $\hat{\pi}$, and the baseline policy $\tilde{\pi}$, normalized by the regret of the baseline relative to the oracle, i.e. $\frac{V(\hat{\pi})-V(\tilde{\pi})}{V(\pi^*)-V(\tilde{\pi})}$. The right panel shows the regret of the safe policy relative to the oracle, scaled by the regret of the baseline relative to the oracle, i.e. $\frac{V(\pi^*)-V(\hat{\pi})}{V(\pi^*)-V(\tilde{\pi})}$. In both panels, the grey dashed line represents the imputation-based policy.

With each simulation draw, we consider finding a safe empirical policy by solving Equation (7) under a Lipschitz restriction on the model as in Appendix C.2 and with the threshold policy class Π_{thresh}. Note that the true model is in fact much smoother than Lipschitz; here we consider using the looser assumption. Following our empirical analysis in Section 5.2, we take the average outcome at each value of x, and compute the largest difference in consecutive averages as pilot estimates for the Lipschitz constants λ_0 and λ_1. We then solve Equation (7) using $\frac{1}{2}$, 1, and 2 times these pilot estimates as the Lipschitz constants, and setting the significance level to 0, 80% and 95%.

We also consider using a model-based imputation estimator without accounting for partial identification. Because the baseline policy assigns 0 for $x \in \{0,1,2,3,4\}$ and 1 for $x \in \{5,6,7,8,9\}$, there are 5 unique values of the covariate when $\tilde{\pi}(x)$ is 0 or 1. Therefore, we fit two separate non-parametric models for $\hat{m}(0,x)$ and $\hat{m}(1,x)$ by fitting a logistic regression of Y on X with a degree four polynomial of X. This creates 5 parameters for each model, one for each unique observed data point. We then use each estimated 4-degree polynomial logistic regression model to extrapolate $\hat{m}(0,x)$ for $x \geq 5$ and $\hat{m}(1,x)$ for $x < 5$ and estimate an imputation-based policy $\hat{\pi}^{\text{impute}}$ solving Equation (E.1). We additionally compute the oracle threshold policy that uses the true model values $m(0,x)$ and $m(1,x)$. We do this for sample sizes $n \in (500, 1000, 1500, 2000)$.

Figure F.1 shows how the empirical safe policy $\hat{\pi}$ and the model-based imputation policy $\hat{\pi}^{\text{impute}}$ compare to both the baseline policy $\tilde{\pi}$ and the oracle policy π^* in terms of expected

utility. First, we see that on average, the empirical safe policy improves over the baseline, no matter the confidence level and the choice of Lipschitz constant. This improvement is larger the less conservative we are, e.g. by choosing a lower confidence level or a smaller Lipschitz constant. Furthermore, as the sample size increases, the utility of the empirical safe policy also increases due to a lower degree of statistical uncertainty. We find similar behavior when comparing it to the oracle policy. Less conservative choices lead to lower regret, and the regret decreases with the sample size. Importantly, the regret does not decrease to zero; even when removing all statistical uncertainty the safe policy can still be suboptimal due to the lack of identification.

In contrast, model-based imputation without accounting for identification issues performs poorly, yielding a policy that has much lower expected utility than the baseline, let alone the oracle. This is because the extrapolation to unseen data does not perform well with the modeling approach that we used. It could have been possible to choose an imputation estimator that performs better in that the extrapolation proved to be correct. However, for any imputation estimator we can come up with an adversarial example where the extrapolation is incorrect and leads to a worse policy than the status quo. Indeed, this is precisely what the maximin criterion is designed to defend against.

G Additional empirical results

In this section, we present additional empirical results for the FTA, NCA, and NVCA scores, as well as the results for the combined bail level and monitoring conditions recommendation. For reference, Table G.1 displays the existing risk-factor weights for the FTA, NCA, and NVCA risk scores.

G.1 Additional results for the NVCA threshold and score

We begin by presenting the results regarding the NVCA threshold. Figure G.1 shows how the maximin threshold changes as we vary the confidence level $1-\alpha$ while setting $C = 3$. The overall relationship between the threshold and the cost is robust to the choice of confidence level. The results show that when the cost of an NVCA is low and/or the confidence level is low the learned safe policy will raise the threshold, implying that fewer arrestees will trigger the NVCA flag.

Figure G.2 shows estimates of the effect of providing the PSA on whether the judge makes a cash bail decision, and on whether the arrestee engages in an NVCA, conditioned on the number of total NVCA points. We find that when the NVCA flag is not triggered (i.e. $x_{\text{nvca}} < 4$) there is little to no effect of providing the PSA on either the judge's decision or the presence of an NVCA. This appears to remain true when $x_{\text{nvca}} = 4$, even though the flag is triggered. For $x_{\text{nvca}} \geq 5$, providing the PSA increases the proportion of decisions

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	

Table G.1: Weights placed on risk factors to construct the failure to appear (FTA), new criminal activity (NCA), and new violent criminal activity (NVCA) scores. The sum of the weights is then thresholded into six levels for the FTA and NCA scores and a binary "Yes"/"No" for the NVCA score.

that are cash bail by over 30 percentage points (though this is not significant for $x_{\text{nvca}} = 6$.) However, NVCAs do not meaningfully change for $x_{\text{nvca}} = 5$, even though there are over 30 percentage points more cash bail decisions, but they decrease for $x_{\text{nvca}} = 6$.

Next, we present several additional empirical results for the NVCA threshold and score.

Second order effect model and model testing. First, Figure G.3a shows how the maximin NVCA flag differs from the original rule as the cost of an NVCA and the confidence level vary under the second order effect model. We find that under the second order effect model, there is too much uncertainty to safely deviate from the original NVCA flag rule with any reasonable degree of confidence if the cost of an NVCA is greater than 1. This is in contrast to the results under the additive effect model shown in Figure 4a; the addition of unidentifiable second order interaction terms precludes safely changing the policy.

To understand whether the additive effects assumption is reasonable for the NVCA rule, we estimate the CATE separately for arrestees with and without the NVCA flag triggered via a similar spirit to the DR-learner (Kennedy, 2022) by regressing the IP-weighted outcomes $\Gamma(1, \boldsymbol{X}, Y) - \Gamma(0, \boldsymbol{X}, Y)$ on the 7 binary risk factors and all observed pair-wise interactions. Note that this partial second order model is point identified because it omits the unidentified terms and so it is only a rough proxy for the full second order model. We then test whether the interaction terms are all zero using a Wald test with Huber-White heteroskedastic robust standard errors. We do not find evidence against the null of the additive model for cases where the flag is not triggered ($p = 0.75$), but there is some evidence for the existence of

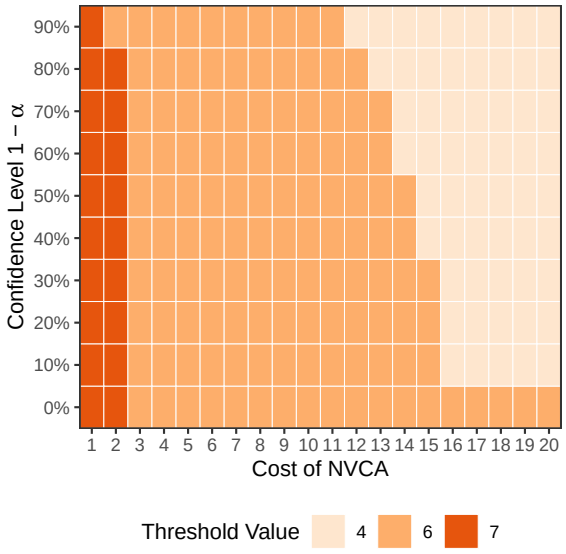


Figure G.1: Learned threshold values solving Equation (7) for the NVCA flag threshold rule as the cost of an NVCA increases from 1 to 20 times the cost of triggering the NVCA flag, and the confidence level varies between 0% and 90%.

interactions when the flag is triggered ($p = 0.067$).

Using a quadratic cost. We consider an alternative value function that assigns a larger marginal utility loss to triggering the NVCA flag for an arrestee if a larger proportion of arrestees have the flag triggered. Formally, defining $\bar{\pi} \equiv \mathbb{E}[\pi(X)]$, the policy value function is given by:

$$V^{\text{quad}}(\pi) \equiv \mathbb{E}\left[\pi(X)\left\{u \times (m^*(1,X) - m^*(0,X)) - (1 + \zeta\bar{\pi})\right\}\right] + \mathbb{E}[m^*(0,X)].$$

This induces a *quadratic* cost, with ζ determining the additional marginal penalization per percent flagged as an NVCA risk. Note that this value function is not an expectation of individual utilities, because the cost of flagging one individual for NVCA risk depends on how many other individuals are also flagged. As with the cost of an NVCA u, it is beyond the scope of this paper to argue for a particular value of the quadratic penalty term ζ, and so we will document how the policy changes as it varies. Note that other forms of such utilities are possible, for example, we could consider a step function that adds an additional penalty if the number of arrestees flagged as an NVCA risk exceeds some threshold.

Figure G.4 shows how the maximin rule compares to the original rule, again in terms of the the performance of the maximin proportion of arrestees flagged for an NVCA risk as we vary both u and ζ while keeping the confidence level fixed to $1 - \alpha = 80\%$. For any given cost of an NVCA, the maximin policy triggers the flag less often as the quadratic penalty increases.

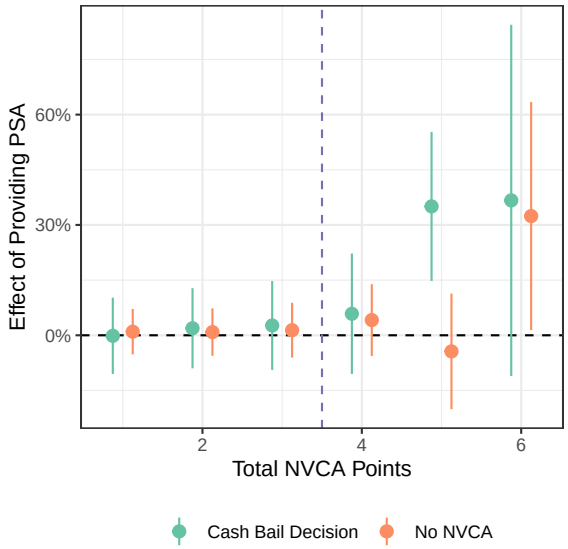


Figure G.2: The effect of providing the PSA on (a) whether the judge makes a cash bail decision and (b) whether the arrestee does not engage in an NVCA, conditioned on the number of total NVCA points. Error bars indicate 95% confidence intervals using heteroskedastic robust standard errors. The vertical dashed line represents the existing NVCA threshold.

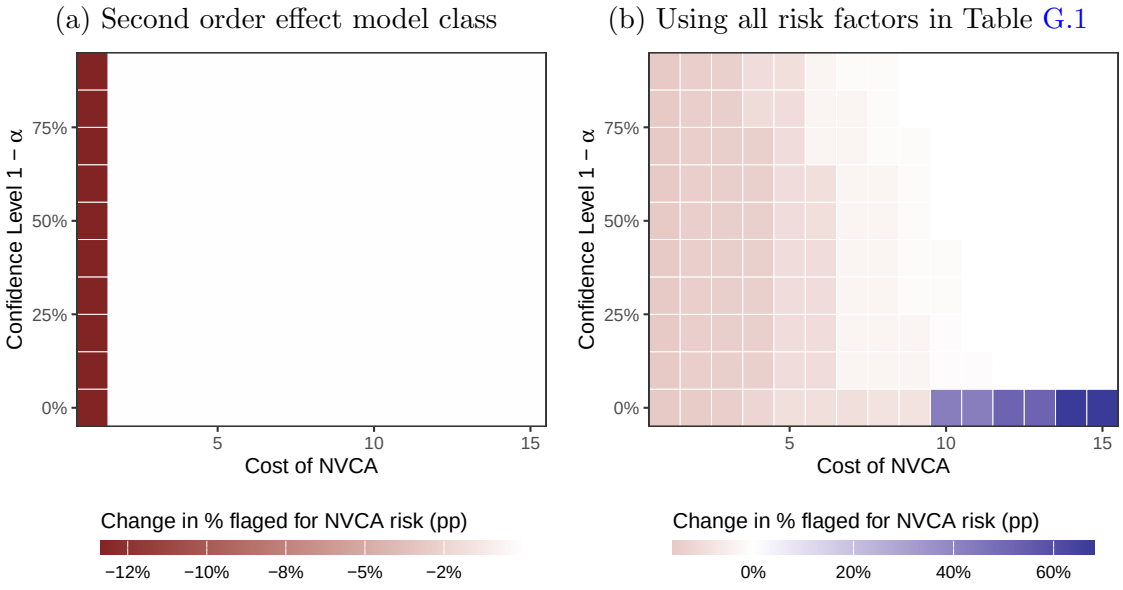


Figure G.3: The percentage point difference in the proportion of arrestees flagged for NVCA risk between the maximin policy and the original NVCA score as the cost of an NVCA increases from 1 to 15 times of the cost of triggering the NVCA flag and the confidence level varies between 0% and 100% (a) in the second order effect model class and (b) under the additive effect model class using all risk factors in Table G.1.

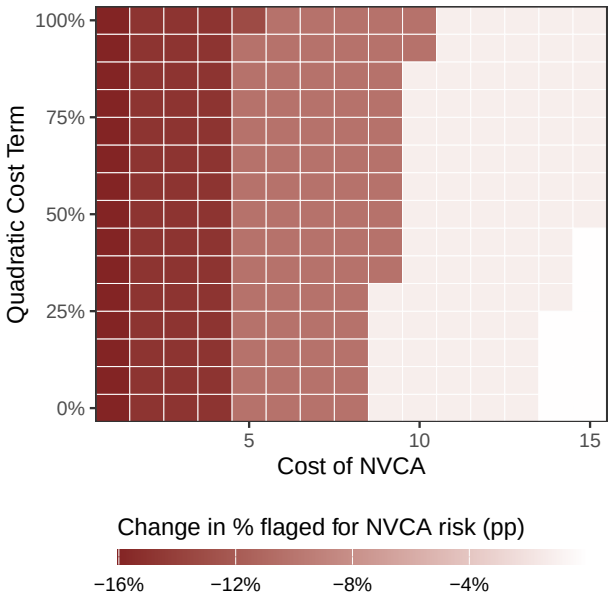


Figure G.4: The percentage point difference in the proportion of arrestees flagged for NVCA risk between the maximin policy and the original NVCA score as the cost of an NVCA increases from 1 to 15 times of the cost of triggering the NVCA flag as ζ varies with a confidence level of 80%.

Figure G.5 shows the integer weights on the risk factors for the maximin policy at the $1 - \alpha = 80\%$ level as the quadratic penalty ζ increases with the cost of an NVCA set to 9. Increasing the quadratic penalty eventually changes the maximin policy back to placing less weight on violent convictions and offenses, similar to the results when we only vary the cost of an NVCA and keep $\zeta = 0$ (e.g. in Figure 4b).

Using the full set of risk factors. We also consider learning a new NVCA flag rule that incorporates the full set of risk factors listed in Table G.1. The scale of the weight placed on each factor is not necessarily meaningful for comparisons across rules that use different risk factors and thresholds. For this reason, we place an upper bound on the weights of 5.

Figure G.3b shows how the resulting maximin rules differ from the original NVCA flag rule, again as the cost of an NVCA and the confidence level vary under the additive effect model, with a quadratic penalty term of zero, i.e., $\zeta = 0$. We find broadly similar results as when using the original reduced set of risk factors. For all confidence levels at lower NVCA costs, the maximin rule classifies fewer arrestees as NVCA risks, eventually collapsing back to the status quo as the cost of an NVCA relative to the cost of triggering the flag increases. Relative to the reduced covariate set, including more risk factors increases the level of statistical uncertainty, and so the maximin rule collapses back to the original rule more quickly.

Relative to the reduced covariate set, including more risk factors increases the level of

statistical uncertainty, and so the maximin rule collapses back to the original rule more quickly. In addition, at a confidence level of 0%, the learned NVCA flag rule eventually begins to flag far more arrestees as NVCA risks than the original rule as the cost of an NVCA increases. However, because more risk factors are included, even when the maximin policy does not differ from the baseline in terms of which arrestees it triggers the flag for, the underlying risk factor weights can be different, as multiple combinations of weights can produce the same recommendations. Figure G.6 shows the set of weights found during the optimization problem with a confidence level of 80%, but as the solutions are not unique and the scales arbitrary, these weights are not directly comparable to the other sets of results.

G.2 Additional results for the FTA and NCA scores

Next, we present additional empirical results for the FTA and NCA scoring systems. We begin by formalizing the FTA and NCA policy classes as follows:

$$\Pi = \left\{ \pi(x) = \sum_{a=1}^{K-1} a \mathbb{1}\left\{\eta_{a-1} < \theta \cdot x \leq \eta_a\right\} \;\middle|\; \theta \in \mathbb{Z}^d,\ \eta_a > \eta_{a-1} \geq 0\ \forall a \in \{1, 2, \ldots, K-1\} \right\},$$

where x are the corresponding risk factors in either the FTA or NCA rule, θ are the integer weights placed on the risk factors, and $\eta_0, \ldots, \eta_{K-1}$ are thresholds that determine what the

Figure G.5: NVCA flag weights θ in Equation (G.2). Change in θ as the quadratic penalty ζ increases from 0 to with a cost an NVCA equal to 9 and a confidence level of 80% (right panel).

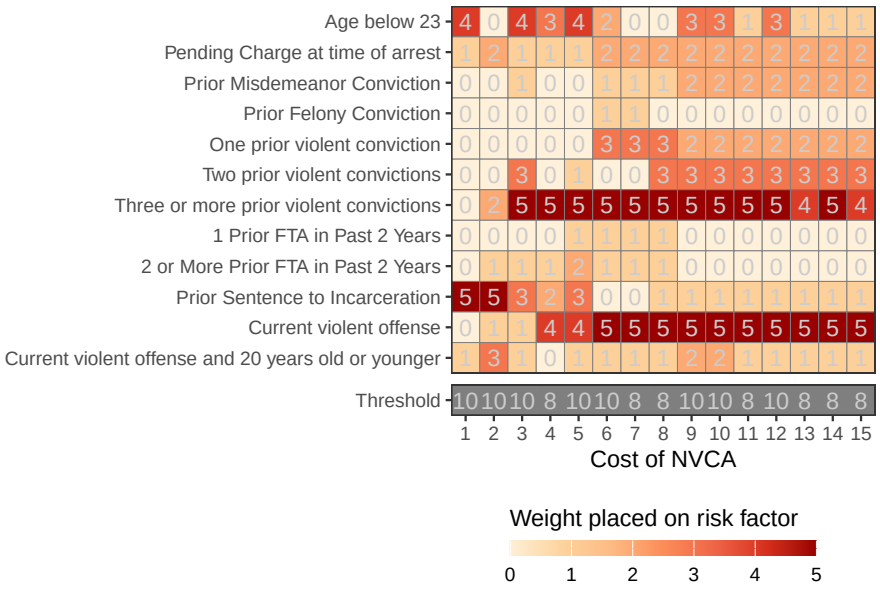


Figure G.6: Change in the NVCA flag weights θ using all of the risk factors in Table G.1 as the cost of an NVCA increases from 1 to 15 times the cost of triggering the NVCA flag, at a confidence level of $1 - \alpha = 80\%$ and no quadratic penalty $\zeta = 0$.

final score is. For example, the baseline FTA rule has thresholds (0, 1, 2, 4, 6, 7) and the baseline NCA rule has thresholds (0, 2, 4, 6, 8, 13).

There are $K = 6$ possible actions for the FTA and NCA scores, each giving scores between 1 and 6. Indexing the cost of the first action to be zero, we must characterize the cost of the remaining 5 actions. There are many potential ways to do so. However, recall from Figure 3 that there is little information to extrapolate from the NCA score and none for the FTA score, so we do not expect to be able to learn maximin policies that are different from the status quo here. Therefore, we extend our utility function from the binary case to a simple linear parameterization of the costs, writing the utility function as $u(y, a) = u \times y - a$ where $|u|$ is the cost of either an FTA or an NCA depending on the risk score. This utility function and these costs are not directly comparable to the binary utility function for the NVCA flag, because the cost for choosing the highest score is indexed to 5 rather than 1 as in the binary case.[2] We note that it is straightforward to encode different cost structures.

Figure G.7 shows how the maximin FTA and NCA scores differ from the original rules as we vary the cost of an FTA or NCA and the confidence level $1 - \alpha$. Overall, we find that with any degree of statistical confidence, if the cost of an FTA or NCA is above 2, the maximin rule collapses to the status quo rule. This is not surprising given the discussion in Section 5.3.

It may be possible, however, to learn simplified versions of the FTA and NCA scores that

[2]Recall that we index the first action to be $a = 0$.

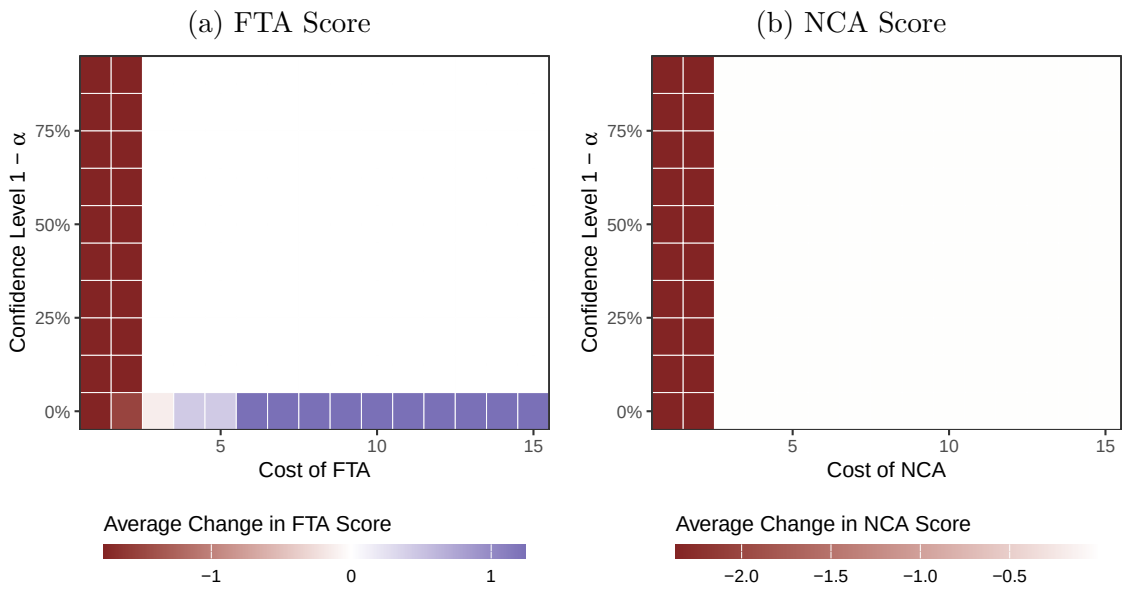


Figure G.7: The average difference in (a) the FTA score and (b) the NCA score for arrestees under the maximin policy and the original FTA and NCA scores as the cost of an FTA (left panel) and NCA (right panel) increases from 1 to 15 and the confidence level varies between 0% and 100%.

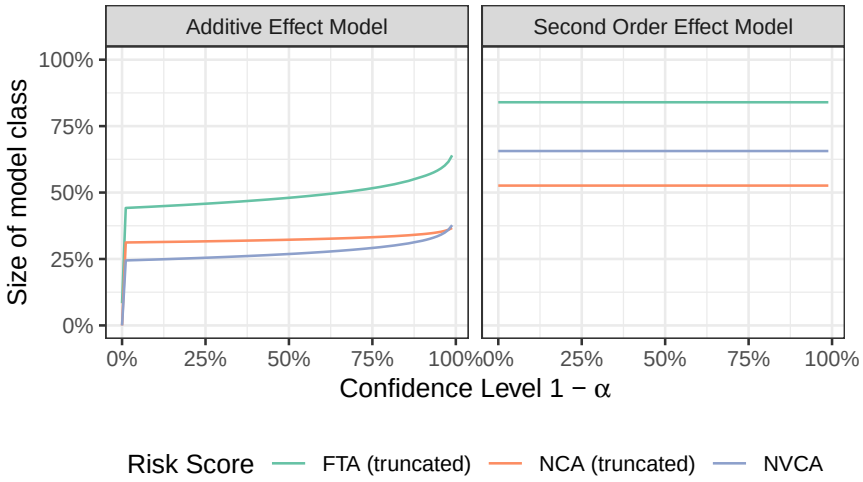


Figure G.8: The size (as a percentage of its maximum value) of two different model classes with respect to the linear threshold policy class versus the confidence level $1-\alpha$ for the FTA (green) and NCA (orange), both truncated into an indicator for high risk (score greater than or equal to 4) and NVCA (purple) scoring rules.

are collapsed into low and high risk. To inspect this, we create truncated versions of the scores that are indicators for whether the scores are greater than or equal to 4. Figure G.8 shows the sizes of the resulting model classes with respect to the truncated policy classes for

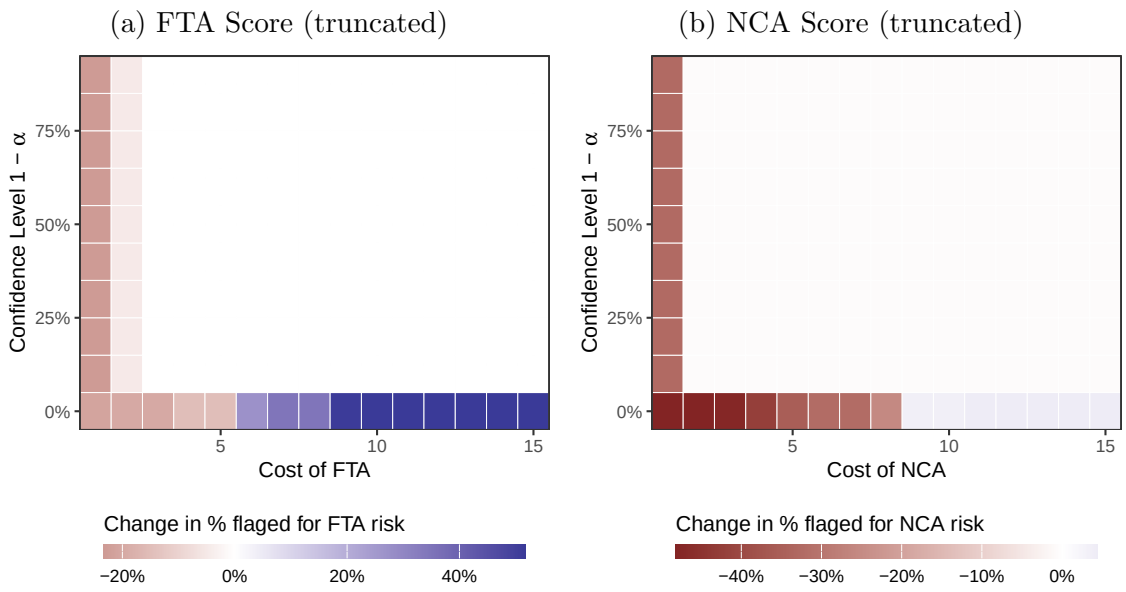


Figure G.9: The percentage point difference in the proportion of arrestees flagged for (a) FTA risk and (b) NCA risk under the maximin policy and the original FTA and NCA scores truncated into low and high risk values as the cost of an FTA (left panel) and NCA (right panel) increases from 1 to 15 and the confidence level varies between 0% and 100%.

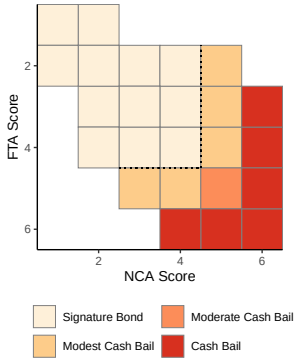
both the additive and second order effect models as the confidence level varies, keeping the NVCA flag for comparison. We find that truncating the scores leads to much smaller model classes. This suggests that it might be possible to learn maximin policies that deviate from the status quo.

We learn such maximin policies using the binary utility function used for the NVCA, and truncating the policy class to output either a low or high risk. Figure G.9 shows how the resulting truncated scores differ from the original truncated scores under the additive effect class as the cost of an FTA or NCA and the confidence level vary. We find the same pattern as in Figure G.7. With any degree of statistical confidence, it is not possible to safely change the underlying scores. Since the sizes of the model classes are smaller with respect to the truncated policy classes, the results suggest that there exist substantial uncertainty as to the heterogeneous effects even for the truncated FTA and NCA scores.

G.3 Additional results for the overall DMF risk score and quaternary and ternary bail recommendations

Testing for interactions. In our main analysis for the binary cash bail recommendation, we use an additive model effective model where $\tau_{\text{add}}(a, \boldsymbol{x}) = \tau_{\text{fta}}(a, x_{\text{fta}}) + \tau_{\text{nca}}(a, x_{\text{nca}})$. We can assess the plausibility of this assumption following the same procedure as in Section G.1

(a) Bail recommendation



(b) Release and monitoring conditions recommendation

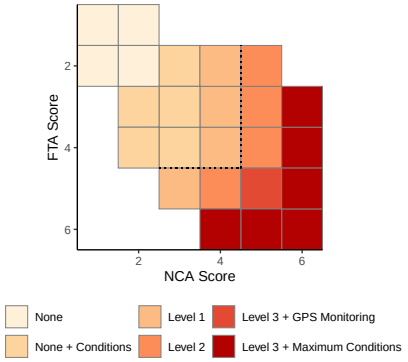


Figure G.10: Decision Making Framework (DMF) matrix recommendation for (a) the cash bail decision, and (b) additional release and monitoring conditions, for cases where the current charge is not a serious violent offense, the NVCA flag is not triggered, and the defendant was not extradited. If the FTA 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 cash bail. The dashed line indicates this boundary. Unshaded areas indicate impossible combinations of FTA and NCA scores. In (b) "Levels" 1,2 and 3 correspond to pre-defined levels of pretrial supervision, "None + Conditions" denotes minor conditions the signature bond if appropriate, "Level 3 + Maximum Conditions" corresponds to the highest level of pretrial supervision along with additional measures such as biweekly face-to-face and phone contact with arrestee.

above. We regress the difference in IP-weighted outcomes $\Gamma(1, \boldsymbol{X}, Y) - \Gamma(0, \boldsymbol{X}, Y)$ on all observed interactions between the FTA and NCA scores separately for the signature bond and cash bail groups. We then again use a heteroskedastic robust Wald test to test whether there is evidence for the coefficients for the interaction terms being non-zero, for each of the signature bond and cash bail groups. We find some weak evidence for interaction terms in the signature bond region ($p = 0.07$), but not in the cash bail region ($p = 0.13$).

Overall DMF risk score. Now we turn to the overall DMF 1–7 risk score that encodes recommendations on both the level of cash bail and the level and type of pre-trial supervision and monitoring conditions. Recall from Section 5.4 that due to the structure of the DMF matrix, it is not possible to identify the CATE for most risk levels at most combinations of FTA and NCA scores. Because we have $K = 7$ possible actions, we again usethe linear utility specification used for the FTA and NCA scores above, though other costs are also possible. For the DMF matrix, we again use the NVCA as the outcome.

Figure G.11 shows the resulting maximin DMF risk score recommendations for different costs of an NVCA and confidence levels. We find that it is not possible to safely change the DMF matrix for the full recommendation if the cost of an NVCA is larger than 5, even without requiring any degree of statistical certainty.

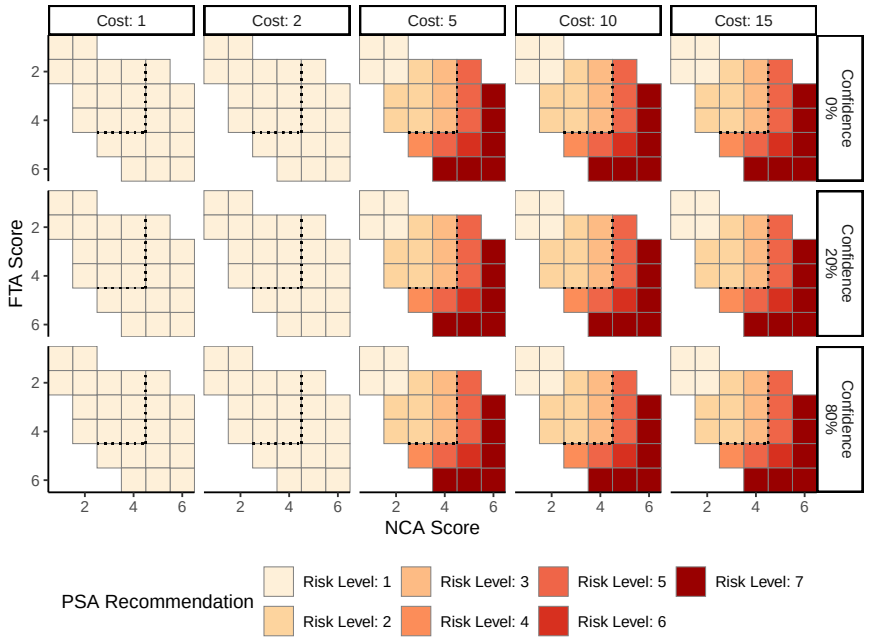


Figure G.11: Maximin monotone risk level cash bail and pre-trial supervision recommendations under an additive model for the treatment effects, as the cost of an NVCA and the confidence level vary. The dashed black line indicates the original decision boundary between a signature bond (above and to the left) and cash bail (below and to the right).

Quaternary cash bail recommendation. We also consider the quaternary cash bail recommendation between a signature bond, modest cash bail, moderate cash bail, and (full) cash bail. Here we have $K = 4$ actions and use the linear utility function. Figure G.12 shows the resulting maximin quaternary cash bail recommendation. This is broadly similar to what we find for the overall DMF risk score.

Ternary cash bail recommendation. We also consider the ternary cash bail recommendation between a signature bond, moderate/modest cash bail, and full cash bail, collapsing the moderate and modest cash bail recommendations. Here we have $K = 3$ actions and use the linear utility function. Figure G.13 shows the resulting maximin ternary cash bail recommendation. This is broadly similar to what we find for the binary cash bail recommendation. When the confidence level is set to zero and the cost of an NVCA is high enough, the maximin policy will extend the region where moderate cash bail is assigned to include the intermediate region between a signature bond and moderate cash bail. However, if any degree of statistical confidence is required, the maximin policy reverts to the status quo. Note that the maximin policy does not change the boundary between modest cash bail and cash bail, only between a signature bond and modest cash bail.

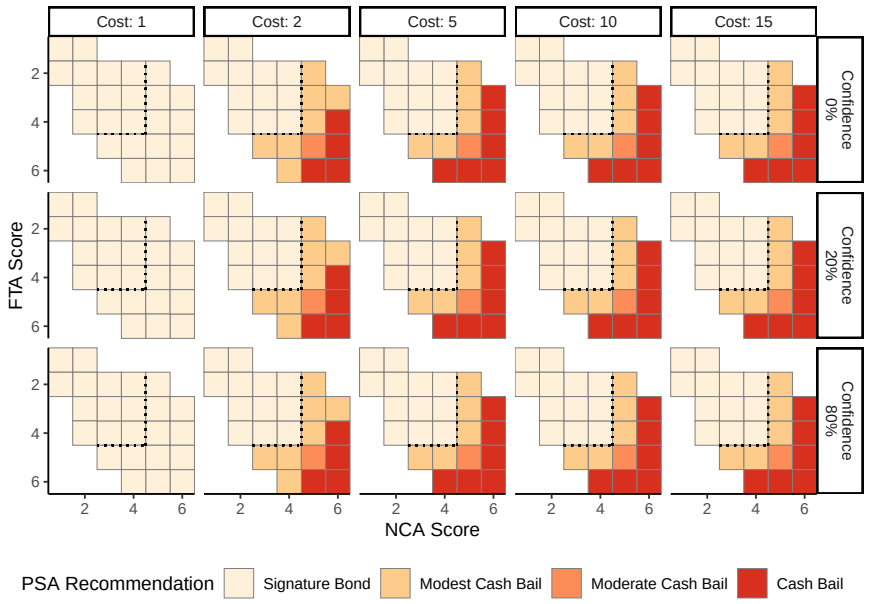


Figure G.12: Maximin monotone risk level ternary cash bail recommendations under an additive model for the treatment effects, as the cost of an NVCA and the confidence level vary. The dashed black line indicates the original decision boundary between a signature bond (above and to the left) and cash bail (below and to the right).

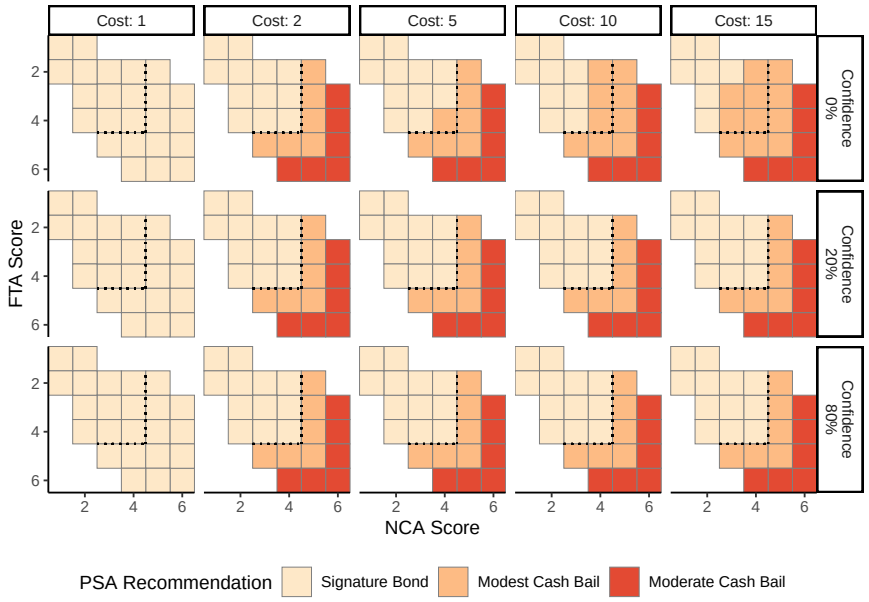


Figure G.13: Maximin monotone risk level ternary cash bail recommendations under an additive model for the treatment effects, as the cost of an NVCA and the confidence level vary. The dashed black line indicates the original decision boundary between a signature bond (above and to the left) and cash bail (below and to the right).

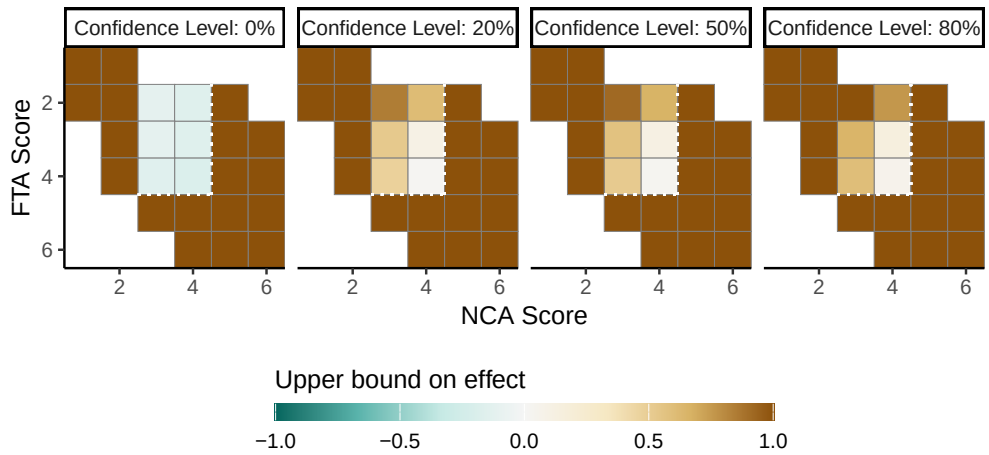


Figure G.14: Upper bound on the treatment effects under the additive model $\tau_{\text{add}}(a, x)$ for FTA and NCA scores. Values below and to the right of the dashed white line are areas where cash bail is recommended, and the bounds are on the effect of recommending a signature bond. Values above and to the left are areas where a signature bond is recommended, and the bounds are on the effect of recommending cash bail.

References

Ben-Michael, E., D. J. Greiner, M. Huang, K. Imai, Z. Jiang, and S. Shin (2024). Does AI help humans make better decisions? A statistical evaluation framework for experimental and observational studies. arXiv:2403.12108.

Chowdhury, S. R. and A. Gopalan (2017). On kernelized multi-armed bandits. *34th International Conference on Machine Learning, ICML 2017 2*, 1397–1422.

Fiedler, C., C. W. Scherer, and S. Trimpe (2021). Practical and Rigorous Uncertainty Bounds for Gaussian Process Regression. In *Association for the Advancement of Artificial Intelligence*.

Imai, K., Z. Jiang, D. J. Greiner, R. Halen, and S. Shin (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 (Statistics in Society) 186*(2), 167–189.

Kennedy, E. H. (2022). Towards optimal doubly robust estimation of heterogeneous causal effects.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces*. Berlin, Heidelberg: Springer.

Rahimi, A. and B. Recht (2008). Random Features for Large-Scale Kernel Machines. In *Advances in Neural Information Processing Systems*, Volume 20.

Srinivas, N., A. Krause, S. M. Kakade, and M. Seeger (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, 1015–1022.

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