On the Lower Confidence Band for the Optimal Welfare

Kirill Ponomarev* Vira Semenova†

This version: September 9, 2024

Abstract

This article addresses the question of reporting a lower confidence band (LCB) for the optimal welfare in a policy learning problem. A straightforward procedure inverts a one-sided t-test based on an efficient estimator of the optimal welfare. We describe a class of DGPs for which this procedure is dominated. Specifically, we show that an LCB corresponding to suboptimal welfare exceeds the straightforward LCB, with the average difference of order $N^{-1/2}$. Revisiting the National JTPA study (Kitagawa and Tetenov, 2018b), we find that an LCB based on a simple "treat everyone" policy exceeds the straightforward LCB by 200-300 USD worth of earnings (30-40%).

Keywords: policy learning, optimal welfare, lower confidence band, margin assumption, uniformity

*Department of Economics, University of Chicago. Email: kponomarev@uchicago.edu

†Department of Economics, University of California, Berkeley. Email: vsemenova@berkeley.edu.
We are grateful to Tim Armstrong, Alberto Abadie, Jiafeng (Kevin) Chen, Bryan Graham, Michael Jansson, Patrick Kline, Elena Manresa, Demian Pouzo, Andres Santos, Liyang (Sophie) Sun, Max Tabord-Meehan, Tiemen Woutersen, and participants of 2024 California Econometrics Conference for valuable comments. Jinglin Yang provided superb research assistance. All errors are ours.

1 Introduction

The problem of personalized treatment assignment has received a lot of attention in recent years (Manski, 2004; Dehejia, 2005; Hirano and Porter, 2009; Stoye, 2009; Chamberlain, 2011; Bhattacharya and Dupas, 2012; Tetenov, 2012; Kitagawa and Tetenov, 2018b; Athey and Wager, 2021; Mbakop and Tabord-Meehan, 2021; Sun, 2021). In addition to the optimal treatment policy, it is customary to report a point estimate of the optimal welfare and a lower confidence band for it (e.g., Johnson et al., 2023). As discussed in Hirano and Porter (2012) and Luedtke and van der Laan (2016), the inference problem is complicated by the potential non-uniqueness of the optimal policy and noisy nonparametric estimators of the unknown regression functions.

In this note, we show that focusing on a simpler policy — such as "treat everyone" or "treat no one" — sometimes results in better estimates and tighter confidence bands for optimal welfare, with the difference of $N^{-1/2}$ order of magnitude. We provide a class of DGPs for which the efficient estimator based on a suboptimal policy dominates the efficient estimator based on the optimal one according to two criteria. First, it has a lower mean squared error for all sample sizes large enough. Second, the corresponding LCB is tighter, on average, than the LCB based on the optimal policy in a large sample, with the difference being of $N^{-1/2}$ order of magnitude. We relate this result to the lack of uniformity in the margin assumption, commonly imposed in policy learning (e.g., Kitagawa and Tetenov (2018b)) and debiased inference (e.g., Luedtke and van der Laan (2016)). We find these considerations relevant in the context of the National JTPA study.

The rest of the note is organized as follows. Section 2 introduces the policy learning problem and sketches the estimators and lower confidence bands. Section 3 states theoretical results. Section 4 contains an empirical application. Section 5 concludes. Appendix A contains proofs. Appendix B contains auxiliary theoretical results. Appendix C contains auxiliary empirical details.

2 Setup

2.1 Assumptions

Let $D \in \{1,0\}$ be a binary indicator of treatment. The subject's potential outcome when treated is $Y(1)$ and when not treated is $Y(0)$. The realized outcome Y is

$$Y = DY(1) + (1 - D)Y(0).$$

The data vector (D,X,Y) consists of the treatment status D, a baseline covariate vector X taking values in $\mathcal{X} \subset \mathrm{R}^{\dim X}$, and the outcome $Y \in \mathrm{R}$. The researcher observes an i.i.d. sample $(D_i,X_i,Y_i)_{i=1}^N$ drawn from a distribution P satisfying the restrictions imposed below.

The social planner's goal is to maximize average welfare by deciding who should be treated based on observed characteristics X. Let $G \subset \mathcal{X}$ be a subset of covariate space. A subject is treated if $X \in G$ and not treated if $X \in G^c$, where $G^c = \mathcal{X} \setminus G$ is the complement of G. The average welfare of a decision rule G is

$$W_G = \mathbb{E}[Y(1)1\{X \in G\} + Y(0)1\{X \in G^c\}]. \tag{1}$$

The object of interest is the first-best unconstrained welfare denoted as

$$W_{G^*} = \max_G W_G.$$

To identify it, we employ the standard unconfoundedness assumption.

Assumption 2.1 (Unconfoundedness)**.** *The potential outcomes $Y(1)$ and $Y(0)$ are independent of treatment D given X:*

$$Y(1), Y(0) \perp D \mid X.$$

If Assumption 2.1 holds, the conditional means of potential outcomes are identified as

$$m(d,x) = \mathbb{E}[Y(d) \mid X = x] = \mathbb{E}[Y \mid D = d, X = x], \quad d \in \{1, 0\} \tag{2}$$

and the conditional average treatment effect (CATE) as

$$\tau(x) = m(1,x) - m(0,x). \tag{3}$$

In turn, the first-best welfare can be identified as

$$W_{G^*} = \mathbb{E}[\max(m(1,X), m(0,X))]. \tag{4}$$

This welfare is attained by the first-best policy

$$G^* = \{X : \tau(X) \geq 0\} \tag{5}$$

of treating those and only those with non-negative values of CATE. Denote the propensity score as

$$\pi(X) = P(D = 1 \mid X).$$

We impose the following regularity conditions on the distribution P of the data.

Assumption 2.2 (Regularity conditions). *For some finite positive constants κ,M the following statements hold. (1). For almost all $x \in \mathcal{X}$, the propensity score $\pi(x)$ obeys $\kappa < \pi(x) < 1 - \kappa$. (2) The outcome is supported on $[-M/2, M/2]$, that is $P(|Y| \leq M/2) = 1$. (3) The first-best policy is unique*

$$P(\tau(X) = 0) = 0. \tag{6}$$

Assumptions 2.2 (1)-(2) are standard in the literature. Assumption 2.2 (3) facilitates the use of regular estimators. If it holds, the semiparametric efficiency bound for W_{G^*} is well-defined (Luedtke and van der Laan (2016)). If it fails, regular estimators of the optimal welfare do not exist (Hirano and Porter (2012)).

2.2 Estimators and Lower Confidence Bands

To construct estimators and lower confidence bands for the first-best welfare, W_{G^*}, in (4), we focus on two policies: the first-best policy, $G = G^*$, which is unknown, and the "treat everyone" policy, $G = \mathcal{X}$.

Consider the first-best policy $G = G^*$ in (5). The semiparametric efficiency bound for W_{G^*} is characterized in Luedtke and van der Laan (2016). The bound is the same as if the first-best policy were known[1] and can be attained by an (infeasible) oracle estimator of the form

$$\widehat{W}_{G^*} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\widehat{\mathbb{E}}[D_iY_i \mid X_i]1\{X_i \in G^*\}}{\widehat{\mathbb{E}}[D_i \mid X_i]} + \frac{\widehat{\mathbb{E}}[(1-D_i)Y_i \mid X_i]1\{X_i \notin G^*\}}{1-\widehat{\mathbb{E}}[D_i \mid X_i]}\right]. \tag{7}$$

Consider the "treat everyone" policy. Its welfare,

$$W_{\mathcal{X}} = \mathbb{E}[Y(1)] = \mathbb{E}[\mathbb{E}[Y \mid D=1, X]],$$

provides a lower bound on W_{G^*}. Hahn (1998) characterizes the efficiency bound for $W_{\mathcal{X}}$ and shows it can be attained by an estimator of the form:

$$\widehat{W}_{\mathcal{X}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{\mathbb{E}}[D_iY_i \mid X_i]}{\widehat{\mathbb{E}}[D_i \mid X_i]}. \tag{8}$$

Define the mean squared error of each estimator

$$MSE(\widehat{W}_{\mathcal{X}}) = \mathbb{E}[\widehat{W}_{\mathcal{X}} - W_{G^*}]^2, \quad MSE(\widehat{W}_{G^*}) = \mathbb{E}[\widehat{W}_{G^*} - W_{G^*}]^2. \tag{9}$$

Since $\widehat{W}_{G^*}$ is an efficient estimator of W_{G^*}, one may expect $MSE(\widehat{W}_{G^*})$ to be smaller than $MSE(\widehat{W}_{\mathcal{X}})$.

Given the two estimators $\widehat{W}_{\mathcal{X}}$ and $\widehat{W}_{G^*}$, let us define the corresponding LCBs. In a large sample, the estimators are approximately distributed as

$$\sqrt{N}(\widehat{W}_{\mathcal{X}} - W_{\mathcal{X}}) \Rightarrow^d N(0, \sigma^2_{\mathcal{X}}) \tag{10}$$

$$\sqrt{N}(\widehat{W}_{G^*} - W_{G^*}) \Rightarrow^d N(0, \sigma^2_{G^*}). \tag{11}$$

A $100(1-\alpha)\%$ LCB for each parameter can be formed as

$$LCB_{\mathcal{X}} = \widehat{W}_{\mathcal{X}} - N^{-1/2}z_{1-\alpha}\widehat{\sigma}_{\mathcal{X}}, \tag{12}$$

$$LCB_{G^*} = \widehat{W}_{G^*} - N^{-1/2}z_{1-\alpha}\widehat{\sigma}_{G^*}, \tag{13}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of $N(0,1)$ and $\widehat{\sigma}_{G^*}, \widehat{\sigma}_{\mathcal{X}}$ are consistent estimators

[1]This property is established in the proof of Luedtke and van der Laan (2016), Theorem 1 (cf. Lemma A.1 in Appendix A).

of σ_{G^*} and $\sigma_{\mathcal{X}}$. Since $W_{\mathcal{X}}$ is a lower bound on the first-best welfare, $LCB_{\mathcal{X}}$ is also a valid LCB for W_{G^*}

$$P(LCB_{\mathcal{X}} \leq W_{G^*}) \geq P(LCB_{\mathcal{X}} \leq W_{\mathcal{X}}) \to 1-\alpha, \text{ as } \quad N \to \infty. \tag{14}$$

Since $W_{\mathcal{X}} \leq W_{G^*}$, one may expect $LCB_{\mathcal{X}}$ to be conservative relative to LCB_{G^*}. We compare $LCB_{\mathcal{X}}$ and LCB_{G^*} by taking expectation of their difference, $\mathbb{E}[LCB_{\mathcal{X}} - LCB_{G^*}]$.

3 Theoretical Results

Our first main result shows that $\widehat{W}_{G^*}$ can be dominated by $\widehat{W}_{\mathcal{X}}$ in terms of MSE in a finite sample. Likewise, LCB_{G^*} can be dominated by $LCB_{\mathcal{X}}$ in terms of expected length, with the difference of order $N^{-1/2}$ in magnitude.

Theorem 1. *There exists a class of DGPs obeying Assumptions 2.1 and 2.2 such that for all N large enough*
(1) Both MSEs in (9) are finite and

$$MSE(\widehat{W}_{\mathcal{X}}) < MSE(\widehat{W}_{G^*}); \tag{15}$$

(2) For any significance level $\alpha \in (0,1)$, there is a constant $C > 0$ such that

$$\mathbb{E}[LCB_{\mathcal{X}} - LCB_{G^*}] > CN^{-1/2}. \tag{16}$$

The proof of Theorem 1 can be found in Supplementary Appendix A. We construct a sequence of DGPs, indexed by a positive scalar ϵ, with a single binary covariate. Along this sequence, the conditional average treatment effect is

$$\tau(1) = -\epsilon, \quad \tau(0) = 1/2, \tag{17}$$

the welfare gap is

$$0 \leq W_{G^*} - W_{\mathcal{X}} \leq \epsilon, \tag{18}$$

and the variance gap is

$$\sigma^2_{G^*} - \sigma^2_{\mathcal{X}} > 5/4, \quad \forall \epsilon > 0. \tag{19}$$

As ϵ approaches zero, the welfare gap vanishes while the efficiency bounds $\sigma^2_{G^*}$ and $\sigma^2_{\mathcal{X}}$ remain separated. Setting $\epsilon = O(N^{-1/2})$ results in a sequence of distributions for which the dominance results (15) and (16) hold.

In the construction above, the first-best policy is unique for each $\epsilon > 0$ but becomes non-unique in the limit $\epsilon = 0$. At the limit distribution, the optimal welfare does not have a well-defined efficiency bound (Hirano and Porter (2012)). Yet, the optimal welfare coincides with the welfare $W_{\mathcal{X}}$ associated with a different policy $G = \mathcal{X}$, which can be regularly efficiently estimated. Theorem 1 demonstrates that, for distributions within a deleted $N^{-1/2}$-neighborhood[2] of $\epsilon = 0$, the efficient estimator $\widehat{W}_{\mathcal{X}}$ dominates an oracle, efficient estimator $\widehat{W}_{G^*}$.

Theorem 1 highlights the distinction between the two-sided and one-sided inferential objectives. In a two-sided case, the bias typically must vanish faster than the variance (through under-smoothing or debiasing) to ensure valid coverage. In a one-sided case, coverage remains valid as long as the direction of the bias matches the direction of the confidence band. Therefore, when the bias has a known sign, the bias-variance trade-off under the one-sided expected length is similar to the trade-off under the mean squared error.

3.1 The role of the margin assumption

In this Section, we connect the result of Theorem 1 to the lack of uniformity in the margin assumption, introduced below.

Assumption 3.1 (Margin assumption). *For some absolute finite constants $\eta \in (0,M)$ and $\delta \in (0,\infty)$,*

$$P(|\tau(X)| < t) \leq (t/\eta)^{\delta}, \quad \forall t \in [0, \eta). \tag{20}$$

Assumption 3.1 requires the first-best policy to be unique. In addition, it controls the intensity with which $\tau(X)$ concentrates in a neighborhood of $\tau(X) = 0$. For

[2] A deleted neighborhood is a neighborhood surrounding a specific point, in this case $\epsilon = 0$, minus the point itself.

a fixed P, the existence of suitable values of δ and η is typically guaranteed if the optimal policy is unique. For example, if $|\tau(X)|$ is continuous and has a density bounded at zero, then (20) holds for any $\delta < 1$ with η small enough. Similarly, if $\tau(X)$ has finite support and $P(\tau(X) = 0) = 0$, then (20) holds for any $\delta > 0$ and a sufficiently small η.

Assumption 3.1 is a standard sufficient condition for achieving fast rates of convergence in classification analysis (e.g., Mammen and Tsybakov, 1999; Tsybakov, 2004), welfare maximization (e.g., Qian and Murphy, 2011; Kitagawa and Tetenov, 2018b; Mbakop and Tabord-Meehan, 2021), and debiased inference on the first-best welfare in (e.g., Luedtke and van der Laan, 2016; Kallus et al., 2020).

We introduce a class of lower confidence bands for the first-best welfare W_{G^*}. Given a policy $G \subseteq \mathcal{X}$, let W_G denote the welfare parameter in (1) and σ_G^2 denote its semiparametric efficiency bound. The corresponding $100(1-\alpha)\%$ LCB is

$$LCB_G = \widehat{W}_G - N^{-1/2} z_{1-\alpha} \widehat{\sigma}_G, \tag{21}$$

$$\widehat{W}_G = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\widehat{\mathbb{E}}[D_i Y_i \mid X_i] 1\{X_i \in G\}}{\widehat{\mathbb{E}}[D_i \mid X_i]} + \frac{\widehat{\mathbb{E}}[(1-D_i) Y_i \mid X_i] 1\{X_i \notin G\}}{1 - \widehat{\mathbb{E}}[D_i \mid X_i]} \right], \tag{22}$$

where $\widehat{\sigma}_G$ is a consistent estimator of σ_G. Since W_G is a lower bound on W_{G^*}, the validity of LCB_G follows from an argument similar to (14).

To abstract away from the biases of estimators and standard deviations, we compare the LCBs based on their approximate expected lengths rather than their exact lengths. Define the population analogs of LCB_G and LCB_{G^*} as

$$\begin{aligned} ELCB_G &= W_G - N^{-1/2} z_{1-\alpha} \sigma_G \\ ELCB_{G^*} &= W_{G^*} - N^{-1/2} z_{1-\alpha} \sigma_{G^*}. \end{aligned}$$

Our second main result bounds the magnitude of $ELCB_G - ELCB_{G^*}$ uniformly over the class of DGPs obeying the margin assumption and an unrestricted policy class $\mathcal{G}$ consisting of all treatment policies $G \subseteq \mathcal{X}$.

Theorem 2. *Let $\mathcal{P}(M, \kappa, \eta, \delta, \underline{\sigma})$ denote the class of DGPs obeying Assumptions 2.1–3.1 and satisfying $\inf_{x \in \mathcal{X}} \inf_{d \in \{1,0\}} \sigma^2(d,x) \geq \underline{\sigma}^2 > 0$. The following statements hold.*

(1) The approximate expected length is bounded by

$$\sup_{P \in \mathcal{P}(M,\kappa,\eta,\delta,\underline{\sigma})} \sup_{G \in \mathcal{G}} \quad ELCB_G - ELCB_{G^*} \leq \bar{C} N^{-1/2(1+\delta)} \tag{23}$$

for some absolute constant $\bar{C} > 0$.

(2) For $\delta \in (0,1]$, the upper bound is tight in the sense that:

$$ELCB_{\mathcal{X}} - ELCB_{G^*} > \underline{C} N^{-1/2(1+\delta)} \tag{24}$$

for some DGP in $\mathcal{P}(M,\kappa,\eta,\delta,\underline{\sigma})$ and $0 < \underline{C} < \bar{C}$.

Theorem 2 suggests that, under the margin assumption, only higher-order improvements over LCB_{G^*} are possible, with the convergence rate determined by δ. The smaller the value of δ, the more individuals can concentrate near the boundary, and the looser the upper bound (23). The lower bound (24) recovers the negative result (16) of Theorem 1 in the limit value $\delta = 0$, which corresponds to the failure of the margin assumption. For small values of δ, the higher-order improvements may still be empirically relevant in finite samples.

To provide some intuition for the upper bound (23), let $P(G^* \Delta G)$ denote the share of people treated differently under the optimal policy G^* and an alternative G. This share links welfare and standard deviation gaps. Specifically, the welfare gap is lower bounded as

$$W_{G^*} - W_G \geq C_B P(G^* \Delta G)^{1+1/\delta}$$

for some $C_B > 0$ (Tsybakov (2004)). Additionally, we prove that the standard deviation gap is upper bounded as

$$\sigma_{G^*} - \sigma_G \leq (3M/2\kappa\underline{\sigma}) P(G^* \Delta G).$$

Thus, the margin assumption excludes DGPs of the form (18)–(19) where only the welfare gap vanishes but variance gap remains positive. Theorem 2 shows that placing a lower bound on the margin parameter δ is just enough to "restore" the optimality of the conventional LCB_{G^*}.

3.2 Implications and Open Questions

We conclude the discussion with the following Remarks.

Remark 1 sketches a test of the uniqueness of optimal policy.

Remark 1. Let X be a single discrete covariate taking J distinct values. The conditional average treatment effect reduces to a vector $(\tau(j))_{j=1}^{J}$. The first-best policy is non-unique if and only if there exists a cell $j \in \{1,2,\ldots,J\}$ with zero treatment effect. The null hypothesis

$$H_0 : \exists j \in \{1,2,\ldots,J\} : \tau(j) = 0 \tag{25}$$

is a union of J scalar hypotheses $H_{0j} : \tau(j) = 0$. As discussed in e.g. Berger (1997), a rejection region for H_0 can be taken as an intersection of scalar rejection regions

$$\mathrm{R} = \cap_{j=1}^{J} \mathrm{R}_j,$$

where R_j is a rejection region for j'th scalar null hypothesis H_{0j}. Further testing of Assumption 3.1 can be conducted using methods from the shape restrictions literature, e.g. Chernozhukov, Lee, and Rosen (2013).

Remark 2 describes another testable implication of margin assumption that we find empirically relevant.

Remark 2. The welfare gap for the policy $G = \mathcal{X}$ can be bounded as

$$C_B(P(\tau(X) < 0))^{1+1/\delta} \leq W_{G^*} - W_{\mathcal{X}} \leq MP(\tau(X) < 0) \tag{26}$$

for some constant $C_B > 0$ (Tsybakov, 2004). In particular, the welfare gap vanishes if and only if the share of people not to be treated also vanishes. For example, the sample version of this share vanishes in Rows 2 and 3 in Table 1, which is consistent with the sample version of (26), but it does not vanish in Row 4. Formalizing the testable implications of (26) is left for future work.

Remark 3 discusses the non-normality of estimators based on a data-driven lower bound.

Remark 3. Inference on the first-best welfare robust to violations of margin condition remains an important open question. An example of a robust alternative is the estimator of $W_{\mathcal{X}}$, proposed by Hahn (1998). Another robust alternative is the estimator $\widehat{W}_G$ defined in (22) where the policy G is data-driven. However, this alternative faces a substantial efficiency cost since the independence of samples used to estimate G and the welfare W_G can only be attained via sample splitting. Cross-fit estimates of $\widehat{W}_G$, based on different data-driven policies, exhibit a non-standard, possibly heavy-tailed, distribution. As a result, Wald-type confidence intervals (CIs) based on cross-fit estimates may no longer be valid. A similar failure of normality has also been pointed out in the context of batched bandits (Zhang et al., 2021).

In light of this discussion, the practitioners may prefer simple, robust alternatives to W_{G^*} as long as these alternatives are not overly conservative. An example of a lower bound is $\max(W_{\mathcal{X}}, W_{\emptyset})$. Appendix B shows that the resulting welfare gap is upper bounded by the variance of CATE

$$0 \leq W_{G^*} - \max(W_{\mathcal{X}}, W_{\emptyset}) \leq \sqrt{\mathbb{V}ar(\tau(X))}. \tag{27}$$

If $\mathbb{V}ar(\tau(X))$ decays faster than N^{-1}, the welfare gap vanishes faster than the parametric rate. As a result, a lower confidence band targeting $\max(W_{\mathcal{X}}, W_{\emptyset})$ may be within $N^{-1/2}$ from the conventional one LCB_{G^*} according to expected length. Detecting such regimes would be possible using the estimator of $\mathbb{V}ar(\tau(X))$ proposed in Levy et al. (2021).

4 Empirical application

In this Section, we revisit the National Job Training Partnership Act (JTPA) study, considered in Heckman et al. (1997) and Abadie et al. (2002) and recently revisited in the context of policy learning by Kitagawa and Tetenov (2018b), Mbakop and Tabord-Meehan (2021), and Athey and Wager (2021), among others. A detailed description of the study is available in Bloom et al. (1997). The study randomized whether applicants would be eligible to receive job training and related services for a period of eighteen months. The treatment D is the indicator of program eligibility. The outcome Y is the applicant's cumulative earnings thirty months after assignment. Two baseline covariates $X = (PreEarn, Educ)$ include pre-program earnings (in USD) and years of education. The propensity score is constant and equals two-thirds.

We are interested in estimating and constructing a LCB for the first-best welfare gain relative to the "treat no one" policy:

$$W_{gain} = W_{G^*} - W_{\emptyset}. \tag{28}$$

Any policy $G \subseteq \mathcal{X}$ provides a lower bound on welfare gain

$$W_{gain} \geq W_G - W_{\emptyset}. \tag{29}$$

In particular, $G = \mathcal{X}$ corresponds to the average treatment effect

$$W_{gain} \geq \underbrace{W_{\mathcal{X}} - W_{\emptyset}}_{ATE}. \tag{30}$$

We focus on ATE because it is well-studied and common to report.

First, we consider a model $(Educ, D, Y)$ with a single discrete covariate $Educ$. We find that eleven out of twelve education groups do not have a significant CATE at $\alpha = 0.05$, so the null hypothesis of non-unique optimal policy in (25) cannot be rejected using the test described in Remark 1. If the optimal policy is non-unique, we expect the conventional LCB for the first-best welfare, based on Gaussian approximation, to be invalid; thus, we do not report it.

Next, we consider a model $(PreEarn, D, Y)$ with a continuous covariate $PreEarn$. We estimate the welfare gain in (28) by cross-fitting the classic doubly-robust score using two different CATE estimators: series regression and random forest. The conventional LCBs, relying on the margin assumption, are plausibly valid since $PreEarn$ is continuously distributed. We estimate the ATE in (30) using a regression adjustment estimator of the form (8). To avoid relying on nonparametric estimators, we bin $PreEarn$ into five cells of similar size and use cell-specific averages.

Finally, we consider a model (X, D, Y) with both covariates, $X = (Educ, PreEarn)$. If the optimal policy is non-unique, the standard Wald CI based on a cross-fit plug-in estimator may not be valid (see Remark 3). To construct a valid data-driven LCB, we randomly split the sample into two parts and use the first part (one-third of the sample) to choose a policy $\hat{G}$ and the second one to estimate $LCB_{\hat{G}}$. We estimate the optimal policy $\hat{G}$ using random forest. Appendix C provides further details on how the estimators are constructed.

Table 1 summarizes our results. First, comparing Row 1 with Rows 2–3, we find that due to the sampling noise, the estimated ATE exceeds the welfare gain by 220 – 246 USD, in contrast with the population inequality (30). Likewise, the first-best LCB is 28 – 48% wider than the LCB for ATE.[3] At the same time, the estimated share of people not recommended for treatment ranges between 1% and 8%. This pattern is consistent with the sample version of the testable implication (26) in Remark 2, where the welfare gap, the variance gap, and the share of individuals not recommended for treatment vanish. In Row 4, the welfare gap is -293 USD, yet the estimated share is 23%, which could be interpreted as a violation of the margin assumption. Across the board, targeting the ATE results in a more precise estimate and a tighter expected confidence band for the first-best welfare.

5 Conclusion

We pose the question of reporting a lower confidence band for the optimal welfare. When the welfare gain, relative to a simpler class of policies, is small compared to sampling uncertainty, a simpler policy delivers better point estimates and tighter confidence bands in certain cases. In particular, if only a small share of population should be treated differently from the majority under the optimal policy, it could be optimal to report welfare of a simpler policy that treats everyone the same.

[3]The series regression with linear, quadratic, and cubic transformations of the $PreEarn$ covariate results in a "treat everyone" assignment rule. Their point estimates and LCBs are close to those for ATE (see Table C.1). The fourth power detects heterogeneity in treatment assignment, where the percentage of people recommended for treatment is 99.2%(Table 1, Row 2). Yet, the respective LCB is 40% wider. The LCBs based on higher powers of series regression are negative and omitted.

Table 1: Welfare Gain Per Capita: Estimates and Lower Confidence Bands

	(1)	(2)	(3)	(4)	(5)
Treatment Rule	Share of People to be Treated	Welfare Gain (St.Error)	95% LCB	Welfare Gap (USD)	LCB Gap (%)
Treat Everyone ($PreEarn$)	1.00	1289.66 (347.82)	717.52		
Series Regression $\sum_{j=1}^{4}(PreEarn)^j$	0.992	1043.22 (394.67)	394.03	-246.44	45%
Random Forest ($PreEarn$)	0.92	1069.50 (335.24)	518.06	-220.16	28%
Random Forest ($PreEarn + Educ$)	0.77	996.43 (393.99)	348.31	-293.23	51%

Notes. The outcome variable is 30-Month Post-Program Cumulative Earnings in USD. Welfare gain is the first-best welfare net of the control average outcome, as defined in (28). Row (1): average treatment effect as in (30). Rows (2)–(3): welfare gain W_{gain} based on plug-in rule estimated via series regression (Row (2)) and random forest (Row(3)). Row (4): sample-split welfare gain W_{gain} estimate based on a plug-in rule estimated via random forest of both covariates. Column (3): 95% LCB defined as $W_{gain} - 1.645 * St.Error(W_{gain})$. Column (4): welfare gap $W_{gain} - ATE$, where $ATE = 1289.66$ (Row 1) and W_{gain} is in subsequent rows. Column (5): relative LCB gap $100(1 - LCB_j/LCB_{ATE})\%$ where $LCB_{ATE} = 717.52$ (Row 1) and LCB_j is in subsequent rows. The sample ($N =$ 9,223) is the same as in Kitagawa and Tetenov (2018b). See text for details.

A Proof of Theorem 1

Section A.1 contains auxiliary statements. For expositional reasons, the proof of Theorem 1(2) is given in Section A.2 and the proof of Theorem 1(1) is given in Section A.3. Section A.4 contains the proof of Theorem 2.

A.1 Auxiliary statements

Lemma A.1 (Luedtke and van der Laan (2016), Theorem 1)**.** *Suppose Assumptions 2.1 and 2.2 hold. Then, the efficient score for the first-best welfare* $\mathbb{E}[\max(m(1,X), m(0,X))]$ *is*

$$\psi(W) = \left(m(1,X) + \frac{D}{\pi(X)}(Y - m(1,X))\right)1\{\tau(X) > 0\} \\ + \left(m(0,X) + \frac{1-D}{1-\pi(X)}(Y - m(0,X))\right)1\{\tau(X) < 0\}. \tag{A.1}$$

Lemma A.2 (Efficiency bound for W_G)**.** *Suppose Assumption 2.1 and 2.2 (1)-(2) hold. Then, the efficient score for the welfare* W_G *of a known policy* G *is*

$$\psi_G(W) = \left(m(1,X) + \frac{D}{\pi(X)}(Y - m(1,X))\right)1\{X \in G\} \\ + \left(m(0,X) + \frac{1-D}{1-\pi(X)}(Y - m(0,X))\right)1\{X \in G^c\}. \tag{A.2}$$

Proof. The parameter $W_G = \mathbb{E}[Y(1)1\{X \in G\} + Y(0)1\{X \in G^c\}]$ is a sum of two potential outcomes weighted by known functions of X, namely, $1\{X \in G\}$ and $1\{X \in G^c\}$. The efficiency bound follows from an argument of Hahn (1998), Theorem 1. ■

A.2 Proof of Theorem 1(2).

Section A.A.1 sketches the DGP that constitutes the proof of Theorem 1. Section A.A.2 describes the key steps of the proof.

A.A.1 The Data Generating Process

The DGP consists of a marginal distribution of X, the distribution of D given X, and the distribution of $(Y(0), Y(1))$ given X. We take X as a single binary covariate

$$P(X = 1) = p, \quad P(X = 0) = 1 - p, \quad p \in (0,1). \tag{A.3}$$

The propensity score reduces to

$$P(D = 1 \mid X = 1) = \pi(1), \quad P(D = 1 \mid X = 0) = \pi(0).$$

Let $F(\mu, \sigma^2)$ be any distribution supported on $[-M/2, M/2]$ with mean μ and variance σ^2. Suppose the potential outcomes are distributed as

$$Y(1) \mid X = 1 \sim F(1/2 - \epsilon, 1) \tag{A.4}$$

$$Y(1) \mid X = 0 \sim F(1/2, 1), \tag{A.5}$$

$$Y(0) \mid X = 1 \sim F(1/2, 10) \tag{A.6}$$

$$Y(0) \mid X = 0 \sim F(0, 10) \tag{A.7}$$

where $\epsilon > 0$ is a positive scalar to be specified. We focus on DGPs with $p \in (1/4,3/4)$ and $\pi(1), \pi(0) \in (1/4, 3/4)$ and $\epsilon \in (0,1)$. The conditional average treatment effect reduces to

$$\tau(1) = -\epsilon < 0, \quad \tau(0) = 1/2 > 0. \tag{A.8}$$

The first-best policy, which is unique, reduces to

$$G^* = \{X = 0\}. \tag{A.9}$$

The regression estimators in (7) and (8) are sample averages. For $d, x \in \{1, 0\}$, define sample counts as

$$N_{dx} = \sum_{i=1}^{N} 1\{D_i = d\}1\{X_i = x\}. \tag{A.10}$$

The numerator of regression estimators is

$$\widehat{\mathbb{E}}[D_i Y_i \mid X_i = 1] = \frac{\sum_{i=1}^N D_i X_i Y_i}{\sum_{i=1}^N X_i + 1}$$
$$\widehat{\mathbb{E}}[D_i \mid X_i = 1] = \frac{\sum_{i=1}^N D_i X_i + 1}{\sum_{i=1}^N X_i + 1} = \frac{N_{11} + 1}{\sum_{i=1}^N X_i + 1},$$

where one is added throughout to prevent division by zero[4]. For zero value of covariate X, similar expressions apply. For d,x define

$$\widehat{m}_{dx} = \frac{\sum_{i=1}^N 1\{D_i = d\} 1\{X_i = x\} Y_i}{\sum_{i=1}^N 1\{D_i = d\} 1\{X_i = x\} + 1} = \frac{\sum_{i=1}^N 1\{D_i = d\} 1\{X_i = x\} Y_i}{N_{dx} + 1}. \tag{A.11}$$

Then, the estimators $\widehat{W}_{G^*}$ and $\widehat{W}_{\mathcal{X}}$ can be expressed as

$$\widehat{W}_{G^*} = \widehat{m}_{01}\widehat{p} + \widehat{m}_{10}(1 - \widehat{p}) \tag{A.12}$$

$$\widehat{W}_{\mathcal{X}} = \widehat{m}_{11}\widehat{p} + \widehat{m}_{10}(1 - \widehat{p}), \quad \widehat{p} = \sum_{i=1}^N X_i / N. \tag{A.13}$$

A.A.2 Two Key Lemmas

Lemma A.3 proves that the welfare gap shrinks to zero while efficiency bounds are strictly separated.

Lemma A.3 (Positive Variance Gap)**.** *(1) The statement* (10) *holds with*

$$W_{\mathcal{X}} = 1/2 - \epsilon p, \quad \sigma^2_{\mathcal{X}} = \frac{p}{\pi(1)} + \frac{(1-p)}{\pi(0)} + \epsilon^2 (1-p)p. \tag{A.14}$$

(2) The statement (11) *holds with*

$$W_{G^*} = 1/2, \quad \sigma^2_{G^*} = \frac{1-p}{\pi(0)} + \frac{10p}{1 - \pi(1)}. \tag{A.15}$$

[4] Inflating the denominators of $\widehat{W}_{G^*}$ and $\widehat{W}_{\mathcal{X}}$ by one prevents the division by zero and ensures that MSEs in (9) are finite. This step introduces bias of order $O(N^{-1})$. Since the results of Theorem 1 are stated at $O(N^{-1/2})$ scale, the bias is negligible for a sufficiently large sample. An alternative option is to work with standard (non-adjusted) denominators on the event where both of them are strictly positive. Under Assumption 2.2 (2), the probability of this event approaches one exponentially fast as N becomes large.

(3) The variance gap is bounded from below

$$\sigma^2_{G^*} - \sigma^2_{\mathcal{X}} > 5p. \tag{A.16}$$

Proof of Lemma A.3. The proof has three steps. Steps 1 and 2 establish (A.14) and (A.15). Step 3 gives a lower bound on variance gap.

Step 1. The estimator (8) is efficient (Hahn (1998), Proposition 4, p. 322). The efficiency bound is

$$\sigma^2_{\mathcal{X}} = \mathbb{E}\left[(m(1,X) - W_{\mathcal{X}})^2 + \frac{\mathbb{V}ar(Y \mid D = 1, X)}{\pi(X)}\right]. \tag{A.17}$$

The first summand is

$$\mathbb{E}\left[(m(1,X) - W_{\mathcal{X}})^2\right] = \epsilon^2(1-p)^2 p + \epsilon^2 p^2 (1-p) = \epsilon^2 p(1-p).$$

The second summand is

$$\mathbb{E}\left[\frac{\mathbb{V}ar(Y \mid D = 1, X)}{\pi(X)}\right] = \frac{1}{\pi(1)}p + \frac{1}{\pi(0)}(1-p).$$

Adding two summands gives $\sigma^2_{\mathcal{X}}$ in (A.14).

Step 2. The first-best efficient score is given in (A.1). Its variance has two summands. The first summand is $\mathbb{V}ar(\max(m(1,X), m(0,X))) = \mathbb{V}ar(1/2) = 0$. The second one is

$$\sigma^2_{G^*} = \mathbb{E}\left[\frac{X\mathbb{V}ar(Y \mid D = 0, X = 1)}{1 - \pi(1)} + \frac{(1-X)\mathbb{V}ar(Y \mid D = 1, X = 0)}{\pi(0)}\right].$$

Step 3. Note that $11\pi(1) > 3/2$ and $\pi(1)(1 - \pi(1)) \leq 1/4$ for $\pi(1) \in (1/4, 3/4)$. Invoking $\epsilon^2 p < 1$ gives (A.16). ∎

Lemma A.4 establishes the lower bound (16), which completes the proof of Theorem 1(2).

Lemma A.4 (Ranking of LCBs)**.** *The LCB ranking (16) holds for any $\epsilon \in (0, z_{1-\alpha}/4N^{-1/2})$ with the constant $C = z_{1-\alpha}/8$.*

Proof of Lemma A.4. Step 1. Decompose

$$\mathbb{E}[LCB_{\mathcal{X}} - LCB_{G^*}] = \Delta_{\mathcal{X}} + (\mathbb{E}[\widehat{W}_{\mathcal{X}} - W_{\mathcal{X}}] - \mathbb{E}[\widehat{W}_{G^*} - W_{G^*}]) \\ + (N^{-1/2} z_{1-\alpha}(\mathbb{E}[\widehat{\sigma}_{G^*} - \sigma_{G^*}] - \mathbb{E}[\widehat{\sigma}_{\mathcal{X}} - \sigma_{\mathcal{X}}])),$$

where the leading term $\Delta_{\mathcal{X}}$ is

$$\Delta_{\mathcal{X}} = N^{-1/2}(\sigma_{G^*} - \sigma_{\mathcal{X}}) - (W_{G^*} - W_{\mathcal{X}}). \tag{A.18}$$

Step 2. Let the estimators $\widehat{\sigma}^2_{\mathcal{X}}$ and $\widehat{\sigma}^2_{G^*}$ be standard sample analogs of $\sigma^2_{\mathcal{X}}$ and $\sigma^2_{G^*}$, respectively. By Assumption 2.2 and the choice of $\widehat{W}_{G^*}, \widehat{W}_{\mathcal{X}}$, these estimators are bounded a.s. Thus, $\widehat{\sigma}_{\mathcal{X}}$ and $\widehat{\sigma}_{G^*}$ are also a.s. bounded and consistent. Then, convergence in probability implies convergence of moments:

$$\mathbb{E}[\widehat{\sigma}_{\mathcal{X}} - \sigma_{\mathcal{X}}] = o(1), \qquad \mathbb{E}[\widehat{\sigma}_{G^*} - \sigma_{G^*}] = o(1).$$

Invoking (A.31) gives $\mathbb{E}[\widehat{W}_{\mathcal{X}} - W_{\mathcal{X}}] - \mathbb{E}[\widehat{W}_{G^*} - W_{G^*}] = o(N^{-1/2}) < \Delta_{\mathcal{X}}/2$ for large enough N.

Step 3. The choice of $\pi(1), \pi(0) \in (1/4, 3/4)$ implies $\sigma^2_{G^*} \leq 44$. In addition, since $\epsilon^2 \leq 1$, $\sigma^2_{\mathcal{X}} \leq 4(p + (1-p)) + 1 = 5$. A lower bound (A.16) implies a lower bound on the standard deviation gap

$$\sigma_{G^*} - \sigma_{\mathcal{X}} = \frac{\sigma^2_{G^*} - \sigma^2_{\mathcal{X}}}{\sigma_{G^*} + \sigma_{\mathcal{X}}} \geq^{(i)} \frac{6p}{\sqrt{5} + \sqrt{44}} > p/2$$

where (i) follows from (A.16). A lower bound on $\Delta_{\mathcal{X}}$ follows

$$\Delta_{\mathcal{X}} > N^{-1/2}(z_{1-\alpha}/2 - \epsilon)p = N^{-1/2} z_{1-\alpha}/4p.$$

Take $\epsilon \in (0, z_{1-\alpha}/4N^{-1/2})$ gives $\Delta_{\mathcal{X}} > N^{-1/2} z_{1-\alpha} p/4$, which implies (16) with $C = z_{1-\alpha} p/8$.

∎

A.3 Proof of Theorem 1(1).

Let (X,D,Y) be a collection of DGPs described in Section A.2. We focus on symmetric DGPs with $p = \pi(1) = \pi(0) = 1/2$. Furthermore, the conditional variances differ from those specified in (A.5)–(A.7)

$$\sigma^2(1,1) = 1/4, \quad \sigma^2(1,0) = 1, \quad \sigma^2(0,1) = 200, \quad \sigma^2(0,0) = 1.$$

Plugging these parameter values into efficiency bounds gives

$$\sigma^2_{\mathcal{X}} = 5/4 + \epsilon^2/4 < 3/2, \quad \sigma^2_{G^*} = 201.$$

Let N_{dx} be sample counts defined in (A.10). The following standard properties of binomial distribution apply. For $N \geq 1$,

$$\mathbb{E}[(N_{dx}+1)^{-1}] = \frac{4}{N+1} - \frac{4}{N+1}(3/4)^{N+1} \leq 4/N \tag{A.19}$$

$$\mathbb{E}[(N_{dx}+1)^{-2}] \leq 32N^{-2}. \tag{A.20}$$

Lemma A.5 bounds the approximation error of expected conditional variance.

Lemma A.5. *For $N \geq 100$ and $C_N = \sqrt{2.25 \ln N/N}$, the following bounds hold for any $d,x \in \{1,0\}$*

$$\sigma^2(d,1)(1 - 10C_N) < \mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{d1} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2] < \sigma^2(d,1)(1 + 10C_N). \tag{A.21}$$

$$\sigma^2(d,0)(1 - 10C_N) < \mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{d0} \mid \mathbf{X}, \mathbf{D})(1-\widehat{p})^2] < \sigma^2(d,0)(1 + 10C_N). \tag{A.22}$$

Proof of Lemma A.5. We prove (A.21), which corresponds to $x = 1$. The symmetry of DGPs implies (A.22) with $x = 0$. Step 1 introduces error terms and the notation, which are bounded in Steps 2 and 3. Let $(\mathbf{X}, \mathbf{D}) = (X_i, D_i)_{i=1}^N$ be stacked realizations of $(X_i)_{i=1}^N$ and $(D_i)_{i=1}^N$.

Step 1. The conditional variance reduces to

$$\mathbb{V}ar(\widehat{m}_{d1} \mid \mathbf{X}, \mathbf{D}) = \sigma^2(d,1)N_{d1}(N_{d1}+1)^{-2}. \tag{A.23}$$

Multiplying both sides by $N\widehat{p}^2$ and taking probability limit gives

$$\text{plim}_{N\to\infty}\sigma^2(d,1)NN_{d1}(N_{d1}+1)^{-2}\widehat{p}^2 = \sigma^2(d,1)\frac{p^2}{\Pr(D = d \mid X = 1)p},$$

which reduces to $\sigma^2(d,1)$ since $p = \pi(1) = \pi(0) = 1/2$. We introduce the following notation

$$\psi_d^1(t) = Nt(N_{d1}+1)^{-1}, \quad \psi_d^2(t) = Nt(N_{d1}+1)^{-2}$$

and decompose error term

$$\begin{aligned}
&N\widehat{p}^2 N_{d1}(N_{d1}+1)^{-2} - 1 \\
&= \psi_d^1(\widehat{p}^2) - \psi_d^2(\widehat{p}^2) - 1 \\
&= \psi_d^1(p^2) + \psi_d^1(\widehat{p}^2 - p^2) - \psi_d^2(\widehat{p}^2) - 1 \\
&= (\psi_d^1(p^2) - N/(N+1)) + \psi_d^1(\widehat{p}^2 - p^2) - \psi_d^2(\widehat{p}^2) - 1/(N+1) \\
&= S_1 + S_2 - S_3 - S_4. \qquad \text{(A.24)}
\end{aligned}$$

Step 2. We bound the term $\mathbb{E}[S_2]$. On the event $\mathcal{M}_N := \{|\widehat{p} - 1/2| < C_N\}$, the error $|\widehat{p}^2 - 1/4| \leq 1.5C_N$ a.s. Then

$$\begin{aligned}
|\mathbb{E}[S_2 1\{\mathcal{M}_N\}]| \leq \mathbb{E}[\psi_d^1(|\widehat{p}^2 - p^2|)|1\{\mathcal{M}_N\}] &\leq \mathbb{E}[\psi_d^1(1.5C_N)1\{\mathcal{M}_N\}] \\
&\leq 1.5C_N \mathbb{E}[\psi^1(1)] \leq^{(i)} (1.5)\cdot 4 \cdot C_N = 6C_N,
\end{aligned}$$

where (i) follows from (A.19). On the event $\mathcal{M}_N^c$, the error $|\widehat{p}^2 - p^2| \leq 1$ a.s.. Then

$$|\mathbb{E}[S_2 1\{\mathcal{M}_N^c\}]| \leq \mathbb{E}[\psi_d^1(|\widehat{p}^2 - p^2|)1\{\mathcal{M}_N^c\}] \leq^{(i)} \mathbb{E}[\psi_d^1(1)1\{\mathcal{M}_N^c\}] \leq NP(\mathcal{M}_N^c)$$

where (i) follows from $N_{d1} \geq 0$ a.s. and $(N_{d1}+1)^{-1} \leq 1$a.s. Chernoff bound for Binomial distribution gives

$$NP(\mathcal{M}_N^c) \leq 2N \exp^{-2C_N^2 N/3} \leq 2N^{-1/2} \leq C_N,$$

where the last inequality holds for $N \geq 6$. Adding the two inequalities gives $|\mathbb{E}[S_2]| \leq 7C_N$.

Step 3. We bound the term $\mathbb{E}[S_1], \mathbb{E}[S_3], \mathbb{E}[S_4]$. Invoking (A.20) gives

$$\mathbb{E}[S_3] = \mathbb{E}[\psi_d^2(\widehat{p}^2)] \leq 32N^{-1} \leq C_N, \quad \forall N \geq 100.$$

A simple argument gives $(N+1)^{-1} \leq C_N$. Invoking (A.19) gives

$$\mathbb{E}[|S_1|] \leq C_N, \quad \forall N \geq 2.$$

Adding the terms gives the result. $\blacksquare$

Lemma A.5 establishes a lower bound for $MSE(\widehat{W}_{G^*})$.

Lemma A.6. *For $N \geq 100$ and $C_N = \sqrt{2.25 \ln N/N}$, $MSE(\widehat{W}_{G^*})$ is lower bounded as*

$$N \cdot MSE(\widehat{W}_{G^*}) > 201(1 - 10C_N). \tag{A.25}$$

Proof of Lemma A.6. Step 1. Let d_1 and d_2 be in $\{1{,}0\}$. Conditional on $(\mathbf{X}, \mathbf{D})$, the only stochastic component of $\widehat{W}_{G^*}$ is $(Y_i)_{i=1}^N$. Thus, the $\text{Cov}(\widehat{m}_{d_1 1}, \widehat{m}_{d_2 0} \mid \mathbf{X}, \mathbf{D})$ consists of N^2 ratios whose numerator is $1\{D_i = d_1\}\{D_j = d_2\}X_i(1 - X_j)\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D})$ and denominator is a function of $N_{d_1 1}$ and $N_{d_2 0}$, where $(i{,}j) \in \{1{,}2, \ldots, N\}^2$.

Step 2. We show that $X_i(1-X_j)\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D}) = 0$ a.s. for any $i{,}j$. For tuples $i \neq j$, $\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D}) = 0$ by independence of the samples i and j. For tuples where $i = j$, the product $X_i(1 - X_j) = X_i(1 - X_i) = 0$ a.s. Thus,

$$\text{Cov}(\widehat{m}_{d_1 1}, \widehat{m}_{d_2 0} \mid \mathbf{X}, \mathbf{D}) = 0,$$

and the variance of each estimator is

$$\mathbb{V}ar(\widehat{W}_{G^*} \mid \mathbf{X}, \mathbf{D}) = \mathbb{V}ar(\widehat{m}_{01} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 + \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2. \tag{A.26}$$

$$\mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}) = \mathbb{V}ar(\widehat{m}_{11} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 + \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2. \tag{A.27}$$

Step 3. We prove (A.25). Invoking Lemma A.5 with parameters $(d{,}x) = (0{,}1)$ and $(d{,}x) = (1{,}0)$ gives a lower bound

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{01} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2] > 200(1 - 10C_N) \tag{A.28}$$

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2] > 1 - 10C_N \tag{A.29}$$

Adding (A.28) and (A.29) gives a lower bound on $\mathbb{E}[\mathbb{V}ar(\widehat{W}_{G^*} \mid \mathbf{X}, \mathbf{D})]$. A lower bound on $MSE(\widehat{W}_{G^*})$ follows.

$\blacksquare$

Lemma A.7 establishes an upper bound for $MSE(\widehat{W}_{\mathcal{X}})$.

Lemma A.7. *For $N \geq 100$ and $C_N = \sqrt{2.25 \ln N / N}$ and $N\epsilon^2 \leq 1$, $MSE(\widehat{W}_{\mathcal{X}})$ is upper bounded by*

$$N \cdot MSE(\widehat{W}_{\mathcal{X}}) < \sigma^2_{\mathcal{X}} + 16C_N + N\epsilon^2/2 \leq 2 + 16C_N. \tag{A.30}$$

Proof of Lemma A.7. Step 1. We show that the bias is bounded from above and below

$$0 \leq W_{G^*} - \mathbb{E}[\widehat{W}_{\mathcal{X}}] \leq \epsilon/2 + 4N^{-1}. \tag{A.31}$$

For $\epsilon \leq 1/2$, the remainder term

$$R := (1/2 - \epsilon)\widehat{p}(N_{11} + 1)^{-1} + 1/2(1 - \widehat{p})(N_{10} + 1)^{-1}$$

is non-negative a.s. The conditional mean is

$$\begin{aligned}\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}] &= (1/2 - \epsilon\widehat{p}) - [(1/2 - \epsilon)\widehat{p}(N_{11} + 1)^{-1} + 1/2(1 - \widehat{p})(N_{10} + 1)^{-1}] \\ &= (1/2 - \epsilon\widehat{p}) - R.\end{aligned} \tag{A.32}$$

Integrating over $\mathbf{X}, \mathbf{D}$ gives

$$\mathbb{E}[\widehat{W}_{\mathcal{X}}] - W_{\mathcal{X}} = -\mathbb{E}[R]$$

where $\mathbb{E}[R] \leq 2/N + 2/N$ follows from $\widehat{p} \in [0,1]$ a.s. and (A.19) and monotonicity of expectation.

Step 2. We show that variance is upper bounded by

$$N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}}) < \sigma^2_{\mathcal{X}} + 16C_N. \tag{A.33}$$

The variance of the conditional mean is

$$\mathbb{V}ar(\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}]) = \mathbb{V}ar(R) - 2\mathrm{Cov}(R, 1/2 - \epsilon\widehat{p}) + \frac{\epsilon^2}{4N}. \tag{A.34}$$

Invoking (A.20) bounds the variance of the remainder

$$N\mathbb{V}ar(R) \leq 2/4\mathbb{E}[N(N_{11}+1)^{-2}]+2/4\mathbb{E}[N(N_{10}+1)^{-2}] \leq 32N^{-1} \leq C_N, \quad \forall N \geq 100.$$

Invoking Cauchy inequality bounds the covariance term

$$2N|\mathrm{Cov}(R, 1/2-\epsilon\widehat{p})| \leq 2\sqrt{32\epsilon^2/(4N)} \leq 4\sqrt{2}N^{-1} \leq C_N, \quad \forall N \geq 8.$$

Invoking (A.21) with $\sigma^2(1{,}1) = 1/4$ and $\sigma^2(1{,}0) = 1$ gives

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D})] \leq 5/4 + 25/2C_N. \tag{A.35}$$

Step 3. Adding (A.34) and (A.35) gives

$$\begin{aligned} N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}}) &= N \cdot \mathbb{V}ar(\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}]) + \mathbb{E}[N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D})] \\ &< \sigma^2_{\mathcal{X}} + 29/2C_N. \end{aligned}$$

Combining (A.31) and(A.33) gives (A.30) since $16N^{-1} \leq C_N$ for all $N \geq 100$.

■

Lemma A.8 summarizes the results of previous Lemma. It completes the proof of Theorem 1(1).

Lemma A.8 (MSE Ranking)**.** *For any* $\epsilon \in (0, N^{-1/2})$ *and* $N \geq N_0 = 1{,}740$*, MSE ranking* (15) *holds.*

Proof of Lemma A.8. Let $N\epsilon^2 \leq 1$ and $C_N = \sqrt{2.25 \ln N/N}$. Lemma A.6 gives a lower bound on $MSE(\widehat{W}_{G^*})$

$$N \cdot MSE(\widehat{W}_{G^*}) > 201(1 - 10C_N).$$

Lemma A.7 gives an upper bound on $MSE(\widehat{W}_{\mathcal{X}})$

$$N \cdot MSE(\widehat{W}_{\mathcal{X}}) < 2 + 16C_N. \tag{A.36}$$

Adding the inequalities gives the result

$$N \cdot MSE(\widehat{W}_{G^*}) - N \cdot MSE(\widehat{W}_{\mathcal{X}}) > 199 - 2026C_N > 0, \quad \forall N \geq N_0 = 1{,}740 < 2000.$$

$\blacksquare$

A.4 Proof of Theorem 2.

Let $G^*\Delta G$ denote the symmetric difference of sets G^* and G. Let $P(X \in G^*\Delta G)$ denote the share of people to be treated differently from the optimal policy, or the non-optimal share. Lemma A.7 in Kitagawa and Tetenov (2018a) (cf. Tsybakov (2004)) bounds the welfare gap in terms of non-optimal share. Lemma A.9 complements this result by adding an upper bound on the standard deviation gap. Define the large-sample analog of the exact expected length $\mathbb{E}[LCB_G - LCB_{G^*}]$ as

$$\Delta_G := N^{-1/2} z_{1-\alpha}(\sigma_{G^*} - \sigma_G) - (W_{G^*} - W_G). \tag{A.37}$$

Lemma A.9. *Suppose Assumptions 2.1–3.1 hold. Then, (1) welfare gap is bounded*

$$C_B(P(X \in G^*\Delta G))^{1+1/\delta} \leq W_{G^*} - W_G \leq MP(X \in G^*\Delta G) \tag{A.38}$$

with $C_B = (\eta\delta\kappa/M)^{1+1/\delta} > 0$*;*
(2) variance gap is bounded

$$\sigma^2_{G^*} - \sigma^2_G \leq (3M^2/\kappa)P(X \in G^*\Delta G). \tag{A.39}$$

(3) the standard deviation gap is bounded

$$\sigma_{G^*} - \sigma_G \leq (3M^2/2\bar{\sigma}\kappa)P(X \in G^*\Delta G). \tag{A.40}$$

Proof of Lemma A.9. The bound (A.38) is established in Tsybakov (2004) (cf., Lemma A.7 in Kitagawa and Tetenov (2018a)). Step 1 introduces notation. Step 2 establishes (A.39). Step 3 establishes (A.40).

Step 1. We introduce extra notation to simplify variance expressions. Given a policy G, let $G_1 = G$ and $G_0 = G^c$. Then, the welfare W_G in (1) can be equivalently rewritten as

$$W_G = \mathbb{E}[\sum_{d \in \{1,0\}} m(d, X)1\{X \in G_d\}].$$

Since X cannot be in G and G^c at the same time, $1\{X \in G\}1\{X \in G^c\} = 0$ a.s. .

Simplifying the variance gives

$$\mathbb{E}(\sum_{d\in\{1,0\}} m(d,X)1\{X\in G_d\})^2 = \mathbb{E}[\sum_{d\in\{1,0\}} m^2(d,X)1\{X\in G_d\}].$$

Invoking law of total variance gives

$$\sigma_G^2 = \sum_{d\in\{1,0\}} \mathbb{E}\Big[\Big(\frac{\sigma^2(d,X)}{P(D=d\mid X)} + m^2(d,X)\Big)1\{X\in G_d\}\Big] - W_G^2. \tag{A.41}$$

For any policy G, define

$$\begin{aligned} T_{1G} &= \mathbb{E}\Big[\Big(\frac{\sigma^2(1,X)}{\pi(X)} - \frac{\sigma^2(0,X)}{1-\pi(X)}\Big)\Big(1\{X\in G^*\setminus G\} - 1\{X\in G\setminus G^*\}\Big)\Big] \\ T_{2G} &= \mathbb{E}\Big[\Big(m^2(1,X) - m^2(0,X)\Big)\Big(1\{X\in G^*\setminus G\} - 1\{X\in G\setminus G^*\}\Big)\Big]. \end{aligned}$$

The gap in squared welfares is

$$T_{3G} = -(W_{G^*}^2 - W_G^2).$$

Invoking (A.41) for $G = G^*$ and any $G\in\mathcal{X}$ gives

$$\sigma_{G^*}^2 - \sigma_G^2 = T_{1G} + T_{2G} + T_{3G}. \tag{A.42}$$

Step 2. By Assumption 2.2(1) and Jensen inequality, $m^2(d,x)\le M^2$ and $\sigma^2(d,x)\le M^2$ for all $x\in\mathcal{X}$. Invoking Assumption 2.2 (2) gives

$$T_{1G} \le M^2/(\kappa)P(X\in G^*\Delta G), \quad T_{2G}\le M^2P(X\in G^*\Delta G).$$

The final term is bounded as

$$\begin{aligned} T_{3G} &\ge -(W_{G^*} - W_G)|W_{G^*} + W_G| \ge^i -M(W_{G^*} - W_G) \\ &\ge^{ii} -M^2P(X\in G^*\Delta G). \end{aligned}$$

where (i) follows from $|W_{G^*} + W_G|\le M$ and (ii) follows from (A.38). Thus, (A.39) holds with $C_V = M^2/(2\kappa) + M^2/\kappa \le 3M^2/(\kappa)$.

Step 3. An upper bound on the standard deviation gap is

$$\sigma_{G^*} - \sigma_G \leq^{i} \frac{\sigma^2_{G^*} - \sigma^2_G}{2\underline{\sigma}} \leq^{ii} 3M^2/(2\underline{\sigma}\kappa)$$

where (i) follows from $\sigma^2_{G^*} \geq \underline{\sigma}^2$ and $\sigma^2_{\mathcal{X}} \geq \underline{\sigma}^2$ and (ii) from (A.39). ■

Proof of Theorem 2(1). Let $C_A = N^{-1/2} z_{1-\alpha} 3M^2/(2\underline{\sigma}\kappa)$ and C_B be as defined in Lemma A.9. Define the function

$$g(x) = C_A x - C_B x^{1/\delta+1}, \quad x \in \mathrm{R}. \tag{A.43}$$

The approximate expected length Δ_G is bounded as

$$\Delta_G \leq^{i} C_A P(X \in G^* \Delta G) - (W_{G^*} - W_G) \leq^{ii} g(P(X \in G^* \Delta G)),$$

where (i) follows from (A.40) and (ii) from the lower bound (A.38). The function $g(x)$ is globally concave. Its global maximum and maximizer are

$$g(x^*) = \left(\frac{C_A \delta}{C_B(1+\delta)} \right)^{\delta} \frac{C_A}{1+\delta}, \quad x^* = \left(\frac{C_A \delta}{C_B(1+\delta)} \right)^{\delta}.$$

The expression $g(x^*)$ matches the upper bound (23) with a constant

$$\bar{C} = \frac{(z_{1-\alpha} 3M^2/\kappa)^{\delta+1}}{\delta+1} \left(\frac{\delta}{\delta+1} \right)^{\delta} C_B^{-\delta}.$$

■

Proof of Theorem 2 (2). The proof of Theorem 2(2) is constructive and consists of three steps. Step 1 describes a class of DGPs. Step 2 shows that the proposed DGPs belong to the class $\mathcal{P}(M, \kappa, \eta, \delta, \underline{\sigma})$. Step 3 establishes the lower bound (24).

Step 1. We take $X \sim U[0,1]$. The propensity score is constant and equals $1/2$. Let $\epsilon \in (0,1/2)$ and ν be an odd integer such that $\delta \geq 1/\nu$. The conditional means and variances are

$$m(1,x) = 0, \quad \sigma^2(1,x) = 1$$
$$m(0,x) = -(x-\epsilon)^{\nu}, \quad \sigma^2(0,x) = 2.$$

The conditional average treatment effect reduces to

$$\tau(x) = m(1,x) - m(0,x) = (x-\epsilon)^{\nu}. \tag{A.44}$$

The first-best policy reduces to

$$G^* = \{X \in [\epsilon, 1]\}. \tag{A.45}$$

Step 2. We show that the proposed DGP belongs to $\mathcal{P}(M, \kappa, \eta, \delta, \underline{\sigma})$ for any $\epsilon \in (0,1/2)$. Note that $\epsilon^{\nu} \leq (1-\epsilon)^{\nu}$ since $\epsilon \leq 1/2$. For small values of $t \in (0, \epsilon^{\nu})$, we have

$$P(|X-\epsilon| \leq t^{1/\nu}) = (\epsilon + t^{1/\nu}) - (\epsilon - t^{1/\nu}) = 2t^{1/\nu}$$

and, for $t \geq \epsilon^{\nu}$,

$$P(|X-\epsilon| \leq t^{1/\nu}) \leq \epsilon + t^{1/\nu} \leq 2t^{1/\nu}.$$

Thus, the restriction (20) holds for any δ smaller than $1/\nu$ and $\eta = 2^{\nu}$.

Step 3. We prove the lower bound (24). The first-best policy differs from $\mathcal{X}$ only for $X \in [0, \epsilon]$. The welfare gap is

$$W_{G^*} - W_{\mathcal{X}} = -\int_0^{\epsilon} (x-\epsilon)^{\nu} dx + \int_{\epsilon}^{1} 0 dx - 0 = \epsilon^{\nu+1}/(\nu+1).$$

The variance gap is obtained by plugging $G = \mathcal{X}$ into (A.42) and noting that both terms $T_{2\mathcal{X}} = \epsilon^{2\nu+2}/(2\nu+1)$ and $T_{3\mathcal{X}} = -\epsilon^{2\nu+2}/(\nu+1)^2$ involve a term $\epsilon^{2\nu+2}$. The variance gap is

$$\sigma^2_{G^*} - \sigma^2_{\mathcal{X}} = 2\epsilon + \epsilon^{2\nu+2}\frac{\nu^2}{(2\nu+1)(\nu+1)^2}.$$

The first-best variance is upper bounded as $\sigma^2_{G^*} < \sigma^2_{\mathcal{X}} + 4\epsilon = 2 + 4\epsilon \leq 4$. The standard deviation gap

$$\sigma_{G^*} - \sigma_{\mathcal{X}} > \epsilon/2.$$

Setting $\epsilon^{\nu} = N^{-1/2} z_{1-\alpha}/2$ gives a lower bound

$$\Delta_{\mathcal{X}} > N^{-1/2(1+1/\nu)}(z_{1-\alpha}/2)^{1+1/\nu}\nu/(\nu+1),$$

which matches the lower bound (24) with $\delta = 1/\nu$ and the constant $\underline{C} = (z_{1-\alpha}/2)^{1+1/\nu}\nu/(\nu+1)$. Thus, $\Delta_{\mathcal{X}} > \underline{C}N^{-1/2(1+1/\nu)} > \underline{C}N^{-1/2(1+\delta)}$ for any $\delta \geq 1/\nu$.

■

B Auxiliary Theoretical Statements

Lemma B.1 gives an upper bound on the welfare gap in terms of variance of CATE.

Lemma B.1 (Upper bound on welfare gap)**.** *The following upper bound holds*

$$0 \leq W_{G^*} - \max(W_{\mathcal{X}}, W_{\emptyset}) \leq \sqrt{\mathbb{V}ar(\tau(X))}. \tag{B.1}$$

Proof of Lemma B.1. For any points $a, b \in \mathrm{R}$,

$$\max(a, 0) - \max(b,0) \leq |a - b|. \tag{B.2}$$

Plugging $a = \tau(X)$ and $b = \mathbb{E}[\tau(X)] = \tau$ into (B.2) and averaging

$$\mathbb{E}[\max(\tau(X), 0)] - \max(\tau, 0) \leq \mathbb{E}|\tau(X) - \tau| \leq \sqrt{\mathbb{V}ar(\tau(X))},$$

where the last step follows from Jensen inequality for $f(t) = t^2$. ■

The inequality (B.1) is useful in two respects. First, it creates a sufficient condition for the welfare gap to be small. If the treatment effect heterogeneity is vanishingly small $\mathbb{V}ar(\tau(X)) = o(N^{-1})$, the resulting welfare gap decays faster than the parametric rate. Second, the inequality (B.1) is also of independent interest. We establish an upper bound on *Jensen gap* for a convex yet non-smooth function $f(t) = \max(t,0)$. In this context, standard tools used for convex differentiable functions (e.g., Abramovich and Persson (2016)) do not apply.

C Auxiliary Empirical Details

Table 1, Rows 2 and 3. We describe the standard doubly-robust cross-fit estimator used in Table 1, Rows 2 and 3 and in Table C.1. The welfare gain parameter is estimated using doubly-robust moment condition

$$W_{gain} = \mathbb{E}\Bigg[\Bigg(\tau(X) + \frac{D}{\pi(X)}(Y - m(1,X)) - \frac{(1-D)}{1-\pi(X)}(Y - m(0,X))\Bigg)1\{\tau(X) \geq 0\}\Bigg]$$
$$=: \mathbb{E}\Bigg[\psi_{gain}(W)\Bigg].$$

For cross-fitting purposes, we partition the data indices $1, \ldots, N$ into $L = 2$ disjoint subsets I_ℓ of about equal size, $\ell = 1, 2$. Let $\hat{\tau}_\ell$ be estimators of CATE constructed using all observations not in I_ℓ. Define the cross-fit estimate of welfare gain as

$$\widehat{W}_{gain} = N^{-1} \sum_{\ell=1}^{2} \sum_{i \in I_\ell} \psi_{gain}(W_i, \hat{\tau}_\ell) \tag{C.1}$$

$$\widehat{\sigma}^2_{gain} = N^{-1} \sum_{\ell=1}^{2} \sum_{i \in I_\ell} (\psi_{gain}(W_i, \hat{\tau}_\ell) - \widehat{W}_{gain})^2, \tag{C.2}$$

where the standard debiased inference is established under the margin condition (20) and $o(N^{-1/4})$ rate conditions on $\widehat{\tau}_\ell$. (e.g., Luedtke and van der Laan, 2016; Kallus et al., 2020). The $100(1-\alpha)\%$ Lower Confidence Band is

$$LCB_{gain} = \widehat{W}_{gain} - N^{-1/2} z_{1-\alpha} \widehat{\sigma}_{gain}. \tag{C.3}$$

Table C.1 presents an efficient estimator of the welfare gain based on series regression with power $d = 1,2,3,4$. The power $d = 4$ corresponds to Row 2 of Table 1.

Table 1, Row 4. To consider a data-driven choice of G, we partition the sample into two parts I_1 and I_2 of sizes $N/3$ and $2/3N$, respectively. Let

$$\widehat{G}_1 := \{X : \widehat{\tau}_1(X) > 0\},$$

where $\widehat{\tau}_1$ is estimated via plug-in rule using random forest regression of earnings of $Educ$ and $PreEarn$. A sample analog of $W_{gain,G}$ is

$$\widehat{W}_{gain,G} = \frac{1}{|I_2|} \sum_{i \in I_2} \left(\frac{D_i}{\pi(X_i)} - \frac{1-D_i}{1-\pi(X_i)} \right) Y_i 1\{X_i \in G\}.$$

Conditional on the data in the partition I_1, we have

$$\sqrt{|I_2|}(\widehat{W}_{gain,\widehat{G}_1} - W_{gain,\widehat{G}_1}) \Rightarrow^d N(0, \sigma^2_{\widehat{G}_1}) \mid (W_i)_{i \in I_1}.$$

The $100(1-\alpha)\%$ Lower Confidence Band

$$LCB_{gain,G_1} = \widehat{W}_{gain,\widehat{G}_1} - |I_2|^{-1/2} z_{1-\alpha} \widehat{\sigma}_{gain,\widehat{G}_1} \tag{C.4}$$

attains correct coverage condition on the data in I_1, and, therefore, unconditionally.

References

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*, 91–117.

Abramovich, S. and L.-E. Persson (2016). Some new estimates of the jensen gap. *Journal of Inequalities and Applications 39*, 285–314.

Athey, S. and S. Wager (2021, January). Policy learning with observational data. *Econometrica 89*, 133–161.

Berger, R. L. (1997). *Likelihood Ratio Tests and Intersection-Union Tests*, pp. 225–237. Boston, MA: Birkhauser Boston.

Bhattacharya, D. and P. Dupas (2012). Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics 167*(1), 168–196.

Bloom, H. S., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos (1997). The benefits and costs of jtpa title ii-a programs: Key findings from the national job training partnership act study. *The Journal of Human Resources 32*(3), 549–576.

Chamberlain, G. (2011, 09). 1011 Bayesian Aspects of Treatment Choice. In *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*, 667–737.

Table C.1: Welfare Gain Per Capita: Estimates and Lower Confidence Bands

	(1)	(2)	(3)	(4)	(5)
Treatment Rule	Share of People to be Treated	Welfare Gain (St.Error)	95% LCB	Welfare Gap (USD)	LCB Gap (%)
Treat Everyone ($PreEarn$)	1.00	1289.66 (347.82)	717.52		
Series Regression $PreEarn$	1.00	1286.058 (346.34)	716.38	≈ 0	$\approx 0\%$
Series Regression $PreEarn + PreEarn^2$	1.00	1296.83 (346.54)	726.82	≈ 0	$\approx 0\%$
Series Regression $\sum_{j=1}^{3} PreEarn^j$	1.00	1162.513 (360.82)	569.01	127.47	20%
Series Regression $\sum_{j=1}^{4} PreEarn^j$	0.992	1043.222 (394.68)	394.03	246.66	45%

Notes. The outcome variable is 30-Month Post-Program Cumulative Earnings in USD. Welfare gain is the first-best welfare net of the control average outcome, as in (28). Each row reports the plug-in rule estimated via series regression with d = 1,2,3,4 polynomials. The columns are the same as in Table 1. See text for details.

Dehejia, R. H. (2005). Program evaluation as a decision problem. *Journal of Econometrics 125*(1), 141–173. Experimental and non-experimental evaluation of economic policy and models.

Hahn, J. (1998, March). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*(2), 315–331.

Heckman, J., H. Ichimura, and P. Todd (1997). Matching as an econometric evaluation estimator:evidence from evaluating a job training programme. *Review of Economic Studies 64*, 605–654.

Hirano, K. and J. Porter (2009). Asymptotics for statistical treatment rules. *Econometrica 77*, 1683–1701.

Hirano, K. and J. Porter (2012). Impossibility results for nondifferentiable functionals. *Econometrica 80*, 1769–1790.

Johnson, M. S., D. I. Levine, and M. W. Toffel (2023, October). Improving regulatory effectiveness through better targeting: Evidence from osha. *American Economic Journal: Applied Economics 15*(4), 3067.

Kallus, N., X. Mao, and A. Zhou (2020). Assessing algorithmic fairness with unobserved protected class using data combination.

Kitagawa, T. and A. Tetenov (2018a). Supplement to: Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica Supplemental Material*.

Kitagawa, T. and A. Tetenov (2018b). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica 86*, 591–616.

Levy, J., M. van der Laan, A. Hubbard, and R. Pirracchio (2021). A fundamental measure of treatment effect heterogeneity. *Journal of Causal Inference 9*(1), 83–108.

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

Mammen, E. and A. B. Tsybakov (1999). Smooth discrimination analysis. *The Annals of Statistics 27*(6), 1808 – 1829.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica 72*, 1221–1246.

Mbakop, E. and M. Tabord-Meehan (2021, March). Model selection for treatment choice: Penalized welfare maximization. *Econometrica 89*, 825–848.

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

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics 151*, 70–81.

Sun, L. (2021). Empirical welfare maximization with constraints.

Tetenov, A. (2012). Statistical treatment choice based on asymmetric minimax regret criteria. *Econometrica 166*, 157–165.

Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics 32*(1), 135 – 166.

Zhang, K. W., L. Janson, and S. A. Murphy (2021). Inference for batched bandits.