We propose a new framework for robust nonparametric estimation of optimal treatment regimes under flexible fairness constraints. Under standard regularity conditions we show that the resulting estimators possess the double robustness property. We use this framework to characterize the trade-off between fairness and the maximum welfare that is achievable by the optimal treatment policy.

1 Introduction

In today's world, an increasing number of decisions that affect people's lives are automatically made by machine learning models. Such decision-making systems are implemented in various settings ranging from financial investment to healthcare policy. Considering the importance of such decisions at an individual and societal level, it is crucial to ensure that the underlying models are not only accurate but fair. In this work, by fairness we mean that the models are not biased so that they do not systematically benefit or harm a specific group of people, such as a minority ethnic group. The need to address such algorithmic biases has given rise to an explosion of works studying algorithmic fairness (e.g., see [3] for a review). However, despite the considerable amount of studies in this area, comparatively little attention has been given to fairness in causal inference. In this work, we propose a novel framework for estimating optimal treatment assignments or regimes in a fair and robust manner, leveraging recent developments in counterfactual optimization [20, 21].

1.1 Related Work

Much of the earlier work on estimating optimal treatment regimes involves postulating a parametric model for the outcome regression function [e.g., 4, 10, 30, 37]. More robust approaches based on the idea of doubly robust estimation have also been proposed, for example, in [45, 46]. In recent studies [1, 13, 22], flexible nonparametric approaches are discussed where an optimal policy is deployed from a pre-specified class that can encode problem-specific constraints. However, they do not provide means to incorporate general fairness constraints.

In order to mitigate algorithmic biases where model performance varies over sensitive features, a wide array of fairness criteria have been developed typically by placing restrictions on the joint distribution of model outcomes and sensitive features. Popular fairness criteria include independence (or statistical parity) [3] and separation (or equalized odds) [9]. In some cases such as risk assessment settings [e.g., 7], counterfactual fairness may be of interest where fairness criteria depend on potential (or counterfactual) outcomes with respect to the sensitive feature [e.g., 23, 31] or a decision variable [e.g., 7, 28, 29]. It is also well known that there exists a fairness-accuracy tradeoff, because in some cases the most accurate models under consideration do not satisfy a chosen fairness criterion [e.g., 27, 29, 36].

Interestingly, little work has been done at the intersection of these two areas. A few important exceptions include [32] which integrates algorithmic fairness and policy learning using tools from
1.2 Contribution

Our method builds on a promising literature at the intersection of algorithmic fairness, causal inference, and stochastic optimization, bridging the gap between algorithmic fairness and optimal treatment regimes. At this intersection, our contribution is twofold. First, we propose a robust estimator of optimal treatment regimes under general fairness constraints. We cast our estimator as a convex quadratic program that can be readily solved with off-the-shelf solvers. We show that the resulting estimators are doubly robust under standard regularity conditions. Our proposed approach contributes to \[42\] in terms of robustness, and to \[44\] in terms of ease of implementation and interpretability. Second, by analyzing the welfare regret bound, we characterize the trade-off between the maximum possible benefit and fairness. This will be useful for understanding, for example, how a desired level of fairness requires a welfare compromise.

2 Setup and Framework

2.1 Optimal Treatment Regimes

Suppose that we have access to an i.i.d. sample \((Z_1, ..., Z_n)\) of \(n\) tuples \(Z = (Y, A, S, X) \sim \mathbb{P}\) for some distribution \(\mathbb{P}\), outcome \(Y \in \mathbb{R}\), binary intervention \(A \in \{0, 1\}\), sensitive feature \(S \in \{0, 1\}\), and additional covariates \(X \in X \subset \mathbb{R}^d\) for some compact subset \(X\). Throughout we assume larger values of \(Y\) are preferred. We let \(W = (S, X) \in W\) represent the measured pre-intervention variables and let \(Y^a\) denote the potential outcome that would have been observed (possibly contrary to fact) under treatment or intervention \(A = a\). A policy maker has to choose a treatment policy or a treatment regime\(^1\) that is a function \(g : W \rightarrow \{0, 1\}\) to determine whether individuals with covariates \(W\) will be assigned to the treatment 0 or 1. For an arbitrary treatment regime \(g\), we define the welfare function for which the treatment regime \(g \in G\) is applied to the population \(\mathbb{P}\) by

\[
\mathcal{R}(g) = \mathbb{E}\left\{ Y^1 g(W) + Y^0 (1 - g(W)) \right\}.
\]

Throughout we assume the standard causal assumptions of consistency, no unmeasured confounding, and positivity [e.g., \[11\] Chapter 12]. Under these assumptions, it is straightforward to show that the optimal treatment regime leading to the largest value of \(\mathcal{R}(g)\) is given by

\[
g^*(W) = \mathbb{I}\{\mu_1(W) > \mu_0(W)\},
\]

where \(\mu_a(W) = \mathbb{E}[Y | W, A = a], \forall a \in \{0, 1\}\); i.e., the optimal regime assigns the treatment that yields the larger mean outcome conditional on the individual characteristics.\(^2\)

2.2 Simple Motivating Example

Sometimes, efficient estimation of \(g^*\) in \[1\] alone can result in unfair treatment policies. Consider the following simple data-generating process

\[
A \sim \text{Bernoulli}(0.5), \quad X \sim \text{Unif}[-1, 1]
\]

\[
\mathbb{P}(S = 1) = \expit(7.5X), \quad \mu_a(W) = AX,
\]

where \(\expit\) and \(\text{Unif}(l, u)\) denote the inverse logit function and the uniform distribution over the interval \([l, u]\). Then the optimal treatment regime is \(\mathbb{I}(X > 0)\). However, when we generate 100 samples, as can be seen in Figure [1], a serious fairness problem is observed; under the optimal treatment regime only less than 7% of individuals with \(S = 0\) are treated, while more than 95% of individuals in the untreated group are \(S = 0\).

Here, group \(S = 0\) is discriminated by the estimated optimal treatment regime that is designed to result in the greatest benefit overall in the population. In data-driven decision-making, this kind of algorithmic bias can lead to critical issues in the real world as illustrated in the following examples.

---

\(^1\)In this work, we use the terms "treatment policy" and "treatment regime" interchangeably to refer to any mapping from the pre-treatment variables to the treatment.

\(^2\)Here, the strict inequality follows from the convention [see, e.g., \[45\]].
Figure 1: When the optimal treatment regime is applied, only less than 7% of individuals with $S = 0$ are treated while more than 95% of individuals in the untreated group are $S = 0$.

- Stop-and-Frisk: if $A$ represents the policing practice of stop-and-frisk program, the established optimal treatment regime could be used as a recipe for discriminatory practice of stop-and-frisk toward specific ethnic groups.

- Medical Resource Allocation: if $A$ represents access to medical treatment or health care resources, many recent studies advocate not only cost-effectiveness but also other ethical values for rationing limited health resources [e.g., 8, 36].

2.3 Proposed Framework

In this section, we lay out a framework for estimating optimal treatment regimes where we can minimize algorithmic unfairness below a particular level. Our strategy is to estimate each outcome regression function $\mu_{a}$ satisfying desired fairness criteria, and then plug back into the formula (1) so that the same fairness criteria are also satisfied in the optimal regime.

Specifically, we aim to estimate a functional approximation of $\mu_{a}$, defined by a projection onto a finite-dimensional parametric model subject to fairness constraints. Our target parameter can be reformulated as the following constrained stochastic optimization problem

$$\min_{\beta \in \mathcal{B}} L_{\text{MSE}} \left( Y^{a}, \beta^T b(W) \right) := \mathbb{E} \left\{ (Y^{a} - \beta^T b(W))^2 \right\}$$

subject to $\beta \in \mathcal{C}_{\text{fair}} := \left\{ \beta \mid \mathbb{E} \left\{ g_{j}(Y^{a}, W) \beta^T b(W) \right\} \leq \delta_{j}, j \in J \right\},$

for some $\delta_{j} \geq 0$ and $J = \{1, ..., m\}, \delta_{j}$ is a prespecified tolerance for the maximum acceptable level of unfairness. The solution of the above program corresponds to the coefficients of the estimated best-fitting function of $\mu_{a}$ on the finite-dimensional model space spanned by the basis functions $b(W) = [b_{1}(W), ..., b_{k}(W)]^T$ subject to $m$ fairness constraints in $\mathcal{C}_{\text{fair}}$. Note that we do not assume anything about the true functional relationship between $Y^{a}$ and $W$. This form of aggregated estimators are widely used in nonparametric regression [e.g., 12, 32].

Following [28], we use the canonical form of fairness function $g_{j} : \mathcal{W} \times \mathcal{Y} \rightarrow \mathbb{R}$ to accommodate a broad range of fairness measures. For example, the criterion of independence that requires our model to be independent of the sensitive feature can be applied by letting

$$g_{j}(Y^{a}, W) = \frac{1 - S}{\mathbb{E}(1 - S)} - \frac{S}{\mathbb{E}(S)}$$

which leads to $|\mathbb{E} \left\{ \beta^T b(W) \mid S = 0 \right\} - \mathbb{E} \left\{ \beta^T b(W) \mid S = 1 \right\}| \leq \delta_{j}$. We refer to [28] Section 3 for more examples.

Similar projection approaches have also been used in causal inference [e.g., 19, 34, 39]. There are several reasons why the above projection approach is preferred in our setting. First, as will be seen shortly, the coefficients $\beta$ may be estimated with flexible nonparametric methods while achieving the

3
property of double robustness and tractable inference, and so does the target parameter $g^*$. It also provides interpretability; it allows practitioners to understand and audit the resulting optimal regimes according to the specified level of unfairness. Further, one may flexibly incorporate not only various fairness constraints but also other practical constraints into estimation. Finally, the optimal solution of $(P_{\alpha})$ can be readily estimated by solving the convex quadratic program that approximates $(P_{\hat{\alpha}})$, which will be described in the following section.

**Remark 1.** Another notable feature of our framework is that we can consider a general setting where only a subset of covariates $V \subseteq W$ can be used for predicting the counterfactual outcome $Y^a$. This allows for runtime confounding, where some factors used by decision-makers are recorded in the training data (used to construct nuisance estimates) but are not available for prediction (see [6] and references therein).

**Notation.** Here we briefly introduce some notation used throughout this paper. For any fixed vector $v$, we let $||v||_q$ denote the $L_q$-norm. Let $P_n$ denote the empirical measure over $(Z_1, ..., Z_n)$. Given a sample operator $h$ (e.g., an estimated function), we let $P$ denote the conditional expectation over a new independent observation $Z$, as in $P(h) = \mathbb{E}\{h(Z)\} = \int h(z) dP(z)$. Then we use $\|h\|_{q, P}$ to denote the $L_q(P)$ norm of $h$ defined by $\|h\|_{q, P} = \left\{ \int |h(z)|^q dP(z) \right\}^{\frac{1}{q}}$. Lastly, we let $\lesssim$ denote less than or equal to up to a nonnegative constant.

## 3 Estimation and Inference

$(P_{\alpha})$ is not directly solvable so we need to find an approximating program of the “true” program $(P_{\alpha})$. A complication arises since standard approaches to stochastic programming such as stochastic approximation (SA) and sample average approximation (SAA) (e.g., [33, 40]) are infeasible in our setting, because i) the relevant sample moments and stochastic (sub)gradients depend on unobserved counterfactuals, and ii) these approaches cannot incorporate efficient semiparametric estimators with cross-fitting [5, 35]. We therefore build our estimators on the recent developments by [20, 21] where counterfactual components are estimated more flexibly.

For convenience, define the following:

$$\pi_a(X) = P[A = a | X],$$

$$\varphi_a(Z; \eta) = \frac{1(A = a)}{\pi_a(X)} \{ Y - \mu_a(X) \} + \mu_a(X).$$

$\varphi_a$ is the uncentered efficient influence function for the parameter $\mathbb{E} \{ Y | X, A = a \}$ with a set of the nuisance components defined by $\eta = \{ \pi_a(X), \mu_a(X) \}$ [15].

First, we provide influence-function-based semiparametric estimators for each component of $(P_{\alpha})$. Following [5, 16, 38, 47], we propose to use sample splitting (or cross fitting) to allow for arbitrarily complex nuisance estimators $\hat{\eta}$. Specifically, we split the data into $K$ disjoint groups, each with size $n/K$ approximately, by drawing variables $(B_1, ..., B_n)$ independent of the data, with $B_i = b$ indicating that subject $i$ was split into group $b \in \{1, ..., K\}$. Then the semiparametric estimators for $C_{\text{MSE}}$ and each element in $C_{\text{fair}}$ based on the efficient influence function and sample splitting are given by

$$\frac{1}{K} \sum_{b=1}^{K} \left\{ \varphi_a(Z; \hat{\eta}_{-b}) - \beta^T b(W) \right\}^2 \equiv P_n \left\{ \left( \varphi_a(Z; \hat{\eta}_{-b}) - \beta^T b(W) \right)^2 \right\},$$

$$\frac{1}{K} \sum_{b=1}^{K} \left\{ g_j(\varphi_a(Z; \hat{\eta}_{-b}), W) \beta^T b(W) \right\} \equiv P_n \left\{ g_j(\varphi_a(Z; \hat{\eta}_{-b}), W) \beta^T b(W) \right\},$$

where we let $P_n^{b}$ denote empirical averages only over the set of units $\{i : B_i = b\}$ in group $b$ and let $\hat{\eta}_{-b}$ denote the nuisance estimator constructed only using those units $\{i : B_i \neq b\}$. Under weak regularity conditions, these sample-splitting-based semiparametric estimators attain the efficiency bound with the double robustness property, and thus allow us to employ flexible machine

---

3When $h$ is a fixed operator, $P$ and $\mathbb{E}$ are used interchangeably.
learning estimation methods while achieving the \(\sqrt{n}\)-rate of convergence and valid inference \cite{15}.

Consequently, our approximating program can be found as the following convex quadratic program (QP)

\[
\begin{aligned}
\text{minimize } & \mathbb{P}_n \left\{ (\varphi_a(Z; \tilde{\eta}_{-b}) - \beta^\top b(W))^2 \right\} \\
\text{subject to } & \mathbb{P}_n \left\{ g_j(\varphi_a(Z; \tilde{\eta}_{-b}), W)\beta^\top b(W) \right\} \leq \delta_j, j \in J.
\end{aligned}
\]

\(\mathbb{P}_n\) can be solved using off-the-shelf QP solvers. Next, we introduce the following assumptions for our counterfactual component estimators.

\(\text{(A1)}\) \(\mathbb{P}(\tilde{\pi}_a \in [\epsilon, 1 - \epsilon]) = 1\) for some \(\epsilon > 0\)

\(\text{(A2)}\) \(|\tilde{\mu}_a - \mu_a|_{2, P} = o_P(1)\) or \(|\tilde{\pi}_a - \pi_a|_{2, P} = o_P(1)\)

\(\text{(A3)}\) \(|\tilde{\pi}_a - \pi_a|_{2, P}||\tilde{\mu}_a - \mu_a|_{2, P} = o_P(n^{-\frac{1}{2}})\)

Assumptions \(\text{(A1)}, \text{(A2)}, \text{(A3)}\) are commonly used in semiparametric estimation in the causal inference literature \cite{14}. In the following theorem, we provide the large-sample properties of our proposed estimator.

**Theorem 3.1.** Let \(\beta^*\) and \(\tilde{\beta}\) denote the optimal solutions to \((\mathbb{P}_{\mu_a})\) and \((\mathbb{P}_{\mu_a}^*)\), respectively. If Assumptions \(\text{(A1)}\) and \(\text{(A2)}\) hold, then

\[
||\tilde{\beta} - \beta^*||_2 = O_P \left( ||\tilde{\pi}_a - \pi_a||_{2, P} ||\tilde{\pi}_a - \pi_a||_{2, P} \vee n^{-\frac{1}{2}} \right).
\]

If we additionally assume \(\text{(A2)}\) uniqueness of \(\beta^*\), and that the Linear Independence Constraint Qualification (LICQ) and Strict Complementarity (SC) hold at \(\beta^*\), then \(\sqrt{n}(\tilde{\beta} - \beta^*)\) converges in distribution to a zero-mean normal random variable. Further, \(\tilde{\beta}\) is efficient, meaning that there exist no other regular asymptotically linear estimators that are asymptotically unbiased and have smaller variance.

The above result immediately follows by Theorems 3.1 and 3.2 of \cite{21}, and gives conditions under which \(\tilde{\beta}\) is \(\sqrt{n}\)-consistent and asymptotically normal. Thus, asymptotically valid confidence intervals and hypothesis tests can be constructed via the bootstrap. LICQ and SC are regularity conditions commonly found in the optimization literature \cite{40, 41}; see Appendix A for the formal definitions.

The uniqueness of \(\beta^*\) simply requires that our basis functions are never perfectly collinear.

Once we obtain \(\tilde{\beta}_1\) and \(\tilde{\beta}_0\) through \((\mathbb{P}_{\mu_a}^*)\), our proposed estimator for \(g^*\) is given by

\[
\hat{g}(W) = 1 \left\{ \tilde{\beta}_1^\top b(W) > \tilde{\beta}_0^\top b(W) \right\}.
\]

### 4 Regret Bounds and Fairness-Welfare Tradeoff

Here, we analyze the welfare regret upper bounds and discuss its implication in incorporating fairness into optimal treatment regimes. To derive the upper bounds we require a margin condition, which restricts the probability that the two outcome regression functions get too close to each other in the neighborhood of \(\mu_1 = \mu_0\).

**Definition 4.1** (Margin Condition). For some \(\alpha > 0\) and for all \(t\), we have that

\[
\mathbb{P}(|\mu_1(W) - \mu_0(W)| \leq t) \lesssim t^\alpha.
\]

The above margin condition is analogous to that used in \cite{17, 22, 25, 26} as well as other problems involving estimation of non-smooth parameters such as classification \cite{2}, clustering \cite{24}.

The following result shows that the welfare regret of the propose estimator \(\hat{g}\) in (2) depends on both the nuisance estimation accuracy and the level of fairness which we would like to attain. Our result is asymptotic in the sample size \(n\).

---

\(^4\) If one is willing to rely on appropriate empirical process conditions (e.g., Donsker-type or low entropy conditions \cite{43}), then \(\gamma\) can be estimated on the same sample without sample splitting. However this would limit the flexibility of the nuisance estimators.
Theorem 4.1. Assume that the margin condition holds with the margin exponent \( \alpha \) in \((3)\) and \( \|\mu_a(W)\|_\infty < \infty \). Then under Assumptions \((A1)\) and \((A2)\) we have the asymptotic regret bounds
\[
\mathcal{R}(g^*) - \mathcal{R}(\hat{g}) = O \left( \left( \max_a \sqrt{\min_b \mathbb{E} \left[ \|Y^a - \beta^T b(W)\|_2^2 \right] | W} \|_{\infty} + R_{1,n} + R_2 \right)^n \right)
\]
where
\[
R_{1,n} = O_p \left( \|\hat{\pi}(W) - \pi(W)\|_{2,p} \max_n \|\hat{\mu}_a(W) - \mu_a(W)\|_{2,p} \vee n^{-\frac{1}{2}} \right),
\]
\[
R_2 = O \left( \sum_{a,j} \lambda_j \|g_j(Y^a, W)b(W)\|_{2,p} \right),
\]
and \( \lambda_j \geq 0 \) is the Lagrange multiplier associated with the \( j \)-th fairness constraint in \((Pa)\).

A sketch of the proof is given in Appendix \[B\]. Note that the same bound derived in the above theorem also holds for \( \mathbb{P} \{\hat{g}(W) \neq g^*(W)\} \), a probability that \( \hat{g} \) differs from the true optimal treatment policy \( g^* \) over a new observation.

The bound in Theorem \[4.1\] consists of three terms. The first term is an un-avoidable modeling error minimized through least square estimation, which will vanish if \( \mu_a(\cdot) \) lies in the function space spanned by the basis functions \( b(\cdot) \). The second term, \( R_{1,n} \), is a doubly robust second-order term that will be small if either \( \pi \) or \( \mu_a \) are estimated accurately. In nonparametric modeling, the condition \( \|\pi - \pi\|_{2,p} \|\mu_a - \mu_a\|_{2,p} = O_p(n^{-\frac{1}{2}}) \) substantially lowers the bar for the nuisance estimator convergence rate, which allows much more flexible methods to be employed while still achieving \( \sqrt{n} \) rates; for example, it suffices that these nuisance functions are estimated consistently at \( n^\frac{1}{4} \) rates.

The third term, \( R_2 \), has particularly important implications. It measures the imbalances in covariate distributions with respect to the sensitive feature, which is closely related to the level of unfairness in the optimal treatment policy; the larger the imbalances, the more likely the estimated optimal policies are unfair. If we use small values of the tolerance level \( \delta_j \) so that the optimum \( \beta^* \) is constrained by the \( j \)-th fairness constraint, then the corresponding Lagrange multiplier, \( \lambda_j \), is positive. On the contrary, if we loosen the standard by using large values of \( \delta_j \) so that the \( j \)-th fairness constraint does not constrain \( \beta^* \), \( \lambda_j \) is set to zero. Therefore, our attempts toward making optimal treatment policies more fair may lead to an additional welfare loss (regret) relative to the universally maximum feasible welfare \( \mathcal{R}(g^*) \). In other words, there is a tradeoff between fairness in the optimal treatment regime and the maximum obtainable welfare.

In short, Theorem \[4.1\] implies that although the proposed approach has considerably reduced the burden on nuisance estimation, regardless how accurately we estimate the nuisance components there is a price that comes with imposing fairness constraints for the optimal treatment regime to achieve the desired fairness level.

5 Discussion

We propose a new framework for fair and robust estimation of optimal treatment regimes. Our method is easily implementable and allows practitioners to flexibly incorporate various fairness constraints to meet the desired level of fairness. This affords new opportunities to leverage the recent development in algorithmic fairness for optimal treatment regimes.

There are two important messages in our regret bound analysis. First, the proposed estimator is robust against model misspecification and allow to use more flexible nonparametric methods while still achieving \( \sqrt{n} \) convergence rates to the maximum obtainable welfare. Second, there is a tradeoff between fairness and the maximum obtainable welfare, which is independent of accuracy of the nuisance estimation.
References


[37] Liliana Orellana, Andrea Rotnitzky, and James M Robins. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. The international journal of biostatistics, 6(2), 2010.


A  Formal Definitions of the Regularity Conditions

First, for a feasible point \( \tilde{\beta} \in \mathcal{C}_{\text{fair}} \) we define the active index set.

**Definition A.1 (Active set).** For \( \tilde{\beta} \in \mathcal{C}_{\text{fair}} \), we define the active index set \( J_0(\tilde{\beta}) \) by

\[
J_0(\tilde{\beta}) = \{1 \leq j \leq m \mid g_j(\tilde{\beta}) = 0\}.
\]

In what follows, we define LICQ and SC with respect to \( \{\mu_a\} \).

**Definition A.2 (LICQ).** Linear independence constraint qualification (LICQ) is satisfied at \( \tilde{\beta} \in \mathcal{S} \) if the vectors \( \nabla_\beta g_j(\tilde{\beta}) \), \( j \in J_0(\tilde{\beta}) \) are linearly independent.

**Definition A.3 (SC).** Let \( L(\beta, \gamma) \) be the Lagrangian. Strict Complementarity (SC) is satisfied at \( \tilde{\beta} \in \mathcal{S} \) if, with multipliers \( \bar{\gamma}_j \geq 0 \), \( j \in J_0(\tilde{\beta}) \), the Karush-Kuhn-Tucker (KKT) condition

\[
\nabla_\beta L(\tilde{\beta}, \bar{\gamma}) := \nabla_\beta L(\beta) + \sum_{j \in J_0(\tilde{\beta})} \bar{\gamma}_j \nabla_\beta g_j(\tilde{\beta}) = 0,
\]

is satisfied such that \( \bar{\gamma}_j > 0, \forall j \in J_0(\tilde{\beta}) \).

LICQ is arguably one of the most widely-used constraint qualifications that admit the first-order necessary conditions. SC means that if the \( j \)-th inequality constraint is active then the corresponding dual variable is strictly positive, so exactly one of them is zero for each \( 1 \leq j \leq m \). SC is widely used in the optimization literature, particularly in the context of parametric optimization [e.g., 40, 41].

B  Sketch of the Proof of Theorem 4.1

**Proof.** Under the margin condition, by Lemma 1 of [18] and the Cauchy–Schwarz inequality, we have

\[
\mathcal{R}(g^*) - \mathcal{R}(\tilde{\beta}) = \mathbb{P}\left[ (\mu_1(W) - \mu_0(W)) \left( \mathbb{I}\{\mu_1(W) > \mu_0(W)\} - \mathbb{I}\{\tilde{\beta}_0^\top b(W) > \tilde{\beta}_0^\top b(W)\} \right) \right]
\leq 2 \max_a \|\mu_a(W)\|_\infty \left( 2 \max_a \|\mu_a(W) - \tilde{\beta}_a^\top b(W)\|_\infty \right)^\alpha.
\]

We let \( \beta_a^* \) denote an optimal solution to the following unconstrained optimization problem

\[
\beta_a^* = \min_{\beta \in \mathbb{R}} \mathbb{E}\left\{ (Y^a - \beta^\top b(W))^2 \right\},
\]

and let \( \tilde{\beta}_a \) be an optimal solution to \( \{P_{\mu_a}\} \). Then \( \forall a \),

\[
\|\mu_a(W) - \tilde{\beta}_a^\top b(W)\|_\infty \leq \|\mu_a(W) - \beta_a^\top b(W)\|_\infty + \|b(W)\|_\infty \left\{ \|\beta_a^* - \bar{\beta}_a\|_2 + \|\bar{\beta}_a - \tilde{\beta}_a\|_2 \right\}.
\]

The unavoidable modeling error can be upper bounded in a more interpretable form by noticing

\[
\|\mu_a(W) - \beta_a^*\top b(W)\|_\infty \leq \sqrt{\min_\beta \mathbb{E}\left[ (Y^a - \beta^\top b(W))^2 \mid W \right]}\]

from the definition of the least square estimator. Further, one may show that

\[
\|\beta_a^* - \bar{\beta}_a\|_2 \leq \sum_j \lambda_j \|g_j(Y^a, W)b(W)\|_{2,\beta}
\]

by noticing that (4) and \( \{P_{\mu_a}\} \) can be viewed as the same form of a parametrized program (after writing (4) as a Lagrange dual form) and then applying the stability results [41] Chapter 6]. Finally, by Theorem 3.1 it follows that

\[
\|\bar{\beta}_a - \tilde{\beta}_a\|_2 = O_\mathbb{P}\left( n^{-\frac{1}{2}} \vee \|\tilde{\pi}(W) - \pi(W)\|_{2,\beta} \max_a \|\tilde{\mu}_a(W) - \mu_a(W)\|_{2,\beta} \right).
\]
Putting the pieces together, we obtain the desired result.

Finally, it immediately follows from Theorem 4.1 by noting that
\[ \mathbb{P} \{ \widetilde{g}(W) \neq g^*(W) \} = E \{ |\widetilde{g}(W) - g^*(W)| \} \]
\[ \leq E \left[ \mathbb{1} \left( |\mu_1(W) - \mu_0(W)| \leq \sum_a |\mu_a(W) - \tilde{\beta}_a^T b(W)| \right) \right] \]
\[ \leq \left( \max_a \| \mu_a(W) - \tilde{\beta}_a^T b(W) \|_\infty \right)^\alpha, \]
where the first inequality follows by Lemma 1 of [18].