On the Lower Confidence Band for the Optimal Welfare

Kirill Ponomarev* Vira Semenova†

This version: October 9, 2024

Abstract

This article addresses the question of reporting a lower confidence band (LCB) for 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 show that under empirically relevant data-generating processes, this procedure can be dominated by an LCB corresponding to suboptimal welfare, with the average difference of the order $N^{-1/2}$. We relate the first-order dominance result to a lack of uniformity in the margin assumption, a standard sufficient condition for debiased inference on the optimal welfare ensuring that the first-best policy is well-separated from the suboptimal ones. Finally, we show that inverting the existing tests from the moment inequality literature produces LCBs that are robust to the non-uniqueness of the optimal policy and easy to compute. We find that this approach performs well empirically in the context of the National JTPA study.

Keywords: policy learning, optimal welfare, lower confidence band, cross-fitting, double machine learning, 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 Alberto Abadie, Isaiah Andrews, Jiafeng (Kevin) Chen, Bryan Graham, Michael Jansson, Patrick Kline, Elena Manresa, Demian Pouzo, Andres Santos, Azeem Shaikh, 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; Sasaki and Ura, 2024; Viviano, 2024). To give the policy-maker confidence that the proposed treatment policy improves upon the status quo, it is customary to report a point estimate and a lower confidence band for the corresponding optimal welfare (e.g., Johnson, Levine, and Toffel, 2023). The inference problem is complicated by the potential non-uniqueness of the optimal policy and noisy nonparametric estimators of the unknown regression functions (Hirano and Porter, 2012; Luedtke and van der Laan, 2016).

In this note, we show that focusing on a simpler, suboptimal policy sometimes results in better estimates and tighter confidence bands for the optimal welfare, with the difference of $N^{-1/2}$ order of magnitude. We characterize the class of data-generating processes for which such first-order dominance is possible and relate it to the lack of uniformity in the margin assumption of Tsybakov (2004), commonly imposed in policy learning and debiased inference. We highlight the importance of using uniformly valid asymptotic approximations in the context of policy learning and show how existing methods for inference in moment inequality models (e.g., Andrews and Soares, 2010; Chernozhukov, Lee, and Rosen, 2013; Romano, Shaikh, and Wolf, 2014; Canay and Shaikh, 2017; Chernozhukov, Chetverikov, and Kato, 2019; Bai, Santos, and Shaikh, 2022) can be used to construct uniformly valid confidence bands for the optimal welfare. Revisiting the National JTPA study (Bloom, Orr, Bell, Cave, Doolittle, Lin, and Bos, 1997), we find the above considerations empirically relevant.

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. Letting $G \subset \mathcal{X}$ be a subset of the 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 (or optimal) welfare denoted as

$$W_{G^*} = \max_{G \subset \mathcal{X}} W_G.$$

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, conditional on X:*

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

Under Assumption 2.1, 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

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

Moreover, 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 the CATE. The object of interest of this note is the first-best welfare W_{G^*}, rather than the policy G^* that attains it.

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 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 for the optimal welfare. 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^*}, we focus on two policies: the first-best policy, $G = G^*$, which is unknown, and the "treat everyone" policy, $G = \mathcal{X}$.

First, consider the first-best policy $G = G^*$. The semiparametric efficiency bound for the corresponding welfare W_{G^*} is characterized in Luedtke and van der Laan (2016). The bound is the same as if the first-best policy was known[1] and can be attained by an oracle estimator

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

of the form

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

where the regression estimators are chosen to prevent the division by zero. Such $\widehat{W}_{G^*}$ is an ideal, infeasible estimator of W_{G^*}.

Next, 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:

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

The mean squared errors of the two estimators are given by:

$$\begin{aligned} MSE(\widehat{W}_{\mathcal{X}}) &= \mathbb{E}[(\widehat{W}_{\mathcal{X}} - W_{G^*})^2];\\ MSE(\widehat{W}_{G^*}) &= \mathbb{E}[(\widehat{W}_{G^*} - W_{G^*})^2]. \end{aligned} \tag{9}$$

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

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

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

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 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 smaller than LCB_{G^*}, on average.

Mimicking the expected length criterion for two-sided confidence intervals, we compare the confidence bands introduced above in terms of the expected values of the LCBs. Similar to the two-sided case, these expected values may not always be finite and may be complicated by the biases of the point estimates. Yet, when the expected difference $\mathbb{E}\left[LCB_{\mathcal{X}} - LCB_{G^*}\right]$ is tractable, it can be first-order approximated as

$$\mathbb{E}\left[LCB_{\mathcal{X}} - LCB_{G^*}\right] = \Delta_{\mathcal{X}} + o(N^{-1/2}),$$

where the approximate difference in expected lengths is

$$\Delta_{\mathcal{X}} = N^{-1/2} z_{1-\alpha}(\sigma_{G^*} - \sigma_{\mathcal{X}}) - (W_{G^*} - W_{\mathcal{X}}). \tag{15}$$

The approximate criterion captures an important trade-off between precision and slackness $W_{G^*} - W_{\mathcal{X}}$ of suboptimal welfare, which we refer to as the *welfare gap* of the policy $\mathcal{X}$.

3 Theoretical Results

3.1 First-Order Dominance

Our first main result shows that $\widehat{W}_{G^*}$ can be dominated by $\widehat{W}_{\mathcal{X}}$, as an estimator of W_{G^*}, in terms of MSE in finite samples. Likewise, LCB_{G^*} can be dominated by $LCB_{\mathcal{X}}$ in terms of expected length at the first-order $N^{-1/2}$ scale.

Theorem 1 (First-order dominance)**.** *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*

$$MSE(\widehat{W}_{\mathcal{X}}) < MSE(\widehat{W}_{G^*}); \tag{16}$$

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

$$\mathbb{E}[LCB_{\mathcal{X}} - LCB_{G^*}] > CN^{-1/2}. \tag{17}$$

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

$$\tau(1) = -\epsilon, \quad \tau(0) = 1/2, \tag{18}$$

the welfare gap is

$$0 \le W_{G^*} - W_{\mathcal{X}} \le \epsilon, \tag{19}$$

and the difference in efficiency bounds (i.e., variance gap) is

$$\sigma^2_{G^*} - \sigma^2_{\mathcal{X}} > 3/2, \quad \forall \epsilon > 0. \tag{20}$$

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 (16) and (17) hold.

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

Theorem 1 has two main takeaways. First, it shows that there may exist better estimators and tighter confidence bands for the optimal welfare, which are based on suboptimal policies. In what follows, we characterize the set of DGP-s for which such results are possible and discuss ways to leverage them in practice. Second, Theorem 1 highlights the distinction between the two-sided and one-sided inferential objectives. In the two-sided case, the bias typically must vanish faster than the variance (through under-smoothing or debiasing) to ensure valid coverage. In the 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.

[2]A deleted neighborhood of a point excludes the point itself.

3.2 Higher-Order Dominance under the Margin Assumption

In this Section, we investigate what features of the data-generating processes drive the first-order dominance result of Theorem 1. We connect this result to the lack of uniformity in the margin assumption, which we introduce below.

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

$$P(|\tau(X)| < t) \leq (t/\eta)^{\delta}, \quad \forall t \in [0, \eta). \tag{21}$$

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 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 (21) holds for any $\delta < 1$ with η small enough. Similarly, if $\tau(X)$ has finite support and $P(\tau(X) = 0) = 0$, then (21) 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) and welfare maximization (e.g., Qian and Murphy, 2011; Kitagawa and Tetenov, 2018b; Mbakop and Tabord-Meehan, 2021) and standard two-sided debiased inference (e.g., Luedtke and van der Laan, 2016; Kallus, Mao, and Zhou, 2020). See Appendix C for a review.

We wish to know if the first-order dominance result of Theorem 1 carries over if the DGPs are further restricted to obey Assumption 3.1 with a uniform lower bound on δ and η. An informative answer to this question requires comparing the first-best LCB_{G^*} to LCBs other than $LCB_{\mathcal{X}}$. 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

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

$$LCB_G = \widehat{W}_G - N^{-1/2} z_{1-\alpha}\widehat{\sigma}_G, \tag{23}$$

where $\widehat{\sigma}_G$ is a consistent estimator of σ_G. Since the welfare gap $W_{G^*} - W_G$ is non-negative, the validity of LCB_G follows from an argument similar to (14). The approximate expected length generalizes to

$$\Delta_G := N^{-1/2} z_{1-\alpha}(\sigma_{G^*} - \sigma_G) - (W_{G^*} - W_G). \tag{24}$$

Our second main result bounds the magnitude of Δ_G uniformly over the class of DGPs obeying the margin assumption and the class of all treatment policies $G \subseteq \mathcal{X}$.

Theorem 2 (Higher-order dominance)**.** *Let $\mathcal{P}(M, \kappa, \eta, \delta, \underline{\sigma})$ denote the class of DGPs obeying Assumptions 2.1, 2.2, 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

$$\sup_{P \in \mathcal{P}(M,\kappa,\eta,\delta,\underline{\sigma})} \sup_{G \subseteq \mathcal{X}} \Delta_G \leq \overline{C} N^{-1/2(1+\delta)} \tag{25}$$

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

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

$$\Delta_G > \underline{C} N^{-1/2(1+\delta)} \tag{26}$$

for a sequence of DGPs in $\mathcal{P}(M, \kappa, \eta, \delta, \underline{\sigma})$ and $0 < \underline{C} < \overline{C}$.

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 (25). The lower bound (26) recovers the negative result (17) 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 (25), let $P(G^* \triangle G)$ 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

$$W_{G^*} - W_G \geq C_1 P(G^* \triangle G)^{1+1/\delta} \tag{27}$$

for some $C_1 = C_1(\delta) > 0$ (Tsybakov, 2004). Additionally, we prove that the standard deviation gap is upper bounded as

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

Thus, the margin assumption excludes DGPs of the form (19)–(20) where only the welfare gap vanishes. Theorem 2 shows that placing a lower bound on the margin parameter δ is just enough to "restore" the optimality of the conventional first-best LCB_{G^*} at $N^{-1/2}$ scale.

Without a lower bound on δ, the LCB_{G^*} can be dominated at a rate arbitrarily close to $N^{-1/2}$.

3.3 Implications and Open Questions

We conclude the discussion with the following Remarks. Remark 1 proposes a test of non-uniqueness of optimal policy in the setting with discrete covariates. Remark 2 describes another testable implication of margin assumption that proves empirically relevant. Remark 3 discusses the failure of standard Gaussian inference when the margin assumption fails to hold uniformly. Remark 4 discusses robust alternatives to LCB_{G^*} obtained via inverting of least-favorable tests. Remark 5 describes an upper bound on welfare gap that can be of independent interest.

Remark 1 (Testing uniquness of the optimal policy)**.** Let X be a discrete covariate taking J distinct values with positive probabilities. Then, 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 $\tau(j) = 0$ for some $j \in \{1, 2, \dots, J\}$. The null hypothesis

$$H_0 : \exists j : \ \tau(j) = 0 \tag{28}$$

is a union of J simple hypotheses $H_{0j} : \tau(j) = 0$. Then, letting R_j denote the rejection region for testing H_{0j}, the test with a rejection region

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

is valid for H_0, although typically conservative (see, e.g., Berger, 1997).

Remark 2 (Testing the margin assumption)**.** In the general case where both discrete and continuous covariates are present, Assumption 3.1 is no longer equivalent to Assumption 2.2(3). We describe a testable implication that we find empirically relevant in Section 4. Consider the lower bound on welfare gap given in (27), where the constant $C_1 = C_1(\delta) = \eta\delta(\frac{1}{1+\delta})^{1+\frac{1}{\delta}}$. Given a lower bound $\underline{\delta} > 0$ and fixing $\eta > 0$, consider a null hypothesis $H_0 : \delta \geqslant \underline{\delta}$. Since both functions $\delta \to C_1(\delta)$ and $\delta \to P(G^* \triangle G)^{1+\frac{1}{\delta}}$ are increasing in δ, the lower bound (27) on welfare gap implies that, for any policy G,

$$C_1(\underline{\delta}) P(G^* \triangle G)^{1+\frac{1}{\underline{\delta}}} - (W_{G^*} - W_G) \leqslant 0. \tag{29}$$

In particular, if the welfare gap $W_{G^*} - W_G$ of some policy G vanishes with sample size, the share of people treated differently under G and G^*, must vanish, too. Existing methods from the moment inequality literature, such as Chernozhukov et al. (2013), can then be applied to construct a test. Pursuing this formally is left for future work.

Remark 3 (Failure of standard inference in the absence of margin assumption)**.** Inference on the first-best welfare robust to violations of margin assumption remains an important open question. The key challenge is that when the margin assumption is violated, the cross-fit estimators of W_{G^*} have a non-standard, heavy-tailed distribution. As a result, the standard Wald-type confidence intervals are no longer valid. A similar failure of normality has also been pointed out in the context of inference in batched bandits in Zhang, Janson, and Murphy (2021).

Remark 4 (Lower confidence band via moment inequalities)**.** One can use the insights from the moment inequality literature to construct a uniformly valid lower confidence band for the first-best welfare $\theta = W_{G^*}$ that is robust to potential violations of margin assumption. Given any class $\mathcal{G}$ of policies $G \subseteq \mathcal{X}$, consider testing the null hypothesis $H_0 : W_G \leqslant \theta, \ \ \forall G \in \mathcal{G}$. For each policy G, let $\widehat{W}_G$ be an efficient estimator of W_G is given in (22) and $\widehat{\sigma}_G$ be estimated standard deviation. The least-favorable test based on the maximum test statistic takes the form:

$$\widehat{\phi}_N = \mathbf{1}\left(\max_{G\in\mathcal{G}} \frac{\sqrt{N}(\widehat{W}_G - \theta)}{\widehat{\sigma}_G} > \widehat{c}_{1-\alpha}\right),$$

where $\widehat{c}_{1-\alpha}$ is the $(1-\alpha)$-quantile of the maximum of a Gaussian vector (or a Gaussian process, if $\mathcal{G}$ is infinite). Inverting the test yields a $100(1-\alpha)\%$-lower confidence band of the form:

$$LCB^{LF}_{\mathcal{G}} = \max_{G\in\mathcal{G}}(\widehat{W}_G - \widehat{c}_{1-\alpha}N^{-1/2}\widehat{\sigma}_G). \tag{30}$$

Notice that $LCB^{LF}_{\mathcal{G}}$ constructs conservative lower confidence bands for each policy G and selects the largest one, thus addressing the trade-off between slackness $W_{G^*} - W_G$ and precision. This procedure can be made less conservative by incorporating the Generalized Moment Selection step from Andrews and Soares (2010). In Appendix B.2, we provide further details and a closed-form expression for the corresponding LCB. Other moment selection procedures are possible: Romano et al. (2014), if $\mathcal{G}$ is low-dimensional, Chernozhukov et al. (2019) or Bai et al. (2022) if $\mathcal{G}$ is finite yet high-dimensional, or Chernozhukov et al. (2013) if $\mathcal{G}$ infinite. The class $\mathcal{G}$ itself may either contain all available policies or be constrained based on economic intuition or anticipated treatment heterogeneity.

Remark 5 (The role of treatment heterogeneity)**.** The magnitude of the welfare gap between the optimal policy and a suboptimal one is directly related to the treatment effect heterogeneity. In Lemma B.1 in Appendix B.1, we prove that the welfare gap is upper bounded by the standard deviation of CATE

$$W_{G^*} - \max(W_{\mathcal{X}}, W_{\emptyset}) \leq \sqrt{\mathbb{V}ar(\tau(X))}. \tag{31}$$

If there is no treatment effect heterogeneity, the welfare gap reduces to zero. If the treatment effect heterogeneity is "locally vanishing", that is, $\mathbb{V}ar(\tau(X))$ decays faster than N^{-1}, the welfare gap of $\max(W_{\mathcal{X}}, W_{\emptyset})$ vanishes faster than the parametric rate. In line with Theorem 1, this indicates a DGP for which targeting $\max(W_{\mathcal{X}}, W_{\emptyset})$ may deliver more precise estimates and tighter confidence bands for W_{G^*}. Empirically detecting such regimes would be possible via inference methods of Levy, van der Laan, Hubbard, and Pirracchio (2021) on $\mathbb{V}ar(\tau(X))$. A similar result holds if there is no treatment effect heterogeneity along some dimensions of the covariate space. Letting $T = T(X)$ be any measurable transformation of X and W_{T^*} denote the optimal welfare in the class of treatment rules that depend only on T, in Lemma B.2 in Appendix B.1 we show that:

$$W_{G^*} - W_{T^*} \leqslant \mathbb{E}[\sqrt{\mathbb{V}ar(\tau(X) \,|\, T)}].$$

For example, if $T(X) = X_1$ is a subvector of $X = (X_1, X_{-1})$, the upper bound reduces to $\mathbb{E}[\sqrt{\mathbb{V}ar(\tau(X) \,|\, X_1)}]$. Thus, if there is no heterogeneity in treatment effects beyond that explained by covariates other than X_1, focusing on W_{T^*} may lead to better estimates and tighter confidence bounds. Regimes of small treatment effect heterogeneity have been found empirically relevant in e.g., Athey, Keleher, and Spiess (2024).

4 Empirical Application

In this Section, we revisit the National Job Training Partnership Act (JTPA) study, considered in Heckman, Ichimura, and Todd (1997) and Abadie, Angrist, and Imbens (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. Since JTPA is a randomized trial with a constant propensity score, the following unconditional independence holds

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

and the first-best welfare is identified in each of the models (D,Y), $(PreEarn, D,Y)$, $(Educ, D,Y)$ and (X,D,Y). We are interested in estimating and constructing a LCB for the first-best welfare gain relative to the "treat no one" policy:

$$W_{gain} = W_{G^*} - W_{\emptyset}. \tag{32}$$

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

$$W_{gain} \geq W_G - W_{\emptyset}. \quad \forall G \subseteq \mathcal{X}. \tag{33}$$

As a result, for any collection $\mathcal{G}$ of policies $G \in \mathcal{G}$, $W_{gain} \geq \max_{G\in\mathcal{G}} W_G - W_{\emptyset}$. As a reference point, we consider the average treatment effect

$$W_{gain} \geq \underbrace{W_{\mathcal{X}} - W_{\emptyset}}_{ATE}. \tag{34}$$

Table 1 summarizes the results. Row 1 reports the ATE in (34) 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 to estimate the propensity score. Rows 2 and 3 report the first-best welfare gain in the model $(PreEarn, D,Y)$. Since $PreEarn$ is continuously distributed, we find the standard efficient inference based on cross-fit doubly robust estimator to be appropriate. In the first stage, the CATE function of $PreEarn$ is estimated via series regression (Row 2) and random forest (Row 3). Row 4 of Table 1 considers the first-best unconstrained welfare in the model (X,D,Y). Out of caution for potential non-uniqueness of the optimal policy, we do not cross-fit the two-stage estimators in the light of Remark 3. Instead, we randomly split the sample into two parts and use the first part (one-third of the sample) to estimate a policy $\widehat{G}$ based on a random forest estimate of CATE and the second part to estimate $LCB_{\widehat{G}}$ as in (23). Appendix C provides further empirical details.

Our findings are as follows. First, we show that the welfare gap is vanishing relative to the sampling uncertainty. Comparing the point estimates (Column 2) in Row 1 to those in Rows 2–3, we find the estimated welfare gap (Column 4) in the model $(PreEarn, D,Y)$ ranges between -246 and -220 USD. Its sign can be negative due to the use of the efficient/doubly-robust estimators in the second stage. Furthermore, its magnitude is within two-thirds of the standard error. Thus, the welfare gap is likely small, highlighting the empirical relevance of DGPs described in Theorem 1.

Second, comparing the LCBs (Column 3) in Row 1 to those in Rows 2–3, we find that

$LCB_{\mathcal{X}} = 717.52$ exceeds the first-best LCB by 40% and 28%, respectively, in the model $(PreEarn, D, Y)$ where margin assumption with some $\delta > 0$ is plausible. According to Theorem 2, the approximate difference in expected lengths of the two LCBs vanishes at rate $N^{-1/2(1+\delta)}$. Thus, the implied margin parameter δ for $(PreEarn, D,Y)$ is such that in the sample of size $N = 9,223$, the scale of $N^{-1/2(1+\delta)}$ is economically relevant. This finding highlights the importance of uniform asymptotics robust to vanishing values of δ in the context of policy learning.

Third, we find that the margin assumption is likely violated as long as $Educ$ is used in the analysis. In the model $(Educ, D,Y)$, for eleven out of twelve education groups, the treatment effect is not significant at $\alpha = 0.05$. Thus, the null hypothesis of non-unique optimal policy in (28) cannot be rejected using the test in Remark 1. For the full model (X, D,Y), we consider the heuristic of Remark 2 in the absence of a formal alternative. Comparing the welfare gap and the share of people not recommended for treatment (i.e., the LHS and RHS in (27)), we notice that the estimated welfare gap has negative sign, yet the estimated share is 23%, indicating that the inequality (27) is likely violated.

To this end, we consider a moment inequality approach to constructing an LCB for the first-best welfare that uses the discrete covariate $Educ$ and remains valid even if some groups have zero treatment effect. To implement the procedure from Remark 4, it remains to specify a policy class. We expect the treatment effect to be non-increasing in education level, with possible jumps at graduation years, $Educ = 12$ and $Educ = 16$. Thus, we consider policy classes of the form $\{Educ \leq C\}$ where the cutoff level C belongs to a finite set.

Table 2 presents the results. The rows correspond to different sets of cutoffs. In Rows 1–2, the cutoff sets are determined by high school and college graduation years.[3] Row 3 takes a more granular set of cutoffs corresponding to all twelve observed education levels. Column (1) inverts a least-favorable test which treats all moment inequalities as binding, and Column (2) incorporates the Generalized Moment Selection procedure of Andrews and Soares (2010).

Our findings are as follows. First, the optimal policy, $\{Educ \leq 15\}$, attains a welfare gain of 1440.25 USD, which exceeds all point estimates in Table 1. Second, the LCBs based on Generalized Moment Selection exceed $LCB_{\mathcal{X}}$ and, therefore, all other LCBs in Table 1. Thus, we find that using the existing methods from the moment inequality literature allows to obtain confidence bands that are robust to the non-uniqueness of the optimal policy and

[3] The policy $\{Educ \leq 11\}$ corresponds to treating only those who did not graduate from high-school (37.3% of the sample); $\{Educ \leq 12\}$ adds those who graduated from high-school but did not attend college (80.0% of the sample); $\{Educ \leq 15\}$ adds those who attended but did not graduate from college (95.9% of the sample); $\{Educ \leq 16\}$ adds college graduates (98.7% of the sample). The policy $\{Educ \leq 18\}$ corresponds to treating everyone and is included in all policy classes.

address the trade-off between the welfare gap and precision.

5 Conclusion

In this article, we addressed the question of reporting a lower confidence band on optimal welfare in policy learning problems. First, we showed that allowing for a first-order bias can be beneficial for constructing a one-sided confidence band as long as the direction of the bias is known and matches the direction of the confidence band. This observation distinguishes one-sided inferential objectives from two-sided ones where the bias must typically vanish at a faster rate. Focusing on the first-best welfare, we connect this observation to the failure of uniformity of Tsybakov (2004)'s margin assumption. We demonstrate how existing inference methods for moment inequalities can be useful for addressing the trade-off between the welfare gap and precision in a way that is robust to the failure of margin assumption.

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

Treatment Rule	(1) Share of People to be Treated	(2) Welfare Gain (s.e.)	(3) 95% LCB	(4) Welfare Gap (USD)	(5) LCB Gap (%)
Treat Everyone	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 (32). Row (1): ATE as in (34). Rows (2)–(3): welfare gain based on plug-in rule estimated via series regression and random forest. Row (4): sample-split welfare gain based on a plug-in rule estimated via random forest. Column (3): 95% LCB defined as $W_{gain} - 1.645 \cdot s.e.(W_{gain})$. Column (4): welfare gap $W_{gain} - ATE$, where $ATE = 1289.66$ (Row 1) and W_{gain} is in subsequent rows $j \in \{2,3,4\}$. Column (5): relative LCB gap $100(1 - LCB_j/LCB_{ATE})\%$ where $LCB_{ATE} = 717.52$ (Row 1) and LCB_j is in rows $j \in \{2,3,4\}$. The sample ($N = 9{,}223$) is the same as in Kitagawa and Tetenov (2018b). See text for further details.

Table 2: 95% LCB for Welfare Gain: Moment Inequality Approach

Treatment Rules	(1) Least-Favorable	(2) Generalized Moment Selection
Cutoff $\in \{11, 15, 18\}$	783.28	783.28
Cutoff $\in \{11, 12, 15, 16, 18\}$	758.25	758.25
Cutoff $\in \{7, 8, \dots, 18\}$	649.53	724.26

Notes. Table reports $LCB_{\mathcal{G}}$ in (30) with distinct choices of policy classes $\mathcal{G}$ in Rows 1–3 and moment selection approaches in Columns 1–2. The classes of treatment rules take the form $\mathcal{G} = \{Educ \leq C : C \in \mathcal{C}\}$ with a set of cutoffs listed above. The policy $\{Educ \leq 18\}$ corresponds to treating everyone. The generalized moment selection procedure is from Andrews and Soares (2010). The critical values $\widehat{c}_{1-\alpha}$ are based on a Gaussian approximation with 10^5 simulation draws. The sample ($N = 9{,}223$) is the same as in Kitagawa and Tetenov (2018b). See text for further details.

A Proofs

Section A.1 contains auxiliary statements. The proof of Theorem 1(1) is given in Section A.3. The proof of Theorem 1(2) is given in Section A.2. Section A.4 contains the proof of Theorem 2.

A.1 Auxiliary statements

Lemma A.1 (Efficiency bound for the first-best welfare, Luedtke and van der Laan (2016), Theorem 1)**.** *Suppose Assumptions 2.1 and 2.2(1)-(3) hold. Then, the the first-best welfare* $\mathbb{E}[\max(m(1,X), m(0,X))]$ *is pathwise differentiable with efficient score*

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

Lemma A.2 (Efficiency bound for W_G)**.** *Suppose Assumption 2.1 and 2.2 (1)-(2) hold. Let*

G be a known policy. Then, the welfare W_G is pathwise differentiable with efficient score

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

Lemma A.2 gives an efficient score for W_G of a known policy G. 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. Plugging $G = G^*$ in (5) recovers the efficient score $\psi(W)$ in (A.1), that is,

$$\psi_{G^*}(W) = \psi^*(W).$$

Thus, the efficiency bound is the same as if the first-best policy G^* were known.

A.2 Proof of Theorem 1(2).

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

A.A.1 The Class of DGPs

It suffices to specify the marginal distribution of X, the distribution of D given X, and the distribution of $Y(1) \mid X$ and $Y(0) \mid X$. We take X as a single binary covariate

$$P(X=1) = p, \quad P(X=0) = 1-p, \quad p \in (0,1).$$

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

$$Y(1) \mid X=1 \sim F(1/2-\epsilon, \sigma^2(1,1)) \tag{A.3}$$
$$Y(1) \mid X=0 \sim F(1/2, \sigma^2(1,0)), \tag{A.4}$$
$$Y(0) \mid X=1 \sim F(1/2, \sigma^2(0,1)) \tag{A.5}$$
$$Y(0) \mid X=0 \sim F(0, \sigma^2(0,0)) \tag{A.6}$$

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 $\epsilon \in (0, 1/2)$. The conditional average treatment effect reduces to

$$\tau(1) = -\epsilon < 0, \quad \tau(0) = 1/2 > 0.$$

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

$$G^* = \{X = 0\}. \tag{A.7}$$

The regression estimators in (7) and (8) take the form of standard cell-specific sample averages. For $d, x \in \{1, 0\}$, define sample counts

$$N_{dx} = \sum_{i=1}^{N} 1\{D_i = d\}1\{X_i = x\}. \tag{A.8}$$

Since the support of X has two points, the regression estimators are based on standard cell-specific sample averages

$$\widehat{m}_{dx} = \widehat{m}(d,x) = \frac{\sum_{i=1}^{N} 1\{D_i = d\}1\{X_i = x\}Y_i}{N_{dx} + 1}, \tag{A.9}$$

where one is added throughout to prevent division by zero[4]. Plugging $\widehat{m}_{dx}$ in (A.9) and G^* in (A.7) gives

$$\begin{aligned}\widehat{W}_{G^*} &= N^{-1} \sum_{i=1}^{N} \widehat{m}(1, X_i)1\{X_i = 0\} + N^{-1} \sum_{i=1}^{N} \widehat{m}(0, X_i)1\{X_i = 1\} \\ &= \widehat{m}_{10}(1 - \widehat{p}) + \widehat{m}_{01}\widehat{p} \end{aligned} \tag{A.10}$$

where the Bernoulli parameter p is estimated as $\widehat{p} = \sum_{i=1}^{N} X_i/N$. A similar argument gives

$$\widehat{W}_{\mathcal{X}} = \widehat{m}_{11}\widehat{p} + \widehat{m}_{10}(1 - \widehat{p}). \tag{A.11}$$

[4]Inflating the denominators of $\widehat{W}_{G^*}$ and $\widehat{W}_{\mathcal{X}}$ by one prevents 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 a $O(N^{-1/2})$ scale, the bias is negligible for a sufficiently large sample. An alternative option is to work with standard (unadjusted) 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.

For the policies $G \in \{G^*, \mathcal{X}\}$, define estimated efficiency bound as

$$\widehat{\sigma}_G^2 := N^{-1} \sum_{i=1}^{N} (\widehat{\psi}_G(W_i) - \widehat{W}_G)^2,$$

where $\psi_G(W)$ is given in (A.2), the regression estimates are in (A.9), and the propensity score $\pi(1), \pi(0)$ is taken as known.

A.A.2 Two Key Lemmas

Lemma A.3 shows that the welfare gap vanishes as $\epsilon = o(1)$ while efficiency bounds remain strictly separated. For the proof of Theorem 1(2), we consider the class of DGPs with the conditional variances

$$\sigma^2(1{,}1) = 1, \quad \sigma^2(1{,}0) = 1, \quad \sigma^2(0{,}1) = 10.$$

Lemma A.3 (Separated efficiency bounds)**.** *(1) The statement* (10) *holds with*

$$W_{\mathcal{X}} = 1/2 - \epsilon p, \quad \sigma_{\mathcal{X}}^2 = \frac{p}{\pi(1)} + \frac{(1-p)}{\pi(0)} + \epsilon^2 (1-p)p. \tag{A.12}$$

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

$$W_{G^*} = 1/2, \quad \sigma_{G^*}^2 = \frac{1-p}{\pi(0)} + \frac{10p}{1-\pi(1)} \tag{A.13}$$

which implies $W_{G^*} - W_{\mathcal{X}} = \epsilon p$.
(3) The efficiency bounds are separated as

$$\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2 > 6p. \tag{A.14}$$

(4) The standard deviations are separated as

$$\sigma_{G^*} - \sigma_{\mathcal{X}} > p/2. \tag{A.15}$$

Proof of Lemma A.3. The proof has three steps. Steps 1 and 2 establish (A.12) and (A.13). Step 3 gives a lower bound on the variance gap.

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

$$\sigma_{\mathcal{X}}^2 = \mathbb{E}\Big[(m(1,X) - W_{\mathcal{X}})^2 + \frac{\mathbb{V}ar(Y \mid D=1, X)}{\pi(X)}\Big]. \tag{A.16}$$

The first summand is

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

The second summand is

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

Adding two summands gives $\sigma_{\mathcal{X}}^2$ in (A.12).

Step 2. The efficiency bound of W_{G^*} is the variance of $\psi^*(W)$ in (A.1). It has two summands where the first one is $\mathbb{V}ar(\max(m(1,X), m(0,X))) = \mathbb{V}ar(1/2) = 0$. The second one is

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

which coincides with $\sigma_{G^*}^2$ in (A.13).

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 $\epsilon^2(1-p) < 1$ gives (A.14). Furthermore, $\sigma_{G^*}^2 \leq 44$ and $\sigma_{\mathcal{X}}^2 \leq 5$. Invoking the lower bound (A.14) gives

$$\sigma_{G^*} - \sigma_{\mathcal{X}} = \frac{\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2}{\sigma_{G^*} + \sigma_{\mathcal{X}}} \geq \frac{6p}{\sqrt{5} + \sqrt{44}} > p/2.$$

■

Lemma A.4 completes the proof of Theorem 1(2).

Lemma A.4 (Ranking of LCBs)**.** *The LCB ranking (17) holds for any $\epsilon \in (0, (z_{1-\alpha}/4)N^{-1/2})$ with the constant $C = z_{1-\alpha}/12$.*

Proof of Lemma A.4. **Step 1.** Decompose exact expected length into the leading term $\Delta_{\mathcal{X}}$ in (15) and the remainders

$$\mathbb{E}[LCB_{\mathcal{X}} - LCB_{G^*}] = \Delta_{\mathcal{X}} + R_1 + R_2,$$

where R_1 and R_2 are defined below. Invoking (A.15) gives $\Delta_{\mathcal{X}} > N^{-1/2}z_{1-\alpha}/4p$ for $\epsilon \in (0, (z_{1-\alpha}/4)N^{-1/2})$.

Step 2. Define the remainder terms as

$$\begin{aligned} R_1 &= \mathbb{E}[\widehat{W}_{\mathcal{X}} - W_{\mathcal{X}}] - \mathbb{E}[\widehat{W}_{G^*} - W_{G^*}] \\ R_2 &= (N^{-1/2}z_{1-\alpha}(\mathbb{E}[\widehat{\sigma}_{G^*} - \sigma_{G^*}] - \mathbb{E}[\widehat{\sigma}_{\mathcal{X}} - \sigma_{\mathcal{X}}])). \end{aligned}$$

By Assumption 2.2(1)-(2), the estimators $\widehat{\sigma}^2_{\mathcal{X}}$ and $\widehat{\sigma}^2_{G^*}$ are a.s. bounded and consistent uniformly over the class of DGPs. For these estimators, the convergence in probability implies convergence of moments, and

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

which implies $|\mathbb{E}[R_2]| \leq \Delta_{\mathcal{X}}/3$ for N large enough.

Step 3. Let $(\mathbf{X}, \mathbf{D}) = (X_i, D_i)_{i=1}^N$ be stacked realizations of $(X_i)_{i=1}^N$ and $(D_i)_{i=1}^N$. The conditional mean is

$$\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}] = (1/2 - \epsilon\widehat{p}) - R, \tag{A.17}$$

where the remainder term

$$R := (1/2 - \epsilon)\widehat{p}(N_{11} + 1)^{-1} + 1/2(1 - \widehat{p})(N_{10} + 1)^{-1} \tag{A.18}$$

is non-negative a.s. for $\epsilon \in (0, 1/2)$. Furthermore, since $\mathbb{E}[\widehat{p}] = p$, the first summand is unbiased for $W_{\mathcal{X}} = 1/2 - \epsilon p$. The remainder is bounded as

$$0 \leq \mathbb{E}[R] \leq^{i} 1/2\mathbb{E}[(N_{11} + 1)^{-1}] + 1/2\mathbb{E}[(N_{10} + 1)^{-1}] =^{ii} O(N^{-1}),$$

where (i) follows from the monotonicity of expectation and $\widehat{p} \leq 1$a.s. and (ii) from the standard property of binomial distribution stated in (A.19). A similar argument applies to $\widehat{W}_{G^*}$. Thus, $|\mathbb{E}[R_1]| < \Delta_{\mathcal{X}}/3$ for a sufficiently large sample size.

■

A.3 Proof of Theorem 1(1).

Notation and Preliminaries. Consider a class of DGPs defined in Section A.2.

Note that $N_{dx} \sim Binom(N, \rho)$ where $\rho = \Pr(D = d \mid X = x)\Pr(X = x)$. For any

$\mathcal{Z} \sim Binom(N, \rho)$, for $N \geq 1$, the following standard properties hold:

$$\mathbb{E}[(\mathcal{Z}+1)^{-1}] = \frac{1}{(N+1)\rho} - \frac{1}{(N+1)\rho}(1-\rho)^{N+1} \leq N^{-1}\rho^{-1}$$
$$\mathbb{E}[(\mathcal{Z}+1)^{-2}] \leq 2\rho^{-2}N^{-2}.$$

We focus on symmetric DGPs with $p = \pi(1) = \pi(0) = 1/2$ so that $\rho = 1/4$ for all pairs (d,x). Then

$$\mathbb{E}[(N_{dx}+1)^{-1}] = \frac{4}{(N+1)} - \frac{4}{(N+1)}(3/4)^{N+1} \leq 4/N \tag{A.19}$$
$$\mathbb{E}[(N_{dx}+1)^{-2}] \leq 32N^{-2}. \tag{A.20}$$

Structure of the proof. The Proof of Theorem 1(1) is established in four steps. Lemma A.5 bounds the approximation error of expected conditional variance. Lemma A.6 establishes a lower bound for $MSE(\widehat{W}_{G^*})$. Lemma A.7 establishes an upper bound for $MSE(\widehat{W}_{\mathcal{X}})$. Lemma A.8 completes the proof.

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

$$(1-10C_N) < \mathbb{E}[\sigma^{-2}(d,1)N \cdot \mathbb{V}ar(\widehat{m}_{d1} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2] < (1+10C_N). \tag{A.21}$$
$$(1-10C_N) < \mathbb{E}[\sigma^{-2}(d,0)N \cdot \mathbb{V}ar(\widehat{m}_{d0} \mid \mathbf{X}, \mathbf{D})(1-\widehat{p})^2] < (1+10C_N). \tag{A.22}$$

Proof of Lemma A.5. **Step 1 (Notation).** Denote the expression inside the expectation of (A.21) as

$$\Xi_d := \sigma^{-2}(d,1)N\mathbb{V}ar(\widehat{m}_{d1} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 = NN_{d1}(N_{d1}+1)^{-2}\widehat{p}^2.$$

whose probability limit reduces to 1. Define the error terms

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

and decompose its asymptotic error

$$\begin{aligned}
\Xi_d - 1 &= NN_{d1}(N_{d1}+1)^{-2}\widehat{p}^2 - 1 \\
&= \psi_d^1(\widehat{p}^2) - \psi_d^2(\widehat{p}^2) - 1 \\
&= \psi_d^1(\widehat{p}^2 - p^2) + \psi_d^1(p^2) - \psi_d^2(\widehat{p}^2) - 1 \\
&= \underbrace{(\psi_d^1(\widehat{p}^2 - p^2))}_{S_2} + \underbrace{(\psi_d^1(p^2) - N/(N+1))}_{S_1} - \underbrace{\psi_d^2(\widehat{p}^2)}_{S_3} - \underbrace{1/(N+1)}_{S_4}.
\end{aligned} \tag{A.23}$$

Step 2. (Leading term 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. On this event, the term S_2 is upper bounded

$$\begin{aligned}
|\mathbb{E}[S_2 1\{\mathcal{M}_N\}]| &\leq \mathbb{E}[\psi_d^1(|\widehat{p}^2 - p^2|)|1\{\mathcal{M}_N\}] \\
&\leq \mathbb{E}[\psi_d^1(1.5C_N)1\{\mathcal{M}_N\}] \\
&\leq 1.5C_N\mathbb{E}[\psi_d^1(1)] \leq^{(i)} (1.5)\cdot 4\cdot C_N = 6C_N,
\end{aligned} \tag{A.24}$$

where the first lines from linearity of $\psi_d^1(\cdot)$ and monotonicity of expectation and (i) follows from (A.19). On the complementary event $\mathcal{M}_N^c$, the error is still bounded $|\widehat{p}^2 - p^2| \leq 1$ a.s. and

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

where (i) follows from $N_{d1} \geq 0$ a.s. and $(N_{d1}+1)^{-1} \leq 1$a.s.
Invoking Chernoff bound for Binomial distribution and $C_N = \sqrt{2.25 \ln N/N}$ gives

$$NP(\mathcal{M}_N^c) \leq 2N \exp^{-2C_N^2 N/3} \leq 2N^{-1/2} \leq C_N, \qquad \forall N \geq 6. \tag{A.25}$$

Adding (A.24) and (A.25) gives $|\mathbb{E}[S_2]| \leq 7C_N$.

Step 3. (Other terms S_1, S_3, S_4) A simple argument gives $S_4 = (N+1)^{-1} \leq C_N$. Invoking (A.19) gives

$$|\mathbb{E}[S_1]| \leq 1/4|\mathbb{E}[\psi_d^1(1) - 4N/(N+1)]| = (N/4)(4/(N+1))(3/4)^{N+1} \leq C_N, \quad \forall N \geq 2.$$

Invoking (A.20) gives

$$0 \leq \mathbb{E}[S_3] = \mathbb{E}[\psi_d^2(\widehat{p}^2)] \leq \mathbb{E}[\psi_d^2(1)] \leq 32N^{-1} \leq C_N, \quad \forall N \geq 100.$$

Combining the bounds gives

$$|\mathbb{E}[\Xi_d - 1]| = |\mathbb{E}[\sigma^{-2}(d,1)N \cdot \mathbb{V}ar(\widehat{m}_{d1} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 - 1]| \leq \sum_{j=1}^{4} |\mathbb{E}[S_d]| \leq 10C_N.$$

Step 4. (Conclusion). Steps 1–3 established (A.21), which corresponds to $x = 1$. The symmetry of DGPs implies (A.22) with $x = 0$. ■

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

$$N \cdot MSE(\widehat{W}_{G^*}) > (\sigma^2(1,0) + \sigma^2(0,1))(1 - 10C_N). \tag{A.26}$$

Proof of Lemma A.6. **Step 1.** Let $(\mathbf{X}, \mathbf{D}) = (X_i, D_i)_{i=1}^N$ be stacked realizations of $(X_i)_{i=1}^N$ and $(D_i)_{i=1}^N$. For any $i,j \in \{1,2,\ldots,N\}$, we show that

$$X_i(1 - X_j)\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D}) = 0 \text{ a.s.}$$

If the indices are distinct, $\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D}) = 0$ by independence of the samples i and j. If the indices coincide, the product $X_i(1 - X_j) = X_i(1 - X_i) = 0$ a.s. Noting that $\text{Cov}(\widehat{m}_{d_1 1}, \widehat{m}_{d_2 0} \mid \mathbf{X}, \mathbf{D})$ consists of N^2 summands involving $X_i(1 - X_j)\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D}) = 0$ a.s., we have

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

Thus, the variance of each estimator is

$$\mathbb{V}ar(\widehat{W}_{G^*} \mid \mathbf{X}, \mathbf{D}) = \mathbb{V}ar(\widehat{m}_{01} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 + \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2 \tag{A.27}$$

$$\mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}) = \mathbb{V}ar(\widehat{m}_{11} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 + \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2. \tag{A.28}$$

Step 2. Invoking Lemma A.5 with parameters $(d,x) = (0,1)$ and $(d,x) = (1,0)$ gives a lower bound

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{01} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2] > \sigma^2(0,1)(1 - 10C_N) \tag{A.29}$$

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2] > \sigma^2(1,0)(1 - 10C_N) \tag{A.30}$$

Adding (A.29) and (A.30) gives a lower bound on $\mathbb{E}[\mathbb{V}ar(\widehat{W}_{G^*} \mid \mathbf{X}, \mathbf{D})]$. A lower bound (A.26) on $MSE(\widehat{W}_{G^*})$ follows.

■

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*

$$N \cdot MSE(\widehat{W}_{\mathcal{X}}) < \sigma^2_{\mathcal{X}} + \frac{N\epsilon^2}{2} + (4 + 10(\sigma^2(1,1) + \sigma^2(1,0)))C_N. \tag{A.31}$$

Proof of Lemma A.7. **Step 1.** **(Bias.)** Let R be the remainder term as defined in (A.18).

The bias is bounded from above and below

$$0 \leq W_{G^*} - \mathbb{E}[\widehat{W}_{\mathcal{X}}] = W_{G^*} - W_{\mathcal{X}} + W_{\mathcal{X}} - \mathbb{E}[\widehat{W}_{\mathcal{X}}] \tag{A.32}$$
$$\leq \epsilon/2 + 4N^{-1}. \tag{A.33}$$

as shown in Step 4 of the proof of Lemma A.4.

Step 2. (Variance.) We show that variance is upper bounded by

$$N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}}) < \sigma^2_{\mathcal{X}} + (\sigma^2(1{,}1) + \sigma^2(1{,}0))10C_N + 2C_N. \tag{A.34}$$

The variance of the conditional mean is

$$\mathbb{V}ar(\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}]) = \mathbb{V}ar(R) - 2\mathrm{Cov}(R, 1/2 - \epsilon\widehat{p}) + \frac{\epsilon^2}{4N}. \tag{A.35}$$

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) gives

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D})] \leq (\sigma^2(1{,}1) + \sigma^2(1{,}0))(1 + 10C_N) \tag{A.36}$$

Adding (A.35) and (A.36) gives

$$N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}}) = N \cdot \mathbb{V}ar(\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}]) + \mathbb{E}[N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D})]$$
$$< \sigma^2_{\mathcal{X}} + (\sigma^2(1{,}1) + \sigma^2(1{,}0))10C_N + 2C_N.$$

Step 3. (MSE.) Combining (A.33) and(A.34) gives (A.31) since $16N^{-1} \leq C_N$ for all $N \geq 100$. ■

Lemma A.8 (MSE Ranking)**.** *For any $\epsilon \in (0, N^{-1/2})$ and N large enough, MSE ranking (16) 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 $MSE(\widehat{W}_{G^*})$

$$N \cdot MSE(\widehat{W}_{G^*}) > (\sigma^2(1,0) + \sigma^2(0,1))(1 - 10C_N).$$

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

$$N \cdot MSE(\widehat{W}_{\mathcal{X}}) < 3/4 + (\sigma^2(1,1) + \sigma^2(1,0)) + [4 + 10(\sigma^2(1,1) + \sigma^2(1,0))]C_N \tag{A.37}$$

Therefore, when $\sigma^2(0,1) - \sigma^2(1,1) - 3/4 > 0$, there exists N_0 that depends on conditional variances such that

$$N \cdot MSE(\widehat{W}_{G^*}) - N \cdot MSE(\widehat{W}_{\mathcal{X}}) > 0.$$

■

Remark A.1 (On a finite-sample result)**.** When the conditional variances are taken as

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

the statement of Lemma A.8 holds for any $N \geq N_0$ where the cutoff $N_0 = 1,745$.

A.4 Proof of Theorem 2.

Let $G^* \triangle G$ denote the symmetric difference of sets G^* and G, i.e., $G^* \triangle G = G^* \backslash G \cup G \backslash G^*$. Let $P(X \in G^* \triangle G)$ 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.

Lemma A.9. *Suppose Assumptions 2.1, 2.2, 3.1 hold. Then, (1) the welfare gap is bounded*

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

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

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

(3) the standard deviation gap is bounded

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

Proof of Lemma A.9. **Step 1.** The bound (A.38) is established in (Tsybakov, 2004) (cf., Lemma A.7 in (Kitagawa and Tetenov, 2018a)).

Step 2. 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

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

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

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

The final term is bounded as

$$|T_{3G}| \leq |(W_{G^*} - W_G)||W_{G^*} + W_G| \leq^i M^2 P(X \in G^* \triangle G).$$

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

Step 4. An upper bound on the standard deviation gap is

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

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

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

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 (25) with a constant

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

■

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 (26).

Step 1. We take $X \sim U[0,1]$. The propensity score equals 1/2 a.s. in X Let $\epsilon \in (0, 1/2)$ and ν be an odd integer such that $\delta \geq 1/\nu$. We focus on distributions with $M/2$-a.s. bounded outcomes whose 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

$$\tau(x) = m(1,x) - m(0,x) = (x-\epsilon)^\nu.$$

The first-best policy reduces to

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

For any $\epsilon \in (0, 1/2)$, this distribution automatically satisfies Assumptions 2.2(1)-(2).

Step 2. We show that the proposed sequence of DGPs satisfies Assumption 3.1 uniformly with $\delta \geq 1/\nu$ and $\eta = 2^\nu$, which suffices to prove that belongs to $\mathcal{P}(M, \kappa, 2^\nu, \delta, \underline{\sigma})$. Note that $\epsilon^\nu \leq (1-\epsilon)^\nu$. For small values of $t \in (0, \epsilon^\nu)$, we have

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

Likewise, for large values of $t \geq \epsilon^\nu$,

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

Since for odd digits $P(|X-\epsilon|^\nu \leq t) = P(|X - \epsilon| \leq t^{1/\nu}) \leq 2t^{1/\nu}$, the restriction (21) holds for any δ larger than $1/\nu$ and $\eta = 2^\nu$ for any value of $\epsilon \in (0, 1/2)$.

Step 3. We prove the lower bound (26). 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 $T_{2\mathcal{X}} = \epsilon^{2\nu+1}/(2\nu+1)$ and $T_{3\mathcal{X}} = -\epsilon^{2\nu+2}/(\nu+1)^2$. The variance gap is

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

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 (26) 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

B.1 Bounds on Welfare Gap

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. 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 provides an upper bound on the *Jensen gap* for a convex yet non-smooth function $f(t) = \max(t,0)$, for which the standard bounds in the literature (for differentiable functions) do not apply (e.g Abramovich and Persson, 2016). Second, the inequality gives a sufficient condition for the welfare gap to be small: if the treatment effect heterogeneity vanishes as $\mathbb{V}ar(\tau(X)) = o(N^{-1})$, the welfare gap decays faster than the parametric rate.

In fact, the following conditional version of the result holds.

Lemma B.2 (Conditional upper bound on welfare gap)**.** *Let $T : \mathcal{X} \rightarrow \mathcal{T}$ be any measurable transformation of X, and W_{T^*} denote the first-best welfare in the restricted class of treatment* *policies that depend on X only through T. Then:*

$$0 \leqslant W_{G^*} - W_{T^*} \leqslant \mathbb{E}[\sqrt{\mathbb{V}ar(\tau(X) \,|\, T)}].$$

Proof. Let $\tau(X) = \mathbb{E}[Y(1) - Y(0) \,|\, X]$ and $\bar{\tau}(T) = \mathbb{E}[Y(1) - Y(0) \,|\, T]$, and note that $\mathbb{E}[\tau(X) \,|\, T] = \bar{\tau}(T)$, by the Law of Iterated Expectations. The restricted first-best welfare W_{T^*} can be written as $W_{T^*} = \mathbb{E}[\max(\bar{\tau}(T), 0)] + \mathbb{E}[Y(0)]$. Therefore,

$$\begin{aligned}
W_{G^*} - W_{T^*} &= \mathbb{E}[\max(\tau(X), 0) - \max(\bar{\tau}(T), 0)] \\
&\overset{(a)}{\leqslant} \mathbb{E}[\max(\tau(X) - \mathbb{E}[\tau(X) \,|\, T], 0)] \\
&\overset{(b)}{\leqslant} \mathbb{E}\left[\mathbb{E}\left[|\tau(X) - \mathbb{E}[\tau(X) \,|\, T]| \,\middle|\, T\right]\right] \\
&\overset{(c)}{\leqslant} \mathbb{E}\left[\mathbb{E}\left[(\tau(X) - \mathbb{E}[\tau(X) \,|\, T])^2 \,\middle|\, T\right]^{1/2}\right] \\
&= \mathbb{E}[\sqrt{\mathbb{V}ar(\tau(X) \,|\, T)}],
\end{aligned}$$

where (a) holds since $\max(a, 0) - \max(b, 0) \leqslant \max(a - b, 0)$, (b) holds due $\max(x, 0) \leqslant |x|$ and the law of iterated expectations, and (c) follows from Jensen's inequality. ■

B.2 Inverting Moment Inequality Tests

This section provides closed-form expression for lower confidence bands based on inverting tests for moment inequalities. The null hypothesis is $H_0 : W_G \leqslant \theta \quad \forall G \in \mathcal{G}$. We assume that the estimators $\widehat{W}_G$ of the form (23) satisfy:

$$(\sqrt{N}(\widehat{W}_G - W_G))_{G \in \mathcal{G}} \quad \Rightarrow^d \quad N(0, \Sigma).$$

Further, suppose the asymptotic covariance matrix can be consistently estimated by $\widehat{\Sigma}_N$, and, setting $\widehat{D}_N = \operatorname{diag}((\widehat{\Sigma}_{N,GG})^{1/2}_{G \in \mathcal{G}})$, let $\widehat{\Omega}_N = \widehat{D}_N^{-1} \widehat{\Sigma}_N D_N^{-1}$, denote the corresponding correlation matrix. We focus on the maximum test statistic:

$$T_N(\theta) = \max_{G \in \mathcal{G}} \frac{\sqrt{N}(\widehat{W}_G - \theta)}{\widehat{\sigma}_G}.$$

The tests take the form:

$$\phi_N(\theta) = \mathbf{1}(T_N(\theta) > \widehat{c}_\alpha(\theta)),$$

where $\widehat{c}_\alpha(\theta)$ is the estimated $(1-\alpha)$ quantile of:

$$\max_{G\in\mathcal{G}}\left(\frac{\sqrt{n}(\widehat{W}_G - W_G)}{\widehat{\sigma}_G} + \frac{s_G(\theta)}{\widehat{\sigma}_N}\right),$$

where $s_G(\theta)$ is the moment selection function, representing a possibly data-dependent upper bound on the quantity $\sqrt{N}(W_G-\theta)$. We consider the least-favorable test:

$$s_G^{LF}(\theta) = 0$$

and the Generalized Moment Selection Test of Andrews and Soares (2010):

$$s_G^{GMS}(\theta) = \begin{cases} 0 & \frac{\sqrt{N}(\widehat{W}_G-\theta)}{\widehat{\sigma}_G} > -\kappa_N \\ -\infty & \text{otherwise} \end{cases}$$

with some sequence $0 < \kappa_N \to \infty$ such that $\kappa_N/\sqrt{N} \to 0$. As discussed in Canay and Shaikh (2017), under standard regularity conditions, the quantile $\widehat{c}_{1-\alpha}(\theta)$ can be estimated using the normal approximation for $(\frac{\sqrt{N}(\widehat{W}_G-W_G)}{\widehat{\sigma}_G})_{G\in\mathcal{G}} \approx N(0,\widehat{\Omega}_N)$. The corresponding LCB can be obtained by test inversion:

$$LCB_{\mathcal{G}} = \{\theta : \phi_N(\theta) = 0\}.$$

The following two lemmas give closed-form expression for the lower confidence bands.

Lemma B.3 (LCB based on LF test inversion)**.** *The lower confidence band obtained by inverting the least-favorable test takes the form:*

$$LCB_{\mathcal{G}}^{LF} = \max_{G\in\mathcal{G}}\left(\widehat{W}_G - \widehat{c}_\alpha\frac{\widehat{\sigma}_G}{\sqrt{N}}\right),$$

where $\widehat{c}_\alpha$ is the $(1-\alpha)$ quantile of the distribution of $\max_{G\in\mathcal{G}}(Z_G)$ with $(Z_G)_{G\in\mathcal{G}} \sim N(0,\widehat{\Omega}_N)$, conditional on the data.

Proof. The critical value of the least-favorable test does not depend on θ. Thus, the set of all values of θ for which the test does not reject is:

$$\max_{G\in\mathcal{G}}\frac{\sqrt{N}(\widehat{W}_G-\theta)}{\widehat{\sigma}_G} \leqslant \widehat{c}_\alpha.$$

Solving for θ gives $LCB_{\mathcal{G}}^{LF}$ above. ■

The following lemma gives a closed-form solution for the lower confidence band based on inverting the Generalized Moment Selection test of Andrews and Soares (2010). Since the critical value of the GMS test is a step function, and the test statistic is a maximum of a finite number of linear functions, the confidence region obtained by test inversion may not be convex (although it can be shown that the probability of such an event approaches zero as N increases). So, in the statement below, we define $LCB_{\mathcal{G}}^{GMS}$ as the lowest point of the confidence set obtained by test inversion.

Lemma B.4 (LCB based on GMS test inversion)**.** *Let $t^{(j)}$ denote the j-th largest value among $\widehat{W}_G + \widehat{\sigma}_G \frac{\kappa_N}{\sqrt{N}}$, with $t^{(|\mathcal{G}|+1)} = -\infty$; let $I^{(j)}$ denote the set of policies G corresponding to $t^{(1)}, \ldots, t^{(j)}$, and $\widehat{c}_\alpha^{(j)}$ denote the $(1-\alpha)$ quantile of $\max_{G \in I^{(j)}}(Z_G)$, where $(Z_G)_{G\in\mathcal{G}} \sim N(0, \widehat{\Omega}_N)$, conditional on the data. Denote:*

$$\theta^{(j)} = \max_{G\in\mathcal{G}} \left(\widehat{W}_G - \widehat{c}_\alpha^{(j)} \frac{\widehat{\sigma}_G}{\sqrt{N}} \right).$$

The lower confidence band obtained by inverting the GMS test takes the form:

$$LCB_{\mathcal{G}}^{GMS} = \min\{\theta^{(j)} : t^{(j)} \geqslant \theta^{(j)} > t^{(j+1)}\}. \tag{B.3}$$

Proof. Under the GMS procedure, by definition, the critical value $\widehat{c}_\alpha(\theta)$ takes the form of a step function:

$$\widehat{c}_\alpha(\theta) = \sum_{j=1}^{\mathcal{G}} \widehat{c}_\alpha^{(j)} \mathbf{1}(t^{(j)} \geqslant \theta > t^{(j+1)}).$$

The function $T_N(\theta)$ is a maximum of a finite number of linear functions of θ. The LCB corresponds to the lowest point of intersection between $T_N(\theta)$ and $\widehat{c}_N(\theta)$ (since the latter is a step function, there can be multiple such points). Each point $\theta^{(j)}$ marks the intersection of $T_N(\theta)$ with a constant function $\widehat{c}_\alpha^{(j)}$. If such $\theta^{(j)}$ is within the relevant "step" $[t^{(j)}, t^{(j+1)})$ of the critical value $\widehat{c}_\alpha(\theta)$, it is one of the intersection points of $T_N(\theta)$ and $\hat{c}_\alpha(\theta)$. The minimum in the expression for $LCB_{\mathcal{G}}^{GMS}$ selects the lower point of intresection. ■

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. The welfare gain parameter is estimated using doubly-robust moment condition

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

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 $\widehat{\tau}_\ell$ be estimators of CATE constructed using all observations not in I_ℓ. Define the cross-fit estimate of welfare gain as

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

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

where the standard debiased inference is established under Assumption 3.1 with some positive parameters and $o(N^{-1/4})$ rate conditions on $\widehat{\tau}_\ell$ (see, 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 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 defined as

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

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.

Andrews, D. W. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Athey, S., N. Keleher, and J. Spiess (2024). Machine learning who to nudge: Causal vs predictive targeting in a field experiment on student financial aid renewal.

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

Bai, Y., A. Santos, and A. M. Shaikh (2022). A two-step method for testing many moment inequalities. *Journal of Business & Economic Statistics 40*(3), 1070–1080.

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.

Canay, I. A. and A. M. Shaikh (2017). Practical and theoretical advances in inference for partially identified models. *Advances in Economics and Econometrics 2*, 271–306.

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

Chernozhukov, V., D. Chetverikov, and K. Kato (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies 86*(5), 1867–1900.

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

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), 30–67.

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.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014). A practical two-step method for testing moment inequalities. *Econometrica 82*(5), 1979–2002.

Sasaki, Y. and T. Ura (2024). Welfare analysis with marginal treatment effects. *Econometric Theory*.

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.

Viviano, D. (2024). Policy targeting under network interference. *Review of Economic Studies*, rdae041.

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