

On the Lower Confidence Band for the Optimal Welfare in Policy Learning

Kirill Ponomarev* Vira Semenova†

September 12, 2025

Abstract

We study inference on the optimal welfare in a policy learning problem and propose reporting a lower confidence band (LCB). A natural approach to constructing an LCB is to invert a one-sided t -test based on an efficient estimator for the optimal welfare. However, we show that for an empirically relevant class of DGPs, such an LCB can be first-order dominated by an LCB based on a welfare estimate for a suitable suboptimal treatment policy. We show that such first-order dominance is possible if and only if the optimal treatment policy is not “well-separated” from the rest, in the sense of the commonly imposed margin condition. When this condition fails, standard debiased inference methods are not applicable. We show that uniformly valid and easy-to-compute LCBs can be constructed analytically by inverting moment-inequality tests with the maximum and quasi-likelihood-ratio test statistics. As an empirical illustration, we revisit the National JTPA study and find that the proposed LCBs achieve reliable coverage and competitive length.

Keywords: Statistical decision theory, policy learning, optimal welfare, lower confidence band, partial identification, sensitivity analysis, cross-fitting, uniformity

JEL Numbers: C14, C31, C54

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

†Department of Economics, University of California, Berkeley. Email: vsemenova@berkeley.edu.

First version: October 2024, <https://arxiv.org/abs/2410.07443>. We are grateful to Alberto Abadie, Timothy Armstrong, Isaiah Andrews, Jiafeng Chen, Victor Chernozhukov, Timothy Christensen, Bryan Graham, Michael Jansson, Patrick Kline, Sokbae Lee, Elena Manresa, Demian Pouzo, Andres Santos, Azeem Shaikh, Liyang Sun, Vasilis Syrgkanis, 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

Treatment assignment problems are ubiquitous in economics, including governments providing subsidies to disadvantaged households, firms offering job training opportunities to their employees, colleges allocating scholarships to students, and online retailers offering discounts to customers. In such settings, a decision-maker (DM) aims to design a treatment rule that determines who should — and who should not — be treated, based on observable individual characteristics, to maximize welfare (Manski, 2004). Since developing good treatment rules may be costly and time-consuming, the DM might want to quantify the potential welfare gains. To this end, the DM may conduct a preliminary experiment and test a hypothesis that the optimal welfare (or welfare gain) exceeds a certain threshold.

Conducting inference for the optimal welfare (and welfare gain) is a challenging task. From a practical perspective, it may require solving complicated non-convex optimization problems, estimating functions of high-dimensional inputs non-parametrically, and dealing with noisy welfare estimates due to suboptimal experiment design. Theoretically, a major complication is the potential non-uniqueness of the optimal policy, which makes standard debiased inference methods inapplicable (Hirano and Porter, 2012; Luedtke and van der Laan, 2016).

In this paper, we show that good estimators and tight lower confidence bands (LCBs) for the optimal welfare (and welfare gain) can be obtained by leveraging suboptimal policies. Our first contribution is to demonstrate a possible trade-off between the welfare level and the precision with which it can be estimated in finite samples. For empirically relevant data-generating processes (DGPs), we provide an example of a slightly suboptimal policy, whose welfare can be estimated substantially more precisely than the optimal one. As a result, an LCB targeting such suboptimal welfare can be first-order tighter — at the $N^{-1/2}$ scale for sample size N — than the LCB targeting the optimal welfare directly. Additionally, such suboptimal policy yields a better estimator of the optimal welfare in terms of mean-squared error, for all N large enough. In particular, this example shows that incorporating asymptotically redundant information can yield first-order improvements for estimators and inference procedures in finite samples.

Our second contribution is to characterize the class of DGPs for which the first-order trade-off between welfare and precision is possible. Intuitively, if the optimal policy is “well-separated” from the rest, the precision gain of any suboptimal policy cannot compensate for the welfare loss. We formalize this intuition using a local asymptotic approximation around a DGP at which “separation” fails, and derive minimax rates for the gap between the two LCBs. As a result, we show that the first-order trade-off is possible if and only if the margin

condition of [Mammen and Tsybakov \(1999\)](#) and [Tsybakov \(2004\)](#) fails to hold uniformly over the relevant DGPs. In such settings, standard debiased inference procedures may be invalid, so alternative inference methods are needed.

To this end, we propose LCBs that address the aforementioned welfare-precision tradeoff and remain valid regardless of the margin condition. The idea is to construct a (possibly large but) finite subclass of test policies, based on economic intuition, within which a “good” suboptimal policy may be found. Each of these policies provides a lower bound on the optimal welfare, yielding a collection of moment inequalities that can be tested using existing methods ([Andrews and Soares, 2010](#); [Chernozhukov, Lee, and Rosen, 2013](#); [Romano, Shaikh, and Wolf, 2014](#); [Canay and Shaikh, 2017](#)). The existing tests combine self-normalization (precision correction in [Chernozhukov et al., 2013](#)) with moment selection, leading to tight LCBs that remain valid under relatively weak conditions. For the problem at hand, the tests can often be inverted analytically, so the LCBs are easy to compute in practice.

To illustrate our theoretical results, we revisit the U.S. National Job Training Partnership Act (JTPA) experiment [Bloom, Orr, Bell, Cave, Doolittle, Lin, and Bos \(1997\)](#). The experiment randomly assigned individuals with distinct education levels and baseline earnings to a job training program and recorded their post-treatment salary. For most education years — apart from graduation thresholds — the respective conditional average treatment effect is statistically insignificant, indicating a violation of the margin condition. Standard procedures that either ignore education or use a holdout sample to estimate the first-best policy suffer from substantial power loss. We consider several classes of test policies based only on education and construct the corresponding LCBs by inverting moment-inequality tests as described above. In line with the theoretical predictions, the LCBs are substantially shorter than the available alternatives.

Related Literature This paper contributes to a large cross-disciplinary literature on optimal treatment choice, following [Manski \(2004\)](#). In econometrics, contributions range from early program-evaluation and partial-identification approaches to modern policy learning ([Dehejia, 2005](#); [Hirano and Porter, 2009](#); [Stoye, 2009](#); [Chamberlain, 2011](#); [Bhattacharya and Dupas, 2012](#); [Tetenov, 2012](#); [Rai, 2019](#); [Kitagawa and Tetenov, 2018b](#); [Mbakop and Tabord-Meehan, 2021](#); [Athey and Wager, 2021](#); [Sun, 2021](#); [Sasaki and Ura, 2024](#); [Kitagawa, Lee, and Qiu, 2022](#); [Yata, 2021](#); [Armstrong and Shen, 2023](#); [Chernozhukov, Lee, Rosen, and Sun, 2025](#); [Moon, 2025](#)). In statistics, optimal treatment regimes are commonly learned via Q-learning and A-learning ([Qian and Murphy, 2011](#); [Murphy, 2003](#); [Robins, 2004](#); [Shi, Fan, Song, and Lu, 2018](#)). This literature focuses primarily on obtaining treatment rules that perform well in terms of expected regret.

In this paper, we consider a complementary problem of inference on the optimal welfare,

also studied in [Luedtke and van der Laan \(2016\)](#).¹ In the absence of ties among the best policies, the authors showed that the optimal welfare is a regular parameter and derived the semiparametric efficiency bound for it. The bound turns out to be the same as if the best policy was known *ex ante*. When ties are present, the optimal welfare is no longer regular ([Hirano and Porter, 2012](#)), but in view of the above, an oracle efficient estimator based on one of the optimal policies still provides a natural benchmark for our analysis. We complement the results of [Luedtke and van der Laan \(2016\)](#) by studying MSEs of the estimators and expected length of the associated LCBs in finite samples, formalizing the necessity of the margin condition for one-sided inference, and proposing simple robust inference procedures. The proposed procedures provide alternatives to the approaches based on smoothing, as in [Chen, Austern, and Syrgkanis \(2023\)](#), [Levis, Bonvini, Zeng, Keele, and Kennedy \(2023\)](#) and [Whitehouse, Austern, and Syrgkanis \(2025\)](#), or entropic regularization, as in [Ben-Michael \(2025\)](#). They also relate to a broader literature on robust policy learning, including decisions under ambiguity ([Ben-Michael, Greiner, Imai, and Jiang, 2021](#); [Cui and Han, 2024](#)) and concerns about external validity ([Adjaho and Christensen, 2022](#)). Although we focus on the utilitarian (linear) formulation of welfare throughout, the proposed approach also applies in non-linear settings, such as inequality-sensitive welfare studied in [Kasy \(2016\)](#); [Kitagawa and Tetenov \(2021\)](#); [Terschuur \(2025\)](#), among others.

This paper also contributes to the literature on inference for partially identified parameters. We show that in finite samples, inference based on sharp bounds may be less precise than inference based on loose bounds, giving rise to a first-order trade-off between sharpness and precision. We argue that existing inference methods are able to address this trade-off by combining self-normalization (precision-correction) and moment selection, while retaining uniform validity ([Andrews and Soares, 2010](#); [Chernozhukov et al., 2013](#); [Romano et al., 2014](#); [Canay and Shaikh, 2017](#); [Bai, Santos, and Shaikh, 2022](#)).²

The rest of the paper is organized as follows. Section 2 introduces the policy learning problem and motivates our target parameters. Section 3 gives a sequence of DGPs exhibiting the first-order dominance. Section 4 discusses the role of the margin assumption. Section 5 proposes robust inference procedures. Section 6 contains an empirical application. Section 7 concludes. Appendix A contains proofs. Appendix B contains auxiliary theoretical results. Appendix C contains auxiliary empirical details.

¹This problem is distinct from the “inference on winners” considered in [Andrews, Kitagawa, and McCloskey \(2024\)](#), [Andrews and Chen \(2025\)](#), and [Chernozhukov et al. \(2025\)](#), and the proposed LCBs are generally not valid in those settings.

²A related question of inference with over-identifying inequality constraints is studied, e.g., in [Cox \(2024\)](#) and [Ketzi and McCloskey \(2025\)](#). Our setting is different in that the target parameter may not be asymptotically Gaussian even when the constraints are not binding.

2 Setup

2.1 Policy Learning Problem

Consider a population of individuals characterized by their potential outcomes in treated and untreated states, $Y(1), Y(0) \in \mathcal{Y} \subseteq \mathbf{R}$, and characteristics $X \in \mathcal{X} \subseteq \mathbf{R}^{d_x}$. A decision-maker (DM) aims to maximize the average welfare in the population by subjecting some individuals to treatment, depending on their observable characteristics X . That is, the DM chooses a treatment rule $G \in \mathcal{G} \subseteq 2^{\mathcal{X}}$ to maximize

$$W_G = \mathbb{E}[Y(1)\mathbf{1}\{X \in G\} + Y(0)\mathbf{1}\{X \in G^c\}], \quad (2.1)$$

where $G^c = \mathcal{X} \setminus G$ denotes the complement of G . The class of feasible treatment rules \mathcal{G} may be *ex ante* restricted for institutional reasons, such as transparency or non-discrimination in treatment, or practical reasons, such as computation and implementation.

We assume that the DM has access to experimental data that identifies W_G . The observable data vector $Z = (D, Y, X)$ contains the assigned treatment $D \in \{0, 1\}$, realized outcome $Y \in \mathcal{Y}$, and covariates $X \in \mathcal{X}$, so that $Y = DY(1) + (1 - D)Y(0)$ and $D \perp (Y(1), Y(0)) \mid X$. The propensity score will be denoted by $\pi(x) = P(D = 1 \mid X = x)$. The conditional mean and variance functions of the potential outcomes are non-parametrically identified as $m(d, x) = \mathbb{E}[Y(d) \mid X = x] = \mathbb{E}[Y \mid D = d, X = x]$ and $\sigma^2(d, x) = \mathbb{V}ar(Y(d) \mid X = x) = \mathbb{V}ar(Y \mid D = d, X = x)$, for $d \in \{0, 1\}$, and the conditional average treatment effect (CATE) function as $\tau(x) = m(1, x) - m(0, x)$. As a result, the average welfare function is identified as $W_G = \mathbb{E}[m(0, X) + \mathbf{1}\{X \in G\}\tau(X)]$ and can be non-parametrically estimated. To this end, the DM observes a random sample $(Z_i)_{i=1}^N$ distributed i.i.d. $Z_i \sim P \in \mathbf{P}$, for a class of distributions \mathbf{P} specified below.

The objects of interest throughout the paper are the maximum (or first-best, or optimal) welfare, denoted by

$$W_{G^*} = \max_{G \in \mathcal{G}} W_G, \quad (2.2)$$

where G^* denotes any policy attaining the maximum,³ and the corresponding welfare gain,

$$W_{G^*}^{\text{gain}} = W_{G^*} - W_{\emptyset}, \quad (2.3)$$

which is non-negative as long as the policy class \mathcal{G} includes the status quo policy \emptyset of not treating anyone.

³For simplicity, we assume that the maximum is well-defined.

2.2 Lower Confidence Bands

In many settings, the DM would naturally be interested in lower confidence bands (LCBs) for the maximum welfare or the corresponding welfare gain. For example, consider a firm deciding whether to build a job-training center. Suppose the firm maximizes the net welfare subject to a “safety” constraint that the risk of false adoption (i.e., incurring negative welfare) must be below level α , for some $\alpha \in (0, 1)$. This leads to testing

$$H_0 : W_{G^*} \leq 0 \quad \text{vs} \quad H_1 : W_{G^*} > 0.$$

In such settings, LCBs are natural inputs to threshold decision rules (see, e.g., Section 3.5 of [Lehmann and Romano, 2005](#)).

As another example, consider an online retailer deciding whether to offer a discount for a certain type of good to its customers. The retailer may first run a small-scale randomized experiment to explore whether any discount rule can lead to increase in profits. This corresponds to testing

$$H_0 : W_{G^*}^{\text{gain}} = 0 \quad \text{vs} \quad H_1 : W_{G^*}^{\text{gain}} > 0,$$

which is equivalent to comparing a $100(1 - \alpha)\%$ LCB for $W_{G^*}^{\text{gain}}$ with zero.

The main input in the construction of LCBs is an estimator for the welfare function W_G . For each policy G , we can express $W_G = \mathbb{E}[\psi_G(Z)]$, where

$$\begin{aligned} \psi_G(Z) = & \left(m(1, X) + \frac{D}{\pi(X)}(Y - m(1, X)) \right) \mathbf{1}\{X \in G\} \\ & + \left(m(0, X) + \frac{1 - D}{1 - \pi(X)}(Y - m(0, X)) \right) \mathbf{1}\{X \in G^c\} \end{aligned} \quad (2.4)$$

is the efficient, doubly robust, moment function ([Robins and Rotnitzky, 1995](#); [Hahn, 1998](#)). For suitable first-stage estimators $\hat{m}(d, x)$ and $\hat{\pi}(x)$, a regular semiparametrically efficient estimator \widehat{W}_G can be constructed using cross-fitting, so that

$$\sqrt{N}(\widehat{W}_G - W_G) \Rightarrow^d \mathcal{N}(0, \sigma_G^2), \quad (2.5)$$

where $\sigma_G^2 = \text{Var}(\psi_G(Z))$. Given a significance level $\alpha \in (0, 1)$, a $100(1 - \alpha)\%$ LCB for W_G can be formed as

$$\widehat{LCB}_G = \widehat{W}_G - N^{-1/2} z_{1-\alpha} \widehat{\sigma}_G, \quad (2.6)$$

where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of $\mathcal{N}(0, 1)$ and $\widehat{\sigma}_G$ is a consistent estimator of the asymptotic standard deviation σ_G .

Since $W_G \leq W_{G^*}$, for any $G \in \mathcal{G}$, an LCB based on any suboptimal policy $G \in \mathcal{G}$ provides valid one-sided coverage for the optimal welfare,

$$P(\widehat{LCB}_G \leq W_{G^*}) \geq P(\widehat{LCB}_G \leq W_G) \rightarrow 1 - \alpha, \quad \text{as } N \rightarrow \infty. \quad (2.7)$$

As a result, \widehat{LCB}_G can be meaningfully compared across distinct policies. As an ideal benchmark, we consider an LCB based on an infeasible efficient estimator of the welfare under a first-best policy,

$$\widehat{LCB}_{G^*} = \widehat{W}_{G^*} - N^{-1/2} z_{1-\alpha} \widehat{\sigma}_{G^*}. \quad (2.8)$$

As discussed in the introduction, such LCB is a valid reference point even when the optimal policy is not unique. Since \widehat{LCB}_{G^*} is based on an efficient estimator for W_{G^*} and the standard deviation is rescaled by $N^{-1/2}$, one might expect that \widehat{LCB}_{G^*} always be preferred to \widehat{LCB}_G in large samples, for any suboptimal policy G . We show, however, that this is not the case. Given the direction of the intended comparison, considering an oracle LCB as a benchmark only strengthens our point. Of course, our recommended inference procedures in Section 5 account for the first-best policy being unknown.

2.3 Asymptotic Criterion for LCB ranking

To compare the candidate LCBs, we consider the *LCB gap*, defined as

$$\Delta_G = N^{-1/2} z_{1-\alpha} (\sigma_{G^*} - \sigma_G) - (W_{G^*} - W_G). \quad (2.9)$$

A positive sign of Δ_G indicates that the policy G is nearly optimal yet the corresponding welfare is substantially more precisely estimated. Consequently, the corresponding LCB_G may be preferred to LCB_{G^*} in large samples.

The motivation for studying LCB gap comes from a local asymptotic approximation along smooth parametric sub-models, standard in the semi-parametric efficiency theory. To elaborate, let \mathbf{P} denote the class of all admissible distributions of the data. Consider a distribution $P_0 \in \mathbf{P}$ such that $W_{G^*(P_0)} = W_G$ for $G \neq G^*(P_0)$. Let $T(P_0)$ denote the tangent space at P_0 ,⁴ and $P_{N,h} = P_{1/\sqrt{N},h}$, for $h \in T(P_0)$, be a sequence of distributions following a

⁴See [Hahn \(1998\)](#) for the derivation of $T(P)$ in the present setting.

smooth parametric submodel $\{t \mapsto P_{t,h}\} \subseteq \mathbf{P}$. Denote

$$\begin{aligned}\mu(h) &= \sqrt{N}(W_{G^*(P_{N,h})} - W_G); \\ s(h) &= \sigma_{G^*(P_{N,h})} - \sigma_G,\end{aligned}$$

where the dependence of $\mu(h)$ and $s(h)$ on N is suppressed for notational convenience, and note that

$$\Delta_G(P_{N,h}) = N^{-1/2}(z_{1-\alpha}s(h) - \mu(h)).$$

The assumed regularity of \widehat{W}_G , consistency of $\widehat{\sigma}_G$, and contiguity of $P_{N,h}$ with respect to P_0 imply that, under $P_{N,h}$,

$$\sqrt{N}(\widehat{LCB}_G - \widehat{LCB}_{G^*(P_{N,h})}) \Rightarrow_d \mathcal{N}(z_{1-\alpha}s(h) - \mu(h), \sigma_\Delta^2(P_0)),$$

for some $\sigma_\Delta^2(P_0) \geq 0$. That is, the distribution of $\sqrt{N}(\widehat{LCB}_G - \widehat{LCB}_{G^*(P_{N,h})})$ under any sequence of “perturbations” $P_{N,h}$ of P_0 , is determined by $z_{1-\alpha}s(h) - \mu(h) = \sqrt{N}\Delta_G(P_{N,h})$. Moreover, under further regularity conditions,

$$\mathbb{E}[\widehat{LCB}_G - \widehat{LCB}_{G^*(P_{N,h})}] = \Delta_G(P_{N,h}) + o(1),$$

so the LCB gap can be interpreted as a large-sample analog to the difference of expected LCBs.⁵ For these reasons, we consider the LCB gap in the formal results below.

3 First-Order Dominance

In this section, we give an example of a model in which the welfare-precision trade-off is of the first order, and discuss the implications of this phenomenon. We focus on welfare throughout, but similar considerations apply to welfare gain. See Remark 1 for the details.

3.1 The Data Generating Process

First, we specify a suitable DGP for $(Y(1), Y(0), D, X)$. It suffices to specify the marginal distribution of X , the propensity score, and the conditional distributions of $Y(1) \mid X$ and

⁵An ideal way to rank LCBs is in terms of the first-order dominance; See [Lehmann \(1959\)](#). Unfortunately, since distinct policies typically result in LCBs with distinct large-sample variances, this criterion does not apply in a Gaussian limit. A natural alternative is to compare LCBs in terms of their expected values, as suggested, e.g., in [Harter \(1964\)](#). While the exact expectations may not exist without further restrictions or be distorted by the biases in first-stage estimators, their large sample analogs remain tractable.

$Y(0) \mid X$. Let X be a binary covariate distributed as

$$P(X = 1) = p; \quad P(X = 0) = 1 - p, \quad \text{for some } p \in (1/4, 3/4).$$

Denote the propensity score by

$$P(D = 1 \mid X = 1) = \pi(1); \quad P(D = 1 \mid X = 0) = \pi(0), \quad \text{for some } \pi(1), \pi(0) \in (1/4, 3/4).$$

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

$$Y(1) \mid X = 1 \sim F\left(\frac{1}{2} - \epsilon, 1\right); \quad Y(1) \mid X = 0 \sim F\left(\frac{1}{2}, 1\right); \quad (3.1)$$

$$Y(0) \mid X = 1 \sim F\left(\frac{1}{2}, 10\right); \quad Y(0) \mid X = 0 \sim F\left(\frac{1}{2} - \epsilon, 10\right), \quad (3.2)$$

where $\epsilon \in (0, 1/2)$ is a vanishing sequence to be specified. Since we focus on the average welfare, the joint distribution of $(Y(1), Y(0)) \mid X$ is immaterial, so we leave it unspecified⁶

Simple algebra shows that the CATE function takes the form

$$\tau(1) = -\epsilon < 0; \quad \tau(0) = \epsilon > 0,$$

the unique first-best policy is

$$G^* = \{0\}, \quad (3.3)$$

and the corresponding welfare is

$$W_{G^*} = 1/2 \cdot p + 1/2 \cdot (1 - p) = 1/2. \quad (3.4)$$

In addition, consider the “treat everyone” policy, $G = \mathcal{X}$ whose welfare is

$$W_{\mathcal{X}} = (1/2 - \epsilon) \cdot p + 1/2 \cdot (1 - p) = 1/2 - \epsilon p. \quad (3.5)$$

Note that the welfare gap between the two policies scales linearly with ϵ

$$0 \leq W_{G^*} - W_{\mathcal{X}} \leq \epsilon p. \quad (3.6)$$

⁶With variance parameters $\sigma^2(1, 1) = 1/4$, $\sigma^2(1, 0) = 1$, $\sigma^2(0, 1) = 200$, $\sigma^2(0, 0) = 1$, the statement holds for all sample sizes exceeding 1745. For the variances in the main text, the minimal cutoff sample size N is approximately 6000.

while the standard deviation gap does not depend on ϵ ,

$$\sigma_{G^*} - \sigma_{\mathcal{X}} > p. \quad (3.7)$$

3.2 Estimators and Lower Confidence Bands

Since X is binary, the average welfare under any fixed policy G can be efficiently estimated using the regression-adjusted estimator. For each $(d, x) \in \{0, 1\}^2$, denote

$$N_{dx} = \sum_{i=1}^N \mathbf{1}\{D_i = d\} \mathbf{1}\{X_i = x\}, \quad (3.8)$$

and define the estimators

$$\begin{aligned} \hat{\pi}(x) &= \frac{N_{1x}}{N_{1x} + N_{0x}}; \\ \hat{m}(d, x) &= \frac{\sum_{i=1}^N Y_i \cdot \mathbf{1}\{D_i = d\} \mathbf{1}\{X_i = x\}}{N_{dx} + 1}, \end{aligned} \quad (3.9)$$

where one is added to the denominator throughout to prevent division by zero.⁷ Recalling from (3.3) that $G^* = \{0\}$, the first-best welfare is estimated as

$$\widehat{W}_{G^*} = \hat{m}(0, 1) \cdot \hat{p} + \hat{m}(1, 0) \cdot (1 - \hat{p}), \quad (3.10)$$

where $\hat{p} = \sum_{i=1}^N X_i / N$. Similarly,

$$\widehat{W}_{\mathcal{X}} = \hat{m}(1, 1) \cdot \hat{p} + \hat{m}(1, 0) \cdot (1 - \hat{p}). \quad (3.11)$$

The mean squared errors of the two estimators, with respect to W_G^* , are given by

$$\begin{aligned} MSE(\widehat{W}_{\mathcal{X}}) &= \mathbb{E}[(\widehat{W}_{\mathcal{X}} - W_{G^*}^*)^2]; \\ MSE(\widehat{W}_{G^*}) &= \mathbb{E}[(\widehat{W}_{G^*} - W_{G^*}^*)^2]. \end{aligned} \quad (3.12)$$

The asymptotic variances of \widehat{W}_G , for $G \in \{G^*, \mathcal{X}\}$, can be estimated as

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

⁷This step introduces bias of order $O(N^{-1})$ which is negligible for a sufficiently large sample. An alternative is to work with unadjusted denominators on the event where both of them are strictly positive.

where $\widehat{\psi}_G(Z_i)$ is obtained by plugging the estimated propensity score and regression functions from (3.9) in (2.4). The corresponding LCBs are obtained as

$$\widehat{LCB}_{\mathcal{X}} = \widehat{W}_{\mathcal{X}} - N^{-1/2} z_{1-\alpha} \widehat{\sigma}_{\mathcal{X}}, \quad (3.13)$$

$$\widehat{LCB}_{G^*} = \widehat{W}_{G^*} - N^{-1/2} z_{1-\alpha} \widehat{\sigma}_{G^*}. \quad (3.14)$$

Following the discussion of Section 2, we compare $LCB_{\mathcal{X}}$ and LCB_{G^*} in terms of LCB gap

$$\Delta_{\mathcal{X}} = \frac{z_{1-\alpha}}{\sqrt{N}} (\sigma_{G^*} - \sigma_{\mathcal{X}}) - (W_{G^*} - W_{\mathcal{X}}). \quad (3.15)$$

3.3 First-Order Dominance

Our first main result shows that $\widehat{W}_{\mathcal{X}}$ dominates \widehat{W}_{G^*} in terms of MSE, and the respective LCB gap is positive.

Proposition 1 (First-Order Dominance). *For all N large enough, for the DGP (3.2) and estimators (3.10) and (3.11), the following statements hold:*

1. Both MSEs in (3.12) are finite and

$$MSE(\widehat{W}_{\mathcal{X}}) < MSE(\widehat{W}_{G^*}); \quad (3.16)$$

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

$$\Delta_{\mathcal{X}} > C_{\alpha} N^{-1/2}. \quad (3.17)$$

Proposition 1 makes three points. First, the trade-off between welfare and precision may be first-order. As a result, suboptimal policies may yield better point estimates and tighter, on average, lower confidence bands for the optimal welfare. That is, the first-best policy — the policy that is best to implement — may differ from the policy whose estimated welfare is best to report.⁸ Similar observations apply to inference on partially-identified parameters, as we further discuss in Remark 2.

Second, there is a distinction between the two-sided and one-sided inferential objectives. In the two-sided case, the bias typically must vanish faster than the standard deviation to ensure valid coverage of the confidence intervals. In the one-sided case, coverage remains

⁸DGPs with treatment effects vanishing at the $N^{-1/2}$ rate have been employed to obtain a meaningful limiting experiment (Hirano and Porter, 2009) or establish minimax rates for expected regret (Kitagawa and Tetenov, 2018b; Athey and Wager, 2021). In this paper, we use DGPs with similar conditional means and carefully chosen variances to establish a lower bound on the LCB gap.

valid as long as the direction of the bias matches the direction of the confidence band, which allows bias and variance to be potentially of the same order. Proposition 1 gives a concrete, empirically relevant example of this distinction⁹.

Third, efficiency arguments in near non-regular settings may be problematic. For each $\epsilon > 0$, the oracle efficient estimator \widehat{W}_{G^*} attains the semiparametric efficiency bound (Luedtke and van der Laan, 2016), but in the limit, $\epsilon = 0$, the optimal welfare is a non-regular parameter, and semiparametric efficiency bounds do not apply (Hirano and Porter, 2012). Proposition 1 demonstrates that, for distributions within a $N^{-1/2}$ -neighborhood of $\epsilon = 0$ (excluding zero), $\widehat{W}_{\mathcal{X}}$ dominates \widehat{W}_{G^*} in terms of MSE, for all N large enough. Thus, the familiar notion of efficiency fails not only at the point of non-regularity, but already in a $N^{-1/2}$ -neighborhood around it.

Remark 1 (Implications for welfare gain). The above example could be modified to obtain a first-order dominance statement for the welfare gain in (2.3). Consider the DGPs

$$Y(1) \mid X = 1 \sim F\left(\frac{1}{2} - \epsilon, 1\right), \quad Y(1) \mid X = 0 \sim F\left(\frac{1}{2}, 10\right), \quad (3.18)$$

$$Y(0) \mid X = 1 \sim F\left(\frac{1}{2}, 1\right), \quad Y(0) \mid X = 0 \sim F\left(\frac{1}{2} - \epsilon, 10\right). \quad (3.19)$$

where asymptotic variance is small for $X = 1$ and large for $X = 0$. Let $G^* = \{0\}$ be the optimal policy and $G = \{1\}$ be the suboptimal policy. Simple algebra shows that the welfare gap and variance gap satisfy

$$W_{G^*}^{\text{gain}} - W_G^{\text{gain}} \leq \epsilon, \quad (3.20)$$

$$(\sigma_{G^*}^{\text{gain}})^2 - (\sigma_G^{\text{gain}})^2 > 7. \quad (3.21)$$

As a result, an analog of (3.17) holds for the LCB gap for welfare gain. ■

Remark 2 (Redundant moment inequalities). The above discussion applies to inference for partially identified parameters. For example, consider the setting of Section 2 with binary potential outcomes and unconditional treatment exogeneity, i.e. $(Y(1), Y(0), X) \perp D$. The share of “always-takers”, $\theta = P(Y(1) = Y(0) = 1)$, can be bounded from above by either $\delta_1 = P(Y = 1 \mid D = 0)$ or $\delta_2 = \mathbb{E}[\min(P(Y = 1 \mid D = 1, X), P(Y = 1 \mid D = 0, X))]$. By Jensen’s inequality, δ_2 gives a tighter bound than δ_1 . A $100(1 - \alpha)\%$ Upper Confidence Band

⁹The one-sided dominance result echoes findings in one-sided nonparametric inference: in adaptive tests and multiscale procedures, directed smoothing bias can be exploited to lower variance while preserving size (Dumbgen and Spokoiny, 2001; Armstrong, 2015). Our setting differs in the target parameter (optimal welfare rather than a function at a point) and mechanism (policy-induced bias $W_{G^*} - W_G$ rather than smoothing bias).

(UCB) for θ can be formed using either of the two bounds

$$\widehat{UCB}_j = \widehat{\delta}_j + N^{-1/2} z_{1-\alpha} \widehat{\sigma}_j,$$

where $\widehat{\sigma}_j$ are consistent estimators of the asymptotic standard deviations σ_j of $\widehat{\delta}_j$, for $j = 1, 2$. Similar to Proposition 1, there exist DGPs such that

$$\delta_2 + N^{-1/2} z_{1-\alpha} \sigma_2 > \delta_1 + N^{-1/2} z_{1-\alpha} \sigma_1, \quad (3.22)$$

for all N large enough. As a result, a UCB based on a non-sharp bound first-order dominates its sharp counterpart in terms of the average length. In other words, inference based on a sharp bound may be less informative in finite samples. ■

4 Margin Condition and Higher-Order Dominance

Next, we investigate whether the conclusions of Proposition 1 carry over when the model is restricted by the following additional assumptions.

Assumption 4.1 (Regularity). *(i) The propensity score $\pi(x)$ satisfies $\kappa < \pi(x) < 1 - \kappa$, for almost all $x \in \mathcal{X}$, for some $\kappa \in (0, 1/2)$; (ii) The outcome is bounded so that $P(|Y| \leq M/2) = 1$, for some $M < \infty$.*

Assumption 4.2 (Margin Condition). *For some $\eta \in (0, M)$ and $\delta \in (0, \infty)$,*

$$P(|\tau(X)| < t) \leq (t/\eta)^\delta, \quad \forall t \in [0, \eta]. \quad (4.1)$$

Assumption 4.1 imposes regularity conditions common in the policy learning literature (see, e.g., Kitagawa and Tetenov, 2018b; Mbakop and Tabord-Meehan, 2021). Assumption 4.2 is the margin condition of Tsybakov (2004). In addition to requiring uniqueness of the first-best policy, it controls the intensity with which $\tau(X)$ concentrates in a neighborhood of zero. When the optimal policy is unique, the existence of suitable values of δ and η is a matter of mild regularity conditions. For example, if $|\tau(X)|$ is continuous and has a density bounded at zero, then (4.1) holds for any $\delta < 1$ with η small enough. If $\tau(X)$ has finite support and $P(\tau(X) = 0) = 0$, then (4.1) holds for any $\delta > 0$ and a sufficiently small η .

The sequence of DGPs in Proposition 1 can be chosen to satisfy Assumption 4.1, but it fails to satisfy Assumption 4.2 with uniform lower bounds on η and δ . As we show below, this is precisely what drives the first-order dominance phenomenon. To state the formal

result, we assume that any $G \subseteq \mathcal{X}$ is feasible.¹⁰ Proposition 2 below characterizes the order of magnitude of the worst-case LCB gap Δ_G over all policies $G \subseteq \mathcal{X}$.

Proposition 2 (Higher-Order Dominance). *Let \mathbf{P} denote the class of DGPs obeying Assumptions 4.1–4.2 for some $0 < \underline{\delta} \leq \delta \leq \bar{\delta} < \infty$, $\eta = \eta(\delta) > 0$, and $\inf_{x \in \mathcal{X}, d \in \{1,0\}} \sigma^2(d, x) \geq \underline{\sigma}^2 > 0$. There exist constants $0 < \underline{C} < \bar{C} < \infty$, depending on $(M, \kappa, \underline{\delta}, \bar{\delta}, \underline{\sigma})$, such that*

$$\underline{C}N^{-(1+\underline{\delta})/2} \leq \sup_{P \in \mathbf{P}} \sup_{G \subseteq \mathcal{X}} \Delta_G \leq \bar{C}N^{-(1+\underline{\delta})/2}. \quad (4.2)$$

Proposition 2 demonstrates that once uniform lower bounds on δ and η are imposed, no suboptimal policy G can lead to first-order dominance in the sense of Proposition 1. The smaller the value of δ , the more $\tau(X)$ concentrates near zero, the looser the upper bound in (4.2). In the limit, $\delta = 0$, which corresponds to failure of the margin condition, the lower bound in (4.2) recovers the first-order dominance result (3.17). In the absence of uniform bounds on the margin parameters, Propositions 1 and 2 imply that the first-best welfare may not be the optimal, or relevant, inferential target. The following remarks discuss testable implications of the margin condition and possible testing procedures, as well as further connections with the literature.

Remark 3 (Testing uniqueness 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

$$H_0 : \exists j : \tau(j) = 0 \quad (4.3)$$

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 4 (Testing the margin assumption). In the general case where both discrete and continuous covariates are present, Assumption 4.2 is no longer equivalent to uniqueness of the optimal policy. We describe a testable implication that we find empirically relevant in

¹⁰The upper bound in Proposition 2 holds for all $G \subseteq \mathcal{X}$, so it applies to any restricted class \mathcal{G} as well. The lower bound holds within restricted classes \mathcal{G} as long as they include threshold policies based on each covariate.

Section 6. 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

$$W_{G^*} - W_G \geq C_1 P(G^* \triangle G)^{1+\frac{1}{\delta}} \quad (4.4)$$

for $C_1 = C_1(\delta) = \eta \delta (\frac{1}{1+\delta})^{1+\frac{1}{\delta}} > 0$ (Tsybakov, 2004). Given a lower bound $\underline{\delta} > 0$ and fixing $\eta > 0$, consider a null hypothesis $H_0 : \delta \geq \underline{\delta}$. Since both functions $\delta \rightarrow C_1(\delta)$ and $\delta \rightarrow c^{1+\frac{1}{\delta}}$ are increasing in δ , the lower bound (4.4) on welfare gap implies that, for any policy G ,

$$C_1(\underline{\delta}) P(G^* \triangle G)^{1+\frac{1}{\underline{\delta}}} - (W_{G^*} - W_G) \leq 0. \quad (4.5)$$

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) and Chernozhukov, Newey, and Santos (2015), can then be applied to construct a test. Pursuing this formally is left for future work.¹¹ ■

Remark 5 (Implications for debiased inference). Propositions 1 and 2 imply that sharp bounds may not be optimal, or relevant, inferential targets in the absence of uniform margins, highlighting the tightness of this condition in the context of covariate-assisted bounds; see Kallus, Mao, and Zhou (2020); Kallus (2022b,a); Levis et al. (2023); Semenova (2020, 2023). We expect this insight to imply the tightness of the margin condition in other settings, such as support function analysis (Chandrasekhar, Chernozhukov, Molinari, and Schrimpf, 2012) and algorithmic fairness (Liu and Molinari, 2024), and other policy-relevant metrics. ■

5 Robust Testing Procedures

In this section we discuss testing procedures that address the welfare-precision trade-off and remain valid regardless of the margin assumption. Let $\mathcal{G}_{\text{test}} \subseteq \mathcal{G}$ be a class of policies, which, based on economic intuition, may contain a good lower bound for the optimal welfare. We

¹¹The lower bound in (4.4) plays a role analogous in spirit to the polynomial minorant condition used in partial identification literature, e.g., Condition C.2 in Chernozhukov, Hong, and Tamer (2007), Condition V in Chernozhukov et al. (2013), and Assumption 4.2 in Armstrong (2014). In its general form, this condition relates the difference in the criterion function to the distance metric on the parameter of interest. In policy learning settings, the decision set \mathcal{G} is a collection of partitions of covariate space. In both Chernozhukov et al. (2013) and Kitagawa and Tetenov (2018b), this condition is imposed to tighten convergence guarantees for the proposed estimators. In contrast to prior work, this paper uses the (failure of) margin assumption to motivate the use of suboptimal policies for constructing lower confidence bands for welfare.

look for a LCB of the form

$$\widehat{LCB} = \max_{G \in \mathcal{G}_{\text{test}}} \left\{ \widehat{W}_G - \hat{c}_\alpha \frac{\hat{\sigma}_G}{\sqrt{N}} \right\}, \quad (5.1)$$

where \hat{c}_α is as small as possible to guarantee the desired coverage. We show that such LCB naturally arise from testing moment inequalities, which allows to use a host of existing testing procedures. Our results take the form of finite-sample algebraic identities, so the coverage properties of the resulting LCBs are inherited from validity of the underlying tests. The latter relies only on the uniform CLT-type assumptions and holds regardless of the margin condition. We refer the reader to [Chernozhukov et al. \(2013\)](#) and [Canay and Shaikh \(2017\)](#) for the details.

5.1 Lower Confidence Bands via Testing Moment Inequalities

Suppose $\mathcal{G}_{\text{test}}$ is finite (potentially growing with sample size). Let $\theta = W_{G^*}$ denote the parameter of interest, and consider testing

$$H_\theta : W_G - \theta \leq 0, \text{ for all } G \in \mathcal{G}_{\text{test}}. \quad (5.2)$$

Suppose the estimator $(\widehat{W}_G)_{G \in \mathcal{G}_{\text{test}}}$ for $(W_G)_{G \in \mathcal{G}_{\text{test}}}$ satisfies

$$(\sqrt{N}(\widehat{W}_G - W_G))_{G \in \mathcal{G}} \Rightarrow^d \mathcal{N}(0, \Sigma), \quad (5.3)$$

for a positive definite covariance matrix $\Sigma = (\Sigma_{G_1 G_2})_{G_1, G_2 \in \mathcal{G}}$, and a consistent estimator $\widehat{\Sigma}$ is available. A test for (5.2) can then be constructed as

$$\hat{\phi}_N(\theta) = \mathbf{1} \left(\hat{T}_N(\theta) > \hat{c}_\alpha(\theta) \right), \quad (5.4)$$

with, e.g., the maximum test statistic

$$\hat{T}_N(\theta) = \max_{G \in \mathcal{G}_{\text{test}}} \frac{\sqrt{N}(\widehat{W}_G - \theta)}{\hat{\sigma}_G}, \quad (5.5)$$

where $\hat{\sigma}_G = (\widehat{\Sigma}_{GG})^{1/2}$ and $\hat{c}_\alpha(\theta)$ is suitable a critical value. A common computationally simple choice is the least-favorable critical value, corresponding to

$$\hat{c}_{\alpha, \text{max}}^{\text{LF}} = \hat{Q}_{1-\alpha} \left(\max_{G \in \mathcal{G}_{\text{test}}} \frac{\sqrt{N}(\widehat{W}_G - W_G)}{\hat{\sigma}_G} \right), \quad (5.6)$$

where the quantile can be estimated using bootstrap or Gaussian approximation.

Given the direction of the inequalities in (5.2), the set of all values of θ for which the test in (5.4) does not reject, $\{\theta \in \mathbf{R} : \hat{\phi}_N(\theta) = 0\}$, provides a LCB for W_{G^*} . For the least-favorable critical value, the test compares the value of a partially linear decreasing function of θ with a constant, which allows to obtain a simple closed form for the LCB.

Proposition 3 (LCB by test inversion). *The LCB obtained by inverting a test in (5.4) with the least-favorable critical value in (5.6) is given by*

$$\widehat{LCB}_{\max}^{LF} = \max_{G \in \mathcal{G}_{\text{test}}} \left\{ \widehat{W}_G - \hat{c}_{\alpha, \max}^{LF} \frac{\hat{\sigma}_G}{\sqrt{N}} \right\}. \quad (5.7)$$

Intuitively, the above procedure corresponds to constructing a candidate LCB for W_{G^*} using each suboptimal policy $G \in \mathcal{G}_{\text{test}}$ separately and taking the shortest one, thus explicitly resolving the welfare-precision trade-off. The least-favorable critical value $\hat{c}_{\alpha, \max}^{LF}$ ensures that the resulting LCB has the desired coverage, but it essentially assumes that all of the moment inequalities in (5.2) are binding, which may be too conservative. The critical value can be reduced using moment selection procedures, such as the Generalized Moment Selection (GMS) of Andrews and Soares (2010), or pre-testing, as in Romano et al. (2014). Although both procedures perform well in practice, we focus on GMS because it allows for closed-form test inversion.

The critical value for the GMS procedure is computed as follows. Define the set of inequalities that are “close to binding,”

$$I_N(\theta) = \left\{ G \in \mathcal{G}_{\text{test}} : \frac{\sqrt{N}(\hat{W}_G - \theta)}{\hat{\sigma}_G} > -\kappa_N \right\},$$

where $\kappa_N > 0$ is a sequence of tuning parameters such that $\kappa_N \rightarrow \infty$ and $\kappa_N/\sqrt{N} \rightarrow 0$, for example, $\kappa_N = \sqrt{\log N}$. Then, the GMS critical value is

$$\hat{c}_{\alpha, \max}^{GMS}(\theta) = \hat{Q}_{1-\alpha} \left(\max_{G \in I_N(\theta)} \frac{\sqrt{N}(\hat{W}_G - \theta)}{\hat{\sigma}_G} \right), \quad (5.8)$$

where the quantile can be estimated using bootstrap or Gaussian approximation. As θ increases, the set $I_N(\theta)$ shrinks, so $\hat{c}_{\alpha, \max}^{GMS}(\theta)$ is a decreasing step-function of θ . Thus, the test in (5.4) with the GMS critical value compares a partially linear decreasing function of θ with a step-function. Since there may be multiple intersections, 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. In the statement below, we conservatively define

the LCB starting from the lowest intersection point.

Proposition 4 (LCB by test inversion with GMS). *The LCB obtained by inverting the test in (5.4) with the critical value (5.8) can be computed as follows. For $j \in \{1, \dots, |\mathcal{G}_{\text{test}}|\}$, let $t^{(j)}$ denote the j -th largest value among $\widehat{W}_G + \kappa_N \widehat{\sigma}_G / \sqrt{N}$, and set $t^{(|\mathcal{G}_{\text{test}}|+1)} = -\infty$. Let $I^{(j)} \subseteq \mathcal{G}_{\text{test}}$ collect the policies G corresponding to $t^{(1)}, \dots, t^{(j)}$, and $\widehat{c}_\alpha^{(j)}$ be computed as in (5.8) with $I^{(j)}$ instead of $I_N(\theta)$. Denote $\widehat{\theta}^{(j)} = \max_{G \in \mathcal{G}} (\widehat{W}_G - \widehat{c}_\alpha^{(j)} \widehat{\sigma}_G / \sqrt{N})$. Then,*

$$\widehat{LCB}_{\max}^{\text{GMS}} = \min\{\theta^{(j)} : t^{(j)} \geq \widehat{\theta}^{(j)} > t^{(j+1)}\}. \quad (5.9)$$

The LCB in (5.9) uses a weakly smaller critical value than the LCB in (5.7), so it is always shorter. Yet, the two LCBs are uniformly valid over the same set of distributions. Taken together, self-normalization of the test statistic and a moment selection procedure allow to resolve the welfare-precision trade-off while ensuring that the resulting LCB is robust to violations of the margin condition.

5.2 Lower Confidence Bands via Intersection Bounds

Inference methods for intersection-bounds-type parameters, such as $\max_{G \in \mathcal{G}} W_G$, have been introduced by Chernozhukov et al. (2013) (CLR for short). The authors pointed out that inference based on the plug-in estimator $\max_{G \in \mathcal{G}} \widehat{W}_G$ may be distorted for two reasons: upward bias and large differences in precision of estimates \widehat{W}_G across $G \in \mathcal{G}$. To address these issues, they introduced a “precision corrected” LCB of the form (5.1) and proposed a different moment selection device, tailoring the analysis to an infinite number of intersection parameters (i.e., infinite \mathcal{G}). In what follows, we derive a new duality result between the procedure of CLR and test inversion in the spirit of Section 5.1 and use it to obtain a computationally simpler LCB.

In the preceding section, to find a good lower bound on W_{G^*} , we restricted attention to policies in the test class $\mathcal{G}_{\text{test}} \subseteq \mathcal{G}$. A better lower bound may potentially be obtained by taking convex combinations of $(W_G)_{G \in \mathcal{G}_{\text{test}}}$, which is equivalent to randomizing over $G \in \mathcal{G}_{\text{test}}$. Specifically, let $\Lambda = \{\lambda \in \mathbf{R}_+^{|\mathcal{G}_{\text{test}}|} : \mathbf{1}'\lambda = 1\}$, where $\mathbf{1} = (1, \dots, 1)' \in \mathbf{R}^{|\mathcal{G}_{\text{test}}|}$, denote the probability simplex, $W_{\text{test}} = (W_G)_{G \in \mathcal{G}_{\text{test}}}$ collect the test policies into a finite vector, and $\widehat{W}_{\text{test}} = (\widehat{W}_G)_{G \in \mathcal{G}_{\text{test}}}$ denote the corresponding estimator vector. Each $\lambda \in \Lambda$ yields a lower bound $\lambda' W_{\text{test}} \leq W_{G^*}$ for the optimal welfare. Therefore, following CLR, we look for a LCB of the form

$$\widehat{LCB}_{\text{mix}} = \max_{\lambda \in \Lambda} \left\{ \lambda' \widehat{W}_{\text{test}} - \widehat{c}_\alpha \frac{\sqrt{\lambda' \widehat{\Sigma} \lambda}}{\sqrt{N}} \right\}, \quad (5.10)$$

where the critical value \hat{c}_α is chosen to ensure correct coverage.

A version of CLR's procedure calibrates \hat{c}_α by approximating the supremum of the self-normalized Gaussian process $(\sqrt{N}(\lambda'\widehat{W}_{\text{test}} - \lambda'W_{\text{test}})/(\lambda'\widehat{\Sigma}\lambda)^{1/2})_{\lambda \in \Lambda}$ in simulations, which can be computationally heavy. We replace that step with a finite-dimensional convex program using convex duality. We show that for any vector T and positive definite matrix Σ ,

$$\max_{\lambda \in \Lambda} \frac{\lambda'T}{\sqrt{\lambda'\Sigma\lambda}} = \left(\min_{t \in \mathbf{R}_{-}^{|\mathcal{G}_{\text{test}}|}} (T-t)'\Sigma^{-1}(T-t) \right)^{1/2}, \quad (5.11)$$

Consequently, an LCB of the form (5.10) actually arises from inverting a test for (5.2) using the so-called Quasi-Likelihood-Ratio (QLR) test statistic,

$$\hat{T}_N(\theta) = \min_{t \leq \mathbf{0}} \left(\sqrt{N}(\widehat{W}_{\text{test}} - \theta\mathbf{1}) - t \right)' \widehat{\Sigma}^{-1} \left(\sqrt{N}(\widehat{W}_{\text{test}} - \theta\mathbf{1}) - t \right) \quad (5.12)$$

also considered in Andrews and Soares (2010). The least-favorable critical value,

$$\hat{c}_{\alpha, \text{QLR}}^{LF} = \hat{Q}_{1-\alpha} \left(\min_{t \leq \mathbf{0}} \left(\sqrt{N}(\widehat{W}_{\text{test}} - W_{\text{test}}) - t \right)' \widehat{\Sigma}^{-1} \left(\sqrt{N}(\widehat{W}_{\text{test}} - W_{\text{test}}) - t \right) \right), \quad (5.13)$$

can be estimated using bootstrap or Gaussian approximation and requires solving one convex program per simulation. Our final Proposition summarizes this discussion.

Proposition 5 (LCB by test inversion with QLR). *The LCB obtained by inverting a test in (5.4) with the QLR test statistic (5.12) and least-favorable critical value (5.13) takes the form*

$$\widehat{LCB}_{\text{mix}} = \max_{\lambda \in \Lambda} \left\{ \lambda'\widehat{W}_{\text{test}} - (\hat{c}_{\alpha, \text{QLR}}^{LF})^{1/2} \frac{\sqrt{\lambda'\widehat{\Sigma}\lambda}}{\sqrt{N}} \right\}. \quad (5.14)$$

In practice, $\widehat{LCB}_{\text{max}}^{LF}$ in (5.7) (and its GMS version (5.9)) are computationally simpler and employ a less conservative critical value than $\widehat{LCB}_{\text{mix}}$. However, $\widehat{LCB}_{\text{mix}}$ involves searching over all convex mixtures of test policies which creates more scope to trade off mean welfare against precision. As discussed in Example 4.1 in Canay and Shaikh (2017), both tests are admissible, so the corresponding LCBs cannot generally be ranked. Depending on the underlying DGP, either of the LCBs may be tighter.

6 Empirical Application

To illustrate the welfare-precision trade-off in practice and showcase the proposed procedures, 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. By design, unconditional independence holds,

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

so the first-best welfare and the corresponding welfare gain are identified in each of the models (D, Y) , $(PreEarn, D, Y)$, $(Educ, D, Y)$ and (X, D, Y) . This fact allows us to compare the estimated optimal welfare gains and corresponding LCBs across the models and highlight connections with our theoretical results.

Table 1 presents the estimates and LCBs for the welfare gain based on first-best policy rules in three different policy classes: no covariates (Row 1), only *PreEarn* (Rows 2–3), and both covariates X (Row 4). The welfare gain from treating everyone (Row 1) corresponds to the Average Treatment Effect. It provides a robust lower bound for the optimal welfare gain based on more complex policy classes, so we use it as a reference point. In Rows 2–3, given that *PreEarn* is continuously distributed and the margin assumption is plausible, we adopt the cross-fitted efficient-score estimator. We consider estimating the CATE function of *PreEarn* via series regression (Row 2) and random forest (Row 3). To estimate the propensity score, we bin *PreEarn* into five cells of similar size and use cell-specific averages as an input into the regression adjustment estimator of the form (3.9).

First, in the full model (X, D, Y) , we find that the margin assumption likely fails. For eleven out of twelve education groups, the CATE is not significant at the 5% level, so ties among the first-best treatment rules based on *Educ* are very likely. Although the continuous covariate *PreEarn* may alleviate the concern, violation of margin assumption can still be detected based on the heuristic in Remark 4. The estimated welfare gap (Row 4, Column 4) is negative yet 23% of individuals would be treated differently than under the optimal policy (Row 4, Column 1), so the inequality (4.4) is violated in-sample. To ensure validity of the reported LCB, we do not cross-fit. Using only two-thirds of the sample to compute the point estimate and its 95% LCB incurs substantial efficiency loss, resulting in the lowest LCB in the Table.

Second, in the model $(PreEarn, D, Y)$, we do not detect sufficient heterogeneity to warrant personalized treatment assignment. Comparing Rows 2–3 with Row 1 yields welfare gaps of -246 and -220 , relative to treating everyone. Since the estimated sign is negative,¹² the true gap is likely of the order sampling error. Moreover, the LCB in Row 1 exceeds those in Rows 2–3 by 40% and 28%, respectively. We attribute these findings to potential biases in the first-stage estimators of regression functions and/or lack of heterogeneity in CATE function of $PreEarn$.

Next, we implement the LCBs proposed in Section 5, choosing the test policies based on education level. We expect the treatment effects to be non-increasing in education level, with possible jumps at graduation years, $Educ = 12$ and $Educ = 16$. Thus, we limit the focus on cutoff policies of the form $\{Educ \leq C\}$. In particular, 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); and $\{Educ \leq 18\}$ corresponds to treating everyone.

Table 2 presents LCBs obtained with the maximum test statistic and different test sets $\mathcal{G}_{\text{test}}$ determined by the cutoffs. In Row 1, the cutoff set corresponds to those who attended but did not graduate from high school and college, as well as everyone in the sample; and Row 2 includes all possible cutoffs. The first-best policy in both classes is $\{Educ \leq 15\}$ with the estimated welfare gain of 1440.25 USD, which exceeds all point estimates in Table 1. For the first test class, the least-favorable and GMS confidence bands coincide and exceed all of the LCBs in Table 1. The second test class contains policies that are far from optimal and thus provide loose lower bounds. As a result, the least-favorable test is conservative, while GMS leads to a tighter LCB.

7 Conclusion

In this paper, we addressed the question of reporting a Lower Confidence Band on the optimal welfare in a policy learning problem. First, we documented the trade-off between welfare and precision and showed that it can be first-order. Second, we connected the first-order trade-off to the lack of uniformity in the margin condition of [Mammen and Tsybakov \(1999\)](#); [Tsybakov \(2004\)](#). Finally, we proposed procedures for reporting Lower Confidence Bands that address the trade-off and remain valid regardless of the margin condition.

¹²The estimated sign is negative due to the use of the efficient/doubly robust estimators (2.4), which are not necessarily ordered in-sample.

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

Treatment Rule	Treated Share	Welfare Gain (s.e.)	95% LCB	Welfare Gap (USD)	Relative 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 (<i>PreEarn</i>)	0.92	1069.50 (335.24)	518.06	-220.16	28%
Random Forest (<i>PreEarn</i> + <i>Educ</i>)	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 defined in (2.3). Row 1: Average Treatment Effect; Rows 2–3: Welfare gain based on the policy $G = \mathbf{1}\{CATE(PreEarn) \geq 0\}$, where CATE is estimated via series regression or random forest. Row 4: Sample-split welfare gain based on a plug-in treatment rule estimated via random forest. The 95% LCBs are given by $W_j^{gain} - 1.645s.e.(W_j^{gain})$. The Welfare Gap is $W_j^{gain} - ATE$, where $ATE = 1289.66$ (Row 1) and W_j^{gain} are in Rows $j = 2, 3, 4$. The relative LCB gap is defined as $100(1 - LCB_j/LCB_{ATE})\%$, 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

Cutoffs for Test Policies	Least-Favorable	GMS
Cutoff $\in \{11, 15, 18\}$	783.28	783.28
Cutoff $\in \{7, 8, \dots, 18\}$	649.53	724.26

Notes: The table reports LCBs based on two different test policy classes of the form $\mathcal{G}_{test} = \{Educ \leq C : C \in \mathcal{C}\}$ with a set of cutoffs \mathcal{C} listed above. The policy $\{Educ \leq 18\}$ corresponds to treating everyone. The Generalized Moment Selection (GMS) procedure is from Andrews and Soares (2010). The critical values $\hat{c}_{1-\alpha}$ 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.

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.
- Adjaho, C. and T. Christensen (2022). Externally valid policy choice.
- Andrews, D. W. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78(1), 119–157.
- Andrews, I. and J. Chen (2025). Certified decisions.
- Andrews, I., T. Kitagawa, and A. McCloskey (2024). Inference on winners. *The Quarterly Journal of Economics* 139(1), 305–358.
- Armstrong, T. B. (2014). Weighted ks statistics for inference on conditional moment inequalities. *Journal of Econometrics* 181(2), 92–116.
- Armstrong, T. B. (2015). Adaptive testing on a regression function at a point. *The Annals of Statistics* 43(5), 2086–2101.
- Armstrong, T. B. and S. Shen (2023). Inference on optimal treatment assignment. *The Japanese Economic Review* 74(4), 471–500.
- 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.
- Ben-Michael, E. (2025). Partial identification via conditional linear programs: estimation and policy learning.
- Ben-Michael, E., D. J. Greiner, K. Imai, and Z. Jiang (2021). Safe policy learning through extrapolation: Application to pre-trial risk assessment.
- 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.
- Chandrasekhar, A., V. Chernozhukov, F. Molinari, and P. Schrimpf (2012, December). Inference for best linear approximations to set identified functions. *arXiv e-prints*, arXiv:1212.5627.
- Chen, Q., M. Austern, and V. Syrgkanis (2023). Inference on optimal dynamic policies via softmax approximation.
- Chernozhukov, V., H. Hong, and E. Tamer (2007, August). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75, 1243–1284.
- Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica* 81, 667–737.
- Chernozhukov, V., S. Lee, A. M. Rosen, and L. Sun (2025). Policy learning with confidence.
- Chernozhukov, V., W. K. Newey, and A. Santos (2015). Constrained conditional moment restriction models.
- Cox, G. F. (2024). A simple and adaptive confidence interval when nuisance parameters satisfy an inequality. *arXiv preprint arXiv:2409.09962*.
- Cui, Y. and S. Han (2024). Policy learning with distributional welfare.
- 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.
- Dumbgen, L. and Spokoiny (2001). Multiscale testing of qualitative hypotheses. *The Annals of Statistics* 29(1), 124–152.
- Hahn, J. (1998, March). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 66(2), 315–331.

- Harter, H. L. (1964). Criteria for best substitute interval estimators, with an application to the normal distribution. *Journal of the American Statistical Association* 59(308), 1133–1140.
- 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.
- Kallus, N. (2022a). Treatment effect risk: Bounds and inference.
- Kallus, N. (2022b). What’s the harm? sharp bounds on the fraction negatively affected by treatment.
- Kallus, N., X. Mao, and A. Zhou (2020). Assessing algorithmic fairness with unobserved protected class using data combination.
- Kasy, M. (2016). Partial identification, distributional preferences, and the welfare ranking of policies. *Review of Economics and Statistics* 98(1), 111–131.
- Ketz, P. and A. McCloskey (2025). Short and simple confidence intervals when the directions of some effects are known. *Review of Economics and Statistics* 107(3), 820–834.
- Kitagawa, T., S. Lee, and C. Qiu (2022). Treatment choice with nonlinear regret. Working Paper, available at <https://arxiv.org/abs/2205.08586>.
- 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.
- Kitagawa, T. and A. Tetenov (2021). Equality-minded treatment choice. *Journal of Business & Economic Statistics* 39(2), 561–574.
- Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley & Sons.

- Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (3 ed.). Springer Texts in Statistics. New York: Springer.
- Levis, A. W., M. Bonvini, Z. Zeng, L. Keele, and E. H. Kennedy (2023). Covariate-assisted bounds on causal effects with instrumental variables.
- Liu, Y. and F. Molinari (2024). Inference for an algorithmic fairness-accuracy frontier.
- 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.
- Moon, S. (2025). Optimal policy choices under uncertainty. Working Paper, available at <https://arxiv.org/abs/2503.03910>.
- Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(2), 331–355.
- Qian, M. and S. A. Murphy (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics* 39(2), 1180 – 1210.
- Rai, Y. (2019). Statistical inference for treatment assignment policies. Working Paper, available at <https://sites.google.com/site/yoshiyasurai/home>.
- Robins, J. and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of American Statistical Association* 90(429), 122–129.
- Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In D. Y. Lin and P. J. Heagerty (Eds.), *Proceedings of the Second Seattle Symposium in Biostatistics*, Volume 179 of *Lecture Notes in Statistics*, pp. 189–326. New York, NY: Springer.
- 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*.
- Semenova, V. (2020). Generalized lee bounds.
- Semenova, V. (2023). Aggregated intersection bounds and aggregated minimax values.
- Shi, C., A. Fan, R. Song, and W. Lu (2018). High-dimensional a-learning for optimal dynamic treatment regimes. *The Annals of Statistics* 46(3), 925–957.
- Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics* 151, 70–81.
- Sun, L. (2021). Empirical welfare maximization with constraints.
- Terschuur, J. (2025). Locally robust policy learning: Inequality, inequality of opportunity and intergenerational mobility. *arXiv preprint arXiv:2502.13868*.
- 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.
- Whitehouse, J., M. Austern, and V. Syrgkanis (2025). Inference on optimal policy values and other irregular functionals via smoothing.
- Yata, K. (2021). Optimal decision rules under partial identification. Working Paper, available at <https://arxiv.org/abs/2111.04926>.

A Proofs for Sections 3–4

Section A.1 contains auxiliary statements and the proof of (3.17). The proof of (3.16) is given in Section A.2. Section A.3 contains the proof of Proposition 2.

A.1 Auxiliary statements

The first Lemma is Theorem 1 in [Luedtke and van der Laan \(2016\)](#).

Lemma A.1 (Efficiency Influence Function for the first-best welfare). *Suppose Assumption 4.1 holds and $P(\tau(X) = 0) = 0$. Then, the first-best welfare $\mathbb{E}[\max(m(1, X), m(0, X))]$ is pathwise differentiable with efficient influence function*

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

The second Lemma is a simple corollary of [Hahn \(1998\)](#).

Lemma A.2 (Efficiency bound for W_G). *Suppose Assumption 4.1 holds and let G be a known policy. Then, the welfare W_G is pathwise differentiable with efficient influence function*

$$\begin{aligned} \psi_G(Z) = & \left(m(1, X) + \frac{D}{\pi(X)}(Y - m(1, X)) \right) \mathbf{1}\{X \in G\} \\ & + \left(m(0, X) + \frac{1-D}{1-\pi(X)}(Y - m(0, X)) \right) \mathbf{1}\{X \in G^c\}. \end{aligned}$$

The corresponding variance is

$$\begin{aligned} \sigma_G^2 = & \mathbb{V}\text{ar}(m(1, X)\mathbf{1}(X \in G) + m(0, X)\mathbf{1}(X \in G^c)) \\ & + \mathbb{E} \left[\frac{\sigma^2(1, X)}{\pi(X)} \mathbf{1}(X \in G) + \frac{\sigma^2(0, X)}{1-\pi(X)} \mathbf{1}(X \notin G) \right], \end{aligned} \quad (\text{A.2})$$

where $\sigma^2(d, x) = \mathbb{V}\text{ar}(Y(d) | X = x) = \mathbb{V}\text{ar}(Y | D = d, D = x)$.

Proof. The parameter $W_G = \mathbb{E}[Y(1)\mathbf{1}\{X \in G\} + Y(0)\mathbf{1}\{X \in G^c\}]$ is a sum of two potential outcomes weighted by known functions of X , namely, $\mathbf{1}\{X \in G\}$ and $\mathbf{1}\{X \in G^c\}$. The form of the efficient influence function $\psi_G(Z) - W_G$ follows immediately from [Hahn \(1998\)](#), Theorem 1. By the Law of Total Variance,

$$\mathbb{V}\text{ar}(\psi_G(Z)) = \mathbb{V}\text{ar}(\mathbb{E}[\psi_G(Z) | X]) + \mathbb{E}[\mathbb{V}\text{ar}(\psi_G(Z) | X)].$$

By the Law of Iterated Expectations,

$$\mathbb{E}[\psi_G(Z) | X] = m(1, X)\mathbf{1}(X \in G) + m(0, X)\mathbf{1}(X \in G^c),$$

and $\mathbb{E}[\psi_G^2(Z) | X]$ takes the form

$$\begin{aligned} \mathbb{E}[\psi_G^2(Z) | X] &= m(1, X)^2\mathbf{1}(X \in G) + m(0, X)^2\mathbf{1}(X \in G^c) \\ &\quad + \frac{\sigma^2(1, X)}{\pi(X)}\mathbf{1}(X \in G) + \frac{\sigma^2(0, X)}{1 - \pi(X)}\mathbf{1}(X \in G^c). \end{aligned}$$

As a result,

$$\text{Var}(\psi_G(Z) | X) = \frac{\sigma^2(1, X)}{\pi(X)}\mathbf{1}(X \in G) + \frac{\sigma^2(0, X)}{1 - \pi(X)}\mathbf{1}(X \in G^c),$$

and the stated formula follows. ■

Note that plugging the unconstrained first best policy, $G^* = \{x \in \mathcal{X} : \tau(x) \geq 0\}$ into $\psi_G(Z)$, that is $\psi_{G^*}(Z) = \psi^*(Z)$. Thus, the efficiency bound is the same as if the first-best policy G^* was known.

The next Lemma states a uniform lower bound on $\text{Var}(\psi_G)$, which is useful in the sequel.

Lemma A.3 (A lower bound on variance). *Suppose Assumption 4.1(1) holds and, for each $d \in \{0, 1\}$,*

$$\text{ess inf}_x \text{Var}(Y(d) | X = x) \geq \underline{\sigma}^2 > 0.$$

Then, for any policy $G \subseteq \mathcal{X}$,

$$\text{Var}(\psi_G(Z)) \geq \underline{\sigma}^2.$$

Proof. Follows immediately from (A.2) and the fact that $\pi(X) \in (0, 1)$. ■

Lemma A.4 shows that for the DGP in Section 3.1, the welfare gap is proportional to $\epsilon = o(1)$ while the corresponding efficiency bounds remain strictly separated. As a result, it gives the proof for the second part of Proposition 1.

Lemma A.4 (Separated efficiency bounds). *The following calculations hold:*

1. $W_{\mathcal{X}} = 1/2 - \epsilon p$, $\sigma_{\mathcal{X}}^2 = \frac{p}{\pi(1)} + \frac{(1-p)}{\pi(0)} + \epsilon^2(1-p)p$.
2. $W_{G^*} = 1/2$, $\sigma_{G^*}^2 = \frac{10p}{1 - \pi(1)} + \frac{1-p}{\pi(0)}$.
3. $\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2 > 8p$ and $\sigma_{G^*} - \sigma_{\mathcal{X}} > p$.

4. $\Delta_{\mathcal{X}} > z_{1-\alpha}p/\sqrt{N}$, for each $N > z_{1-\alpha}^2$

Proof. Part 1. The value of $W_{\mathcal{X}}$ is computed in the main text. The efficiency bound for $W_{\mathcal{X}}$ in (A.2) consists of two summands. The first summand is

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

and the second 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 them up gives $\sigma_{\mathcal{X}}^2$.

Part 2. The value of W_{G^*} is computed in the main text. The efficiency bound of W_{G^*} in (A.2) consists of two summands. The first summand is

$$\mathbb{V}ar(\max(m(1, X), m(0, X))) = \mathbb{V}ar(1/2) = 0,$$

and the second one is

$$\mathbb{E}\left[\frac{X\mathbb{V}ar(Y \mid D = 0, X = 1)}{1 - \pi(1)} + \frac{(1 - X)\mathbb{V}ar(Y \mid D = 1, X = 0)}{\pi(0)}\right] = \frac{10p}{1 - \pi(1)} + \frac{1 - p}{\pi(0)}.$$

Adding them up yields $\sigma_{G^*}^2$.

Part 3. Recall that $\pi(0), \pi(1), p \in (1/4, 3/4)$. From Parts (1) and (2) it follows that

$$\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2 = p \frac{11\pi(1) - 1}{\pi(1)(1 - \pi(1))} - \epsilon^2(1 - p)p \stackrel{(i)}{>} p \frac{\pi(1)^2 + 10\pi(1) - 1}{\pi(1)(1 - \pi(1))} \stackrel{(ii)}{\geq} 8p,$$

where (i) holds by $\epsilon^2(1 - p) < 1$ and (ii) is attained at $\pi(1) = 1/4$, which can be verified numerically.

Part 4. Note that,

$$\sigma_{G^*}^2 = \frac{10p}{1 - \pi(1)} + \frac{1 - p}{\pi(0)} \leq 40p + 4(1 - p) = 4 + 36p \leq 31.$$

Similarly,

$$\sigma_{\mathcal{X}}^2 = \frac{p}{\pi(1)} + \frac{1 - p}{\pi(0)} + \epsilon^2(1 - p)p \leq 4p + 4(1 - p) + \epsilon^2(1 - p)p \leq 4 + \frac{1}{4}\epsilon^2 \leq 5.$$

Therefore,

$$\sigma_{G^*} - \sigma_{\mathcal{X}} = \frac{\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2}{\sigma_{G^*} + \sigma_{\mathcal{X}}} \geq \frac{8p}{\sqrt{31} + \sqrt{5}} > p.$$

Part 5 Combining the above results, we obtain

$$\Delta_{\mathcal{X}} = \frac{z_{1-\alpha}}{\sqrt{N}}(\sigma_{G^*} - \sigma_{\mathcal{X}}) - (W_{G^*} - W_{\mathcal{X}}) > \frac{z_{1-\alpha}}{\sqrt{N}}p - \epsilon p \geq \frac{pz_{1-\alpha}}{\sqrt{N}},$$

for $\epsilon \leq z_{1-\alpha}/\sqrt{N}$. It remains to ensure that $\epsilon^2(1-p) < 1$ which results in a bound on N . ■

A.2 Proof of Proposition 1

Notation and Preliminaries. Recall the class of DGPs defined in Section 3.1 and the notation introduced in Section 3.2. Note that $N_{dx} \sim \text{Binom}(N, \rho)$, $\rho = P(D = d, X = x)$, and $\hat{p} = \sum_{i=1}^N X_i/N \sim \text{Binom}(N, p)/N$. For any $\mathcal{Z} \sim \text{Binom}(N, \rho)$, for $N \geq 1$, the following standard properties hold:

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

Moreover, the following Chernoff's bounds hold, with $\mu = \mathbb{E}[\mathcal{Z}]$ and $\delta \in (0, 1)$,

$$P(\mathcal{Z} \geq (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{3}\right); \quad P(\mathcal{Z} \leq (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{2}\right) \quad (\text{A.3})$$

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

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

$$\mathbb{E}[(N_{dx} + 1)^{-2}] \leq 32N^{-2}. \quad (\text{A.5})$$

For any sequence $C_N \in (0, 1)$, letting $S = \sum_{i=1}^N X_i \sim \text{Binom}(N, 1/2)$ with $\mathbb{E}[S] = N/2$,

$$\begin{aligned} P(|\hat{p} - 1/2| \geq C_N) &= P(\hat{p} \geq 1/2 + C_N) + P(\hat{p} \leq 1/2 - C_N) \\ &= P(S \geq (1 + 2C_N)\frac{N}{2}) + P(S \leq (1 - 2C_N)\frac{N}{2}) \\ &\leq \exp\left(-\frac{2}{3}N(C_N)^2\right) + \exp\left(-N(C_N)^2\right) \\ &\leq 2\exp\left(-\frac{2}{3}N(C_N)^2\right). \end{aligned} \quad (\text{A.6})$$

Structure of the proof. 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 $d, x \in \{1, 0\}$*

$$(1 - 10C_N) < \mathbb{E}[\sigma^{-2}(d, 1)N \cdot \mathbb{V}ar(\widehat{m}_{d1} \mid \mathbf{X}, \mathbf{D}) \cdot \widehat{p}^2] < (1 + 10C_N). \quad (\text{A.7})$$

$$(1 - 10C_N) < \mathbb{E}[\sigma^{-2}(d, 0)N \cdot \mathbb{V}ar(\widehat{m}_{d0} \mid \mathbf{X}, \mathbf{D}) \cdot (1 - \widehat{p})^2] < (1 + 10C_N). \quad (\text{A.8})$$

Proof. Step 1 (Notation). Denote the expression inside the expectation of (A.7) by

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

and note that its probability limit as $N \rightarrow \infty$ equals 1. Denoting

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

we can decompose the asymptotic error as

$$\begin{aligned} \Xi_d - 1 &= NN_{d1}(N_{d1} + 1)^{-2}\widehat{p}^2 - 1 \\ &= \psi_d^1(\widehat{p}^2) - \psi_d^2(\widehat{p}^2) - 1 \\ &= \psi_d^1(\widehat{p}^2 - p^2) + \psi_d^1(p^2) - \psi_d^2(\widehat{p}^2) - 1 \\ &= \underbrace{\psi_d^1(\widehat{p}^2 - p^2)}_{S_2} + \underbrace{\psi_d^1(p^2) - N/(N + 1)}_{S_1} - \underbrace{\psi_d^2(\widehat{p}^2)}_{S_3} - \underbrace{1/(N + 1)}_{S_4}. \end{aligned} \quad (\text{A.9})$$

Step 2 (Leading term S_2). Recall that $p = 1/2$. On the event $\mathcal{M}_N = \{|\widehat{p} - 1/2| < C_N\}$, the error $|\widehat{p}^2 - 1/4| \leq |\widehat{p} - 1/2||\widehat{p} + 1/2| \leq 1.5C_N$. As a result,

$$\begin{aligned} |\mathbb{E}[S_2 \mathbf{1}\{\mathcal{M}_N\}]| &\leq \mathbb{E}[\psi_d^1(|\widehat{p}^2 - p^2|) \mathbf{1}\{\mathcal{M}_N\}] \\ &\leq \mathbb{E}[\psi_d^1(1.5C_N) \mathbf{1}\{\mathcal{M}_N\}] \\ &\leq 1.5C_N \mathbb{E}[\psi_d^1(1)] \\ &\leq 6C_N, \end{aligned} \quad (\text{A.10})$$

where the first three lines follow from linearity of $\psi_d^1(\cdot)$ and monotonicity of expectation and

the last one follows from (A.4). On the event \mathcal{M}_N^c , we can bound $|\widehat{p}^2 - p^2| \leq 1$, a.s., so that

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

where the last line follows from $N_{d1} \geq 0$ and $(N_{d1} + 1)^{-1} \leq 1$, a.s.. Using (A.6) and $C_N = \sqrt{2.25 \ln N/N}$,

$$NP(\mathcal{M}_N^c) \leq 2N \exp(-\frac{2}{3}N(C_N)^2) = 2N^{-1/2} \leq C_N, \quad \forall N \geq 6. \quad (\text{A.11})$$

Adding (A.10) and (A.11) gives $|\mathbb{E}[S_2]| \leq 7C_N$.

Step 3 (Terms S_1, S_3, S_4). Note that $S_4 = (N + 1)^{-1} \leq C_N$. Invoking (A.4) gives

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

Invoking (A.5) 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

$$|\mathbb{E}[\Xi_d - 1]| \leq \sum_{j=1}^4 |\mathbb{E}[S_d]| \leq 10C_N.$$

Step 4 (Conclusion). Steps 1–3 established (A.7), which corresponds to $x = 1$. The symmetry of DGPs implies (A.8) 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

$$N \cdot MSE(\widehat{W}_{G^*}) > (\sigma^2(1, 0) + \sigma^2(0, 1))(1 - 10C_N). \quad (\text{A.12})$$

Proof. **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 $i, j \in \{1, 2, \dots, N\}$, we show that

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

If the indices are distinct, $\mathbb{C}ov(Y_i, Y_j \mid \mathbf{X}, \mathbf{D}) = 0$ 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 of the form $X_i(1 - X_j)\text{cov}(Y_i, Y_j \mid \mathbf{X}, \mathbf{D})$, we obtain

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

Thus, the variance of each estimator is

$$\mathbb{V}ar(\widehat{W}_{G^*} \mid \mathbf{X}, \mathbf{D}) = \mathbb{V}ar(\widehat{m}_{01} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 + \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2 \quad (\text{A.13})$$

$$\mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}) = \mathbb{V}ar(\widehat{m}_{11} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2 + \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2. \quad (\text{A.14})$$

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

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{01} \mid \mathbf{X}, \mathbf{D})\widehat{p}^2] > \sigma^2(0, 1)(1 - 10C_N) \quad (\text{A.15})$$

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{m}_{10} \mid \mathbf{X}, \mathbf{D})(1 - \widehat{p})^2] > \sigma^2(1, 0)(1 - 10C_N) \quad (\text{A.16})$$

Adding (A.15) and (A.16) gives a lower bound on $\mathbb{E}[\mathbb{V}ar(\widehat{W}_{G^*} \mid \mathbf{X}, \mathbf{D})]$. A lower bound (A.12) on $MSE(\widehat{W}_{G^*})$ follows. \blacksquare

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

$$N \cdot MSE(\widehat{W}_{\mathcal{X}}) < \sigma_{\mathcal{X}}^2 + \frac{N\epsilon^2}{2} + (4 + 10(\sigma^2(1, 1) + \sigma^2(1, 0)))C_N. \quad (\text{A.17})$$

Proof. **Step 1 (Bias).** The remainder term $R = W_{\mathcal{X}} - \widehat{W}_{\mathcal{X}}$ takes the form

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

and is non-negative a.s. for $\epsilon \in (0, 1/2)$. Furthermore, it is bounded as

$$0 \leq \mathbb{E}[R] \stackrel{(i)}{\leq} 1/2\mathbb{E}[(N_{11} + 1)^{-1}] + 1/2\mathbb{E}[(N_{10} + 1)^{-1}] \stackrel{(ii)}{\leq} 4/N,$$

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.4). Next, note that $\mathbb{E}[1/2 - \epsilon\widehat{p}] = 1/2 - \epsilon p = W_{\mathcal{X}}$ since $\mathbb{E}[\widehat{p}] = p$. The bias is bounded from above and below

$$\begin{aligned} 0 \leq |W_{G^*} - \mathbb{E}[\widehat{W}_{\mathcal{X}}]| &\leq |W_{G^*} - W_{\mathcal{X}}| + |W_{\mathcal{X}} - \mathbb{E}[\widehat{W}_{\mathcal{X}}]| \\ &= |W_{G^*} - W_{\mathcal{X}}| + |\mathbb{E}[R]| \\ &\leq \epsilon/2 + 4N^{-1}. \end{aligned} \quad (\text{A.19})$$

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

$$N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}}) < \sigma_{\mathcal{X}}^2 + (\sigma^2(1, 1) + \sigma^2(1, 0))10C_N + 2C_N. \quad (\text{A.20})$$

The variance of the conditional mean is

$$\mathbb{V}ar(\mathbb{E}[\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D}]) = \mathbb{V}ar(R) - 2\mathbb{C}ov(R, 1/2 - \epsilon\widehat{p}) + \frac{\epsilon^2}{4N}. \quad (\text{A.21})$$

Invoking (A.5) 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

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

Invoking (A.7) gives

$$\mathbb{E}[N \cdot \mathbb{V}ar(\widehat{W}_{\mathcal{X}} \mid \mathbf{X}, \mathbf{D})] \leq (\sigma^2(1, 1) + \sigma^2(1, 0))(1 + 10C_N) \quad (\text{A.22})$$

Adding (A.21) and (A.22) gives

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

Step 3 (MSE). Combining (A.19) and (A.20) gives (A.17) 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 (3.16) 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 $MSE(\widehat{W}_{\mathcal{X}})$

$$N \cdot MSE(\widehat{W}_{\mathcal{X}}) < 3/4 + (\sigma^2(1, 1) + \sigma^2(1, 0)) + [4 + 10(\sigma^2(1, 1) + \sigma^2(1, 0))]C_N \quad (\text{A.23})$$

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

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

■

A.3 Proof of Proposition 2

We start with an auxiliary Lemma. Let $G^* \triangle G = G^* \setminus G \cup G \setminus G^*$ denote the symmetric difference of sets G^* and 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), borrowing from 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 4.1 and 4.2 hold. Then, (1) The welfare gap is bounded*

$$C_B(P(X \in G^* \triangle G))^{1+1/\delta} \leq W_{G^*} - W_G \leq MP(X \in G^* \triangle G), \quad (\text{A.24})$$

where $C_B = \eta\delta(\frac{1}{1+\delta})^{1+1/\delta} > 0$;

(2) The variance gap is bounded as

$$\sigma_{G^*}^2 - \sigma_G^2 \leq \frac{5}{4} \frac{M^2}{\kappa} P(X \in G^* \triangle G). \quad (\text{A.25})$$

(3) The standard deviation gap is bounded as

$$\sigma_{G^*} - \sigma_G \leq \frac{5}{4} \frac{M^2}{2\bar{\sigma}\kappa} P(X \in G^* \triangle G). \quad (\text{A.26})$$

Proof. Step 1. The lower bound (A.24) is stated as Lemma A.7 in Kitagawa and Tetenov (2018a) and originally established in Tsybakov (2004). The upper bound is straightforward.

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 (2.1) can be equivalently rewritten as

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

Since $\mathbf{1}\{X \in G\} \mathbf{1}\{X \in G^c\} = 0$ a.s., we have

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

Thus, by the Law of Total Variance,

$$\sigma_G^2 = \sum_{d \in \{1,0\}} \mathbb{E} \left[\left(\frac{\sigma^2(d, X)}{P(D=d | X)} + m^2(d, X) \right) \mathbf{1}\{X \in G_d\} \right] - W_G^2. \quad (\text{A.27})$$

Denoting

$$\begin{aligned} T_{1G} &= \mathbb{E} \left[\left(\frac{\sigma^2(1, X)}{\pi(X)} - \frac{\sigma^2(0, X)}{1 - \pi(X)} \right) \left(\mathbf{1}\{X \in G^* \setminus G\} - \mathbf{1}\{X \in G \setminus G^*\} \right) \right]; \\ T_{2G} &= \mathbb{E} \left[\left(m^2(1, X) - m^2(0, X) \right) \left(\mathbf{1}\{X \in G^* \setminus G\} - \mathbf{1}\{X \in G \setminus G^*\} \right) \right]; \\ T_{3G} &= -(W_{G^*}^2 - W_G^2), \end{aligned}$$

we can write

$$\sigma_{G^*}^2 - \sigma_G^2 = T_{1G} + T_{2G} + T_{3G}. \quad (\text{A.28})$$

Step 3. By Assumption 4.1 and Jensen inequality, $\sigma^2(d, x) \leq m^2(d, x) \leq M^2/4$, for all $d \in \{0, 1\}$, $x \in \mathcal{X}$. Thus, the first two terms are bounded as

$$\begin{aligned} T_{1G} &\leq M^2/(2\kappa) \cdot P(X \in G^* \triangle G) \\ T_{2G} &\leq M^2/2 \cdot P(X \in G^* \triangle G). \end{aligned}$$

The final term is bounded as

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

where the second inequality follows from $|W_{G^*} + W_G| \leq M$ and (A.24). Collecting the terms and using the fact that $\kappa < 1/2$,

$$T_{1G} + T_{2G} + T_{3G} \leq \frac{M^2}{\kappa} \frac{1+3\kappa}{2} \leq \frac{5}{4} \frac{M^2}{\kappa},$$

so (A.25) follows.

Step 4. The corresponding bound on the standard derivation gap is

$$\sigma_{G^*} - \sigma_G \stackrel{(i)}{\leq} \frac{\sigma_{G^*}^2 - \sigma_G^2}{2\underline{\sigma}} \stackrel{(ii)}{\leq} \frac{5}{4} \frac{M^2}{2\underline{\sigma}\kappa} P(X \in G^* \triangle G)$$

where (i) follows from $\sigma_{G^*}^2 \geq \underline{\sigma}^2$ and $\sigma_G^2 \geq \underline{\sigma}^2$ (from Lemma A.3) and (ii) from (A.25). \blacksquare

A.3.1 Proof of Proposition 2: Upper Bound

Let $C_A = N^{-1/2} 1.25 z_{1-\alpha} M^2 / (2\sigma\kappa)$ and C_B be as defined in Lemma A.9. Define a function $g : \mathbf{R} \rightarrow \mathbf{R}$ as

$$g(x) = C_A x - C_B x^{1/\delta+1}. \quad (\text{A.29})$$

The LCB gap Δ_G is bounded as

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

where (i) follows from (A.26) and (ii) from the lower bound in (A.24). Note that the function $g(x)$ is globally concave. Its' global maximum and maximizer are

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

Therefore, for any $G \subseteq \mathcal{X}$,

$$\Delta_G \leq g(x^*) = \overline{C} N^{-\frac{1+\delta}{2}},$$

where

$$\overline{C}_{\delta,\eta} = \left(\frac{1.25 z_{1-\alpha} M^2}{2\sigma\kappa} \right)^{1+\delta} \frac{1}{\eta^\delta}.$$

Under our assumptions, $\overline{C} = \max_{\delta,\eta} \overline{C}_{\delta,\eta} < \infty$ and the slowest rate is attained at $\underline{\delta}$.

A.3.2 Proof of Proposition 2: Lower Bound

The proof is constructive and consists of three steps. Step 1 describes a class of DGPs. Step 2 shows that the proposed DGPs belong to the model \mathbf{P} . Step 3 establishes the lower bound.

Step 1. Let $X \sim U[0, 1]$, and the propensity score be constant, $\pi(x) = 1/2$, \mathcal{X} -a.s. Let $\epsilon \in (0, 1/2)$ and $\nu > 0$ be a rational number for which the function $a \mapsto a^\nu$ is well-defined for both positive and negative values of a .¹³ Let Y be a random variable supported on $[-M/2, M/2]$ so that the conditional means and second moments are bounded as $|m(d, x)| \leq M/2$ and $m^2(d, x) \leq M^2/4$. Consider the following specification:

$$\begin{aligned} m(1, x) &= 0; & \sigma^2(1, x) &= M^2/10; \\ m(0, x) &= -(x - \epsilon)^\nu M/5; & \sigma^2(0, x) &= M^2/5. \end{aligned} \quad (\text{A.30})$$

¹³This is the case if and only if $\nu = \frac{p}{q}$, where p, q are natural numbers with $\gcd(p, q) = 1$.

The CATE function is given by

$$\tau(x) = m(1, x) - m(0, x) = (x - \epsilon)^\nu M/5,$$

so the first-best policy is

$$G^* = [\epsilon, 1].$$

Evidently, this distribution satisfies Assumption 4.1.

Step 2. We show that the proposed sequence of DGPs satisfies Assumption 4.2 for a suitable choice of ν . Note that for t such that $(5t/M)^{1/\nu} \leq \epsilon$,

$$P(|X - \epsilon|^\nu M/5 \leq t) = P(|X - \epsilon| \leq (5t/M)^{1/\nu}) = (\epsilon + (5t/M)^{1/\nu}) - (\epsilon - (5t/M)^{1/\nu}) = 2(5t/M)^{1/\nu}.$$

For t such that $(5t/M)^{1/\nu} \geq \epsilon$,

$$P(|X - \epsilon| \leq (5t/M)^{1/\nu}) \leq \epsilon + (5t/M)^{1/\nu} \leq 2(5t/M)^{1/\nu}.$$

Thus, choosing ν such that $\delta = 1/\nu \geq \underline{\delta}$ but arbitrarily close to it,¹⁴

$$P(|X - \epsilon|^\nu M/5 \leq t) \leq 2(5t/M)^{1/\nu} = \left(\frac{t}{\eta}\right)^\delta,$$

so that (4.1) holds for any $\varepsilon \in (0, 1/2)$.

Step 3. The first-best policy differs from \mathcal{X} only for $x \in [0, \epsilon]$. Thus, the welfare gap is

$$W_{G^*} - W_{\mathcal{X}} = - \int_0^\epsilon (x - \epsilon)^\nu M/5 dx = \frac{\epsilon^{\nu+1}}{\nu+1} M/5.$$

The variance gap is obtained by plugging $G = \mathcal{X}$ into (A.28). We have

$$T_{1\mathcal{X}} = \frac{\epsilon M^2}{5}; \quad T_{2\mathcal{X}} = \frac{\epsilon^{2\nu+1}}{2\nu+1} \frac{M^2}{25}; \quad T_{3\mathcal{X}} = - \frac{\epsilon^{\nu+1}}{\nu+1} \left(2 \frac{(1-\epsilon)^{\nu+1}}{\nu+1} - \frac{\epsilon^{\nu+1}}{\nu+1} \right) \frac{M^2}{25},$$

¹⁴Since rationals are dense in reals, it is without loss of generality to assume $\underline{\delta}$ is rational. Then, it can be expressed either as $\frac{p}{2^d q}$ where p, d, q are natural numbers and $\gcd(p, q) = 1$, or as $\frac{2^d p}{q}$ with the same conditions. In the former case, setting $\delta = 1/\nu = \underline{\delta}$ leads to ν satisfying the requirement of footnote 13. In the latter case, setting $\delta = 1/\nu = \left(\frac{k}{k-1}\right)^d \frac{2^d p}{q}$ for any prime number k corresponds to $\nu = \left(\frac{k-1}{2}\right)^2 \frac{q}{2^d p}$, which is also satisfies the requirement of footnote 13. For arbitrarily large k , $1/\nu$ will be arbitrarily close to $\underline{\delta}$.

so that

$$\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2 = \frac{4M^2}{25}\epsilon + \underbrace{\left(\frac{\epsilon^{2\nu+1}}{2\nu+1} + \frac{\epsilon^{2\nu+2}}{(\nu+1)^2} \right) \frac{M^2}{25}}_{\geq 0} + \underbrace{\left(\epsilon - \frac{2}{(\nu+1)^2} \epsilon^{\nu+1} (1-\epsilon)^{\nu+1} \right) \frac{M^2}{25}}_{=f(\epsilon)} > \frac{4M^2}{25}\epsilon,$$

where the final inequality follows from the fact that $f'(\epsilon) \geq 0$ so $f(\epsilon) \geq f(0) = 0$. On the other hand, recalling the DGP in (A.30), we can bound $\sigma_{G^*}^2 < M^2/5(1+\epsilon) < 3M^2/10$ and $\sigma_{\mathcal{X}}^2 < M^2/5$. Thus, $\sigma_{G^*} + \sigma_{\mathcal{X}} < M$, and

$$\sigma_{G^*} - \sigma_{\mathcal{X}} = \frac{\sigma_{G^*}^2 - \sigma_{\mathcal{X}}^2}{\sigma_{G^*} + \sigma_{\mathcal{X}}} > \frac{4M}{25}\epsilon.$$

Setting $\epsilon^\nu = (4z_{1-\alpha}/5)N^{-1/2}$ and recalling that $\nu = 1/\delta$ gives a lower bound

$$\Delta_{\mathcal{X}} > \underline{C}N^{-\frac{1+\delta}{2}},$$

where $\underline{C}_\delta = \frac{4M}{5} \left(\frac{4z_{1-\alpha}}{5} \right)^{1+\delta} \frac{1}{1+\delta}$. Since the above inequality holds for all δ arbitrarily close to $\underline{\delta}$, the stated result follows. \blacksquare

B Inverting Moment Inequality Tests

B.1 Proof of Proposition 4

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 conservatively define $LCB_{\mathcal{G}}^{GMS}$ as the lowest point of the confidence set obtained by test inversion.

Lemma B.1 (LCB based on GMS test inversion). *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:

$$\widehat{LCB}_{\max}^{GMS} = \min\{\theta^{(j)} : t^{(j)} \geq \theta^{(j)} > t^{(j+1)}\}. \quad (\text{B.1})$$

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

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

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 $\widehat{c}_\alpha(\theta)$. The minimum in the expression for $\widehat{LCB}_{\max}^{GMS}$ selects the lower point of intersection. ■

B.2 Proof of Proposition 5

To simplify notation, we write $X - \theta \mathbf{1}$ instead of $\sqrt{N}(\widehat{W}_{\text{test}} - \theta \mathbf{1})$. For the strictly convex minimization problem:

$$f^* = \min_{t \in \mathbb{R}^d, t \leq 0} \{(X - \theta \mathbf{1} - t)' \Sigma^{-1} (X - \theta \mathbf{1} - t)\},$$

consider the dual objective function:

$$g(u) = \min_{t \in \mathbb{R}^d} \{(X - \theta \mathbf{1} - t)' \Sigma^{-1} (X - \theta \mathbf{1} - t) + u' t\},$$

where $u \geq 0$ is a vector of the Lagrange multipliers. Since the Slater condition holds, strong duality applies, so $f^* = \max_{u \geq 0} g(u)$. Simple algebra yields

$$g(u) = (X - \theta \mathbf{1})' u - \frac{1}{4} u' \Sigma u,$$

so the event of not rejecting the LR test can be equivalently written as:

$$\max_{u \geq 0} \left\{ (X - \theta \mathbf{1})' u - \frac{1}{4} u' \Sigma u \right\} \leq c_{\alpha, LR}^{LF}.$$

For $u = 0$, the inequality trivially holds, and for all $u \geq 0$ with $u \neq 0$ it is equivalent to

$$\theta \geq \frac{1}{(\sum_{j=1}^d u_j)} \{X' u - \frac{1}{4} u' \Sigma u - c_{\alpha, LR}^{LF}\}.$$

Any $u \geq 0$ with $u \neq 0$ can be written as $u = \lambda \cdot \gamma$, where $\lambda \geq 0$ satisfies $\sum_{j=1}^d \lambda_j = 1$, and $\gamma > 0$. Thus, the above display is equivalent to

$$\theta \geq X' \lambda - \frac{1}{4} \lambda' \Sigma \lambda \cdot \gamma - \frac{c_{\alpha, LR}^{LF}}{\gamma}.$$

Since this inequality holds for all $\lambda \geq 0$ with $\lambda' \mathbf{1} = 1$, and all $\gamma > 0$,

$$\theta \geq \max_{\lambda \in \Lambda, \gamma > 0} \left\{ X' \lambda - \frac{1}{4} \lambda' \Sigma \lambda \cdot \gamma - \frac{c_{\alpha, LR}^{LF}}{\gamma} \right\}.$$

Concentrating out γ yields the stated result. ■

C Auxiliary Empirical Details

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

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

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.