cemmap
centre for microdata methods and practice

Inference for an algorithmic fairness-accuracy frontier

Yiqi Liu
Francesca Molinari

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP13/25

UKRI Economic and Social Research Council

INFERENCE FOR AN ALGORITHMIC FAIRNESS-ACCURACY FRONTIER

YIQI LIU
Department of Economics, Cornell University

FRANCESCA MOLINARI
Department of Economics, Cornell University

Algorithms are increasingly used to aid with high-stakes decision making. Yet, their predictive ability frequently exhibits systematic variation across population subgroups. To assess the trade-off between fairness and accuracy using finite data, we propose a debiased machine learning estimator for the fairness-accuracy frontier introduced by Liang, Lu, Mu, and Okumura (2024). We derive its asymptotic distribution and propose inference methods to test key hypotheses in the fairness literature, such as (i) whether excluding group identity from use in training the algorithm is optimal and (ii) whether there are less discriminatory alternatives to a given algorithm. In addition, we construct an estimator for the distance between a given algorithm and the fairest point on the frontier, and characterize its asymptotic distribution. Using Monte Carlo simulations, we evaluate the finite-sample performance of our inference methods. We apply our framework to re-evaluate algorithms used in hospital care management and show that our approach yields alternative algorithms that lie on the fairness-accuracy frontier, offering improvements along both dimensions.

KEYWORDS: Algorithmic fairness, statistical inference, support function.

Yiqi Liu: yl3467@cornell.edu

Francesca Molinari: fm72@cornell.edu

This draft: June 13, 2025

We thank Levon Barseghyan, Gillian Hadfield, Hiroaki Kaido, Nathan Kallus, Jens Ludwig, Chuck Manski, Alice Qi, Chen Qiu, Andres Santos, Vira Semenova, Rahul Singh, Alex Tetenov, Lars Vilhuber, Davide Viviano, reviewers for the EC24 conference, seminar participants at Chicago, Cornell, Geneva, JHU, LSE, MSU, Munich, SciencesPo, Stanford, Toulouse, UCL, Warwick, EC24, ESIF: Economics and AI+ML Meeting, the 2024 Brown University workshop "Using Data to Make Decisions," and especially José Montiel-Olea and Thomas Russell for helpful comments. All data and replication files can be accessed at github.com/yiqi-liu/TestAlgFair.

1. INTRODUCTION

Algorithms are increasingly used in many aspects of life, often to guide or support high-stake decisions, for example by predicting job performance, re-offense risk, loan default, college success, or patient health. These predictions feed, respectively, into the determination of who should be hired; which defendants should receive bail; who should be granted a loan; which students should be admitted to college; and which patients to treat. Yet, a growing body of literature documents that algorithms may exhibit bias against legally protected groups, both in their predictive accuracy and in the decisions they lead to (see, e.g., Angwin et al., 2016, Arnold et al., 2021, Obermeyer et al., 2019, Berk et al., 2021). The bias may arise, for example, due to the choice of labels the algorithm is trained on, the objective function that the algorithm optimizes, the training procedure, and many other factors involved in the design of the algorithm (see, e.g., Cowgill and Tucker, 2020).

Designing an algorithm often entails a trade-off between making it more *fair*, i.e., less likely to disproportionately harm a protected class, and more *accurate*, e.g., better at assigning treatment to those who benefit from it and withholding it from those who do not. As a result, improving fairness often comes at the cost of accuracy. Regulators, policymakers, algorithm designers, and actors affected by algorithmic predictions all have an interest in assessing various aspects of this trade-off.

We provide a set of tools for estimation of and statistical inference on a *fairness-accuracy* (FA) frontier recently characterized by Liang, Lu, Mu, and Okumura (2024, LLMO henceforth), where fairness is measured by the gap between group-specific expected losses. The theoretical analysis in LLMO assumes perfect knowledge of the population distribution of the observable variables and formalizes the trade-off between accuracy and fairness, shading light on how to use properties of the data distribution to determine whether it is optimal for the designer of the algorithm to exclude certain inputs from use. However, in practice, regulators and policymakers typically have access to only finite data. Hence, statistical inference tools are crucial for analyzing properties of algorithms and for their regulation.

We put forward a consistent estimator for LLMO's FA-frontier and derive its asymptotic distribution. For each point on the FA-frontier, we characterize an algorithm that achieves

it. We then develop a method to test hypotheses such as: Is it optimal to fully exclude group identity from use in an algorithm? Does a particular algorithm lead to group-specific expected losses that are on the FA-frontier? How far from the fairest point on the FA-frontier are the group-specific expected losses associated with a given algorithm?

Answers to the first two questions inform the regulation of algorithms and the determination of whether discrimination occurred. The law recognizes two main categories of discrimination: *disparate treatment*, where individuals are deliberately treated differently based on their membership in a protected class; and *disparate impact*, where protected classes are adversely affected disproportionately, no matter the intent (Kleinberg et al., 2018b, Blattner and Spiess, 2022). Often, as part of an effort to avoid disparate treatment, algorithms are designed so that they do not take race, gender, or other sensitive attributes as input. Even class-blind algorithms, however, may lead to disparate impact. Our first test informs a fairness-minded policymaker interested in assessing whether banning group identity has the potential to mitigate disparate impact.[1]

Our second test evaluates whether a given algorithm lies on the frontier—and thus whether a less discriminatory alternative (LDA) exists. This test is relevant to both plaintiffs (e.g., job applicants) and defendants (e.g., hiring companies) in disparate impact disputes. For example, if a selection process yields disparate impact, the hiring company may invoke business necessity to justify it. The challenger must then show the existence of an LDA, i.e., a fairer algorithm that is just as accurate. If our test rejects the null that the current algorithm is on the frontier, it supports the plaintiff's claim. Conversely, if the test fails to reject, there is no statistical evidence that the hiring company can build a fairer algorithm without sacrificing accuracy, supporting the business necessity defense. When a given algorithm is not on the frontier, we characterize alternative algorithms that improve upon it in terms of accuracy or fairness (or both).

The third inferential method yields an estimator of the distance from a given algorithm to the fairest point on the frontier and constructs a confidence interval around it. This tool

[1] This question is of interest, e.g., when assessing the recent U.S. Supreme Court decision to rule out the use of race in college admissions; for an overview, see College Board.

may interest any fairness-minded agent (e.g., a college) willing to trade some accuracy for reduced disparate impact (e.g., via affirmative action), as it provides a measure of the trade-off between promoting equity and achieving accuracy.

The key insight underlying our proposed inference methods is that since the feasible set of group-specific expected losses associated with all possible algorithms is convex, it can be fully represented by its *support function*. As the FA-frontier is a portion of the boundary of the feasible set, we characterize and estimate it through this support function. We express the hypotheses listed above as restrictions on the support function, yielding easy to understand test statistics that essentially rely on judicious use of the separating hyperplane theorem. Throughout our analysis, the support function serves as a unifying tool for inference on properties of algorithms and of the FA-frontier.

We provide a consistent debiased machine learning (DML) estimator of the support function and establish that it converges to a tight Gaussian process as sample size increases, building on and extending results in Beresteanu and Molinari (2008), Chandrasekhar et al. (2018) and Semenova (2023). We show how to allow for infimum-type test statistics that are directionally-differentiable mappings of the support function, building on results of Fang and Santos (2019). Earlier uses of the support function for inference in partially identified models (e.g., Beresteanu and Molinari, 2008, Bontemps et al., 2012, Kaido and Santos, 2014, Kaido, 2016, Chandrasekhar et al., 2018, Molinari, 2020, Semenova, 2023) did not include tests for hypotheses such as the ones we consider. Expressing these hypotheses in terms of restrictions on the support function is one of our main contributions.

We evaluate the finite-sample performance of our inference toolkit using extensive Monte Carlo simulations. We then demonstrate its empirical value by reanalyzing algorithms for high-risk care assignment at a research hospital studied by Obermeyer, Powers, Vogeli, and Mullainathan (2019). We fail to reject the hypothesis that the hospital's status-quo algorithm admits an LDA, and document fairness and accuracy gains from several alternative algorithms on the frontier that we characterize.

Related Literature. A growing literature in computer science and statistics studies algorithmic fairness; see Chouldechova and Roth (2018), Barocas et al. (2023), and Corbett-

Davies et al. (2024) for comprehensive overviews and open questions. Models have been developed to explain algorithmic bias by decomposing disparity sources (e.g., Rambachan et al., 2020a) or incorporating taste-based discrimination and unobservables in label generation (e.g., Rambachan and Roth, 2020). Fairness has been modeled as a constraint or regularizer in the objective function that maximizes predictive accuracy (e.g., Dwork et al., 2012, Berk et al., 2017) and incorporated in the preferences of a social planner that uses algorithms in their decision-making process (Kleinberg et al., 2018a, Rambachan et al., 2020b). In optimal policy targeting, fairness has been set as the criterion to be maximized when choosing a policy from the set of welfare-maximizing rules (e.g., Viviano and Bradic, 2023). When protected class membership is not observed in the data but proxy variables are available, data combination methods have been proposed to partially identify disparity measures (Kallus et al., 2022). Yet, tests of hypotheses for properties of the trade-off between fairness and accuracy of algorithms are scant in the literature. Auerbach et al. (2024) propose a test, which can incorporate exogenous constraints on the algorithm space, using sample splitting for the union null hypothesis that a status quo algorithm can be weakly improved in terms of both fairness and accuracy. Their test is based on finding another algorithm within a user-specified subclass, subject to the constraint that the alternative algorithm is at least as accurate as the status quo algorithm.

In contrast, our test for existence of an LDA is a one-shot test, valid across all algorithms rather than a specific subclass, that reveals if the status quo algorithm is on the FA-frontier and does not require first estimating an alternative algorithm. Characterizing the entire FA-frontier allows us to provide a comprehensive toolkit for statistical inference that can be useful for regulators to determine what algorithm design restrictions and reporting requirements to impose on entities making decisions using algorithms.

Outline. Section 2 lays out notation and summarizes the derivation of the FA-frontier in LLMO. Section 3 characterizes the support function of interest and uses it to describe the FA-frontier. Section 4 derives the asymptotic properties of our DML estimator for the support function. Section 5 uses these results to obtain a consistent estimator and an asymptotically valid confidence set for the FA-frontier, and characterizes algorithms attaining points

on it. Section 6 formulates hypotheses of interest in the fairness literature as restrictions on the support function and proposes asymptotically valid tests. Section 7 provides an estimator and inference method for the distance between the expected group-specific losses associated with a given algorithm and the fairest point on the frontier. Section 8 presents our Monte Carlo simulations and re-evaluation of Obermeyer et al. (2019)'s study on a research hospital's use of algorithms for assigning patients to a high-risk care management program. Section 9 concludes. Our main proofs are in Appendix A; Appendix B reports auxiliary results and extensions, and Appendix C includes supplemental empirical results.

2. SETUP

Let a population of individuals be described by an outcome $Y \in \mathcal{Y} \subset \mathbb{R}$, a binary group identity $G \in \{r, b\}$ (red or blue), and a vector of covariates $X \in \mathcal{X} \subset \mathbb{R}^{d_X}$, with the population distribution of (Y, G, X) denoted $\mathbb{P}$. For example, Y may denote an individual's number of active chronic illnesses in the subsequent year, G may denote their race, and X may include age, gender, biomarkers, comorbidity, costs and medication variables. The relation between G and X is left unspecified, but G is not part of X. Throughout, we assume G is binary, though the results extend to multiple groups. Each individual receives a binary decision $D \in \{0, 1\}$, e.g., whether they are automatically enrolled in a high-risk care management program. An algorithm $a : \mathcal{X} \mapsto [0, 1]$ assigns a probability distribution to D; e.g., the algorithm assigns each patient a health risk score in $[0, 1]$, which for simplicity we take to be the only input to the enrollment decision, and hence to coincide with the enrollment probability. Let $\mathcal{A}(\mathcal{X})$ denote the set of all algorithms that map from the input space $\mathcal{X}$ to a probability distribution over D, and $\ell : \{0, 1\} \times \mathcal{Y} \mapsto \mathbb{R}$ be a function that measures the loss associated with decision $d \in \{0, 1\}$ for an individual with outcome $y \in \mathcal{Y}$. In the example discussed so far, the algorithm designer observes training data consisting of covariates X and a binary outcome Y indicating whether someone has a number of chronic illnesses exceeding a given threshold; the loss function ℓ may be the classification loss, $\ell(D, Y) = \mathbb{1}\{D \neq Y\}$, which returns the value 1 if the algorithm mistakenly enrolls a healthy person in the high-risk care program or fails to enroll someone who is very sick. We assume throughout that the training data is drawn from the same distribution as

the population that we eventually apply the algorithm to (i.e., the subpopulation for which labels are observed is representative of the entire population).

Given an algorithm $a \in \mathcal{A}(\mathcal{X})$, let the population expected loss for group $g \in \{r, b\}$ be

$$e_g(a) \equiv \mathbb{E}\left[a(X)\ell(1, Y) + (1 - a(X))\ell(0, Y)|G = g\right], \tag{1}$$

where the expectation is taken with respect to $\mathbb{P}$. We refer to the group-specific expected loss in Eq. (1) as *group risk*. Following LLMO, we define a preference ordering over group risk pairs so that $e = (e_r, e_b)$ is preferred to $e' = (e'_r, e'_b)$, denoted $e >_{FA} e'$, if

$$e_r \leq e'_r, \quad e_b \leq e'_b, \quad \text{and} \quad |e_r - e_b| \leq |e'_r - e'_b|, \tag{2}$$

with at least one strict inequality. As shown in LLMO, all of utilitarian, Rawlsian, egalitarian, and various other preferences are consistent with this ordering. One can then define the *feasible set* of group risk pairs across algorithms from the class $\mathcal{A}(\mathcal{X})$ as

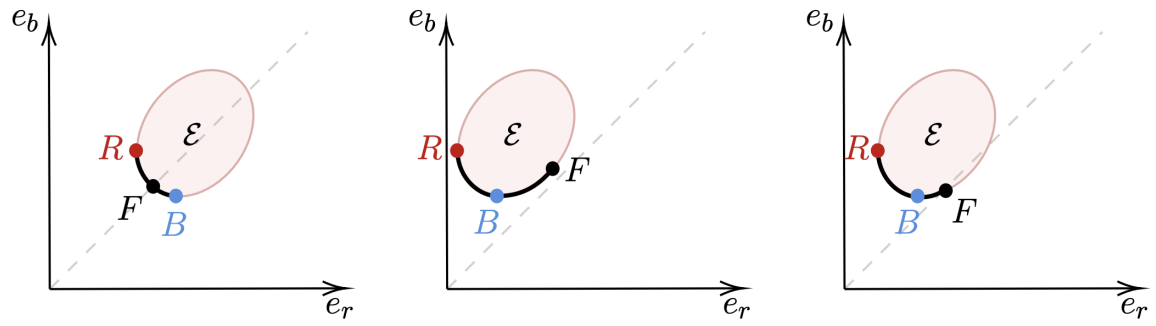
$$\mathcal{E}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big) \equiv \left\{\big(e_r(a), e_b(a)\big) \in \mathbb{R}^2 : a \in \mathcal{A}(\mathcal{X})\right\}, \tag{3}$$

and the *fairness-accuracy (FA) frontier* as

$$\mathcal{F}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big) \equiv \left\{e \in \mathcal{E}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big) : \nexists\, e' \in \mathcal{E}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big) \text{ such that } e' >_{FA} e\right\}. \tag{4}$$

For finite $(\mathcal{X}, \mathcal{Y})$, LLMO show that $\mathcal{E}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ is a closed convex set (we extend this convexity result to general $(\mathcal{X}, \mathcal{Y})$ in the proof of Proposition 3.1) and $\mathcal{F}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ is a specific portion of its boundary connecting three points: the feasible point that minimizes the risk for group r, denoted R; the feasible point that minimizes the risk for group b, denoted B, and the feasible point that minimizes the absolute difference in group risks, denoted F.[2] Adapting LLMO nomenclature, call $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ *group-balanced* if $\mathcal{E}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ has R and B such that either $R = B = F$, or $e_r < e_b$ at R and $e_r > e_b$ at B; call $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ *r-skewed* if $e_r < e_b$ at R and $e_r \leq e_b$ at B, and *b-skewed* if $e_r \geq e_b$ at R and $e_r > e_b$ at B.

[2] Ties are broken in favor of the other group's risk for R or B. If there are multiple feasible points that minimize the absolute difference in group risks, F is chosen to be the one that has the lowest risk for both groups.



(a) Group-balance (b) r-skew, no intersection (c) r-skew, intersection

FIGURE 1.—The feasible set $\mathcal{E}$ in pink and the frontier $\mathcal{F}$ in black under different configurations of $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$.

To ease notation, we drop the dependence of $\mathcal{E}$ and $\mathcal{F}$ on $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ unless explicitly needed. Figure 1 illustrates these sets and key points on them under a smoothness condition stated in Assumption 2. LLMO (Theorem 1) show that the shape of $\mathcal{F}$ depends entirely on whether $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ is group-balanced or g-skewed. If group-balanced, $\mathcal{F}$ is the curve connecting R and B, coinciding with the Pareto frontier (panel (a)); if g-skewed, $\mathcal{F}$ connects F and the feasible point minimizing risk for group g (panels (b) and (c) for the r-skewed case; omitted panels for the b-skewed case). LLMO (Proposition 6) further show that excluding group identity as an algorithmic input is uniformly welfare-reducing under strict group balance, where R and B are strictly separated by the 45-degree line.

Notation. We denote by $\|\cdot\|_E$, $\|\cdot\|_{L^2(\mathbb{P})}$, $\|\cdot\|_\infty$, respectively, the Euclidean norm, the L^2-norm under the probability measure $\mathbb{P}$, and the L^∞-norm (or sup-norm). For a vector $\mathbf{a}$, let $\|\mathbf{a}\|_{L^2(\mathbb{P})} \equiv \big\|\|\mathbf{a}\|_E\big\|_{L^2(\mathbb{P})}$ and $\|\mathbf{a}\|_\infty$ be the supremum over the largest component of $\mathbf{a}$. For a matrix $\mathbf{A}$, let $\|\mathbf{A}\|_{\max}$ denote its max norm (the maximum absolute value among its entries). For two sequences a_n and b_n, $a_n \lesssim b_n$ means $a_n \leq c \cdot b_n$ for some constant $c > 0$.

3. SUPPORT FUNCTION BASED CHARACTERIZATIONS

We leverage the convexity of the feasible set $\mathcal{E}$ to characterize it by its support function and express the points R, B, F, and the FA-frontier $\mathcal{F}$ in Eq. (4) through this support function. We begin by observing that Eq. (1) and the law of iterated expectations yield

$$e_g(a) = \mathbb{E}\left[a(X)\mathbb{E}[\ell(1,Y)\mathbb{1}(G=g)|X] + (1-a(X))\mathbb{E}[\ell(0,Y)\mathbb{1}(G=g)|X]\right]/\mathbb{P}(G=g)$$

$$\equiv \mathbb{E}\left[a(X)\frac{\theta_1^g(X)}{\mu_g} + (1-a(X))\frac{\theta_0^g(X)}{\mu_g}\right], \tag{5}$$

where we denote $\theta_d^g(X) \equiv \mathbb{E}[\ell(d,Y)\mathbb{1}\{G=g\}|X]$ the (measurable) conditional expectation of $L_d^g \equiv \ell(d,Y)\mathbb{1}\{G=g\}$ given X; $\mu_g \equiv \mathbb{P}(G=g)$ the population proportion of group $g \in \{r,b\}$; and the expectation in Eq. (5) is taken with respect to the population marginal distribution of the covariates, $\mathbb{P}(X)$. To make sure that Eq. (5) is well defined, we assume:

ASSUMPTION 1—(Moment Restrictions): *For some constants* $0 < c_1 < 1$ *and* $0 < c_2 < \infty$, $\mu_g \in (c_1, 1-c_1)$ *and* $\operatorname{ess\,sup}_{X\in\mathcal{X}} \mathbb{E}\left[\left(L_d^g\right)^2 \middle| X\right] < c_2$, *for all* $d \in \{0,1\}, g \in \{r,b\}$.

Throughout, we let $\boldsymbol{\theta}(X) \equiv [\theta_1^r(X)\ \ \theta_0^r(X)\ \ \theta_1^b(X)\ \ \theta_0^b(X)]^\intercal$ and

$$\boldsymbol{\theta}_d(X) \equiv [\theta_d^r(X)\ \ \theta_d^b(X)]^\intercal, \quad d \in \{0,1\}, \tag{6}$$

$$\mathcal{M} \equiv \operatorname{diag}(1/\mu_r, 1/\mu_b). \tag{7}$$

3.1. *Support Function of the Feasible Set*

Given Eqs. (5)-(6)-(7), $\mathcal{E}$ can be written as

$$\begin{aligned}\mathcal{E} &\equiv \left\{\left(e_r(a), e_b(a)\right) \in \mathbb{R}^2 : a \in \mathcal{A}(\mathcal{X})\right\} \\ &= \left\{\mathbb{E}[\mathcal{M}\vartheta(X)] : \vartheta(X) \in \operatorname{conv}\left(\{\boldsymbol{\theta}_0(X), \boldsymbol{\theta}_1(X)\}\right)\right\} = \mathbf{E}\left[\mathcal{M}\Lambda(X)\right],\end{aligned} \tag{8}$$

with $\operatorname{conv}(\cdot)$ the convex hull of the set in parentheses, $\Lambda(X) \equiv \operatorname{conv}\left(\{\boldsymbol{\theta}_0(X), \boldsymbol{\theta}_1(X)\}\right)$ a random interval, $\mathcal{M}$ in Eq. (7), and $\mathbf{E}\left[\mathcal{M}\Lambda(X)\right]$ the *Aumann expectation* of the scaled random interval $\mathcal{M}\Lambda(X)$ (Molchanov and Molinari, 2018, Example 1.11 and Def. 3.1).

As the set $\mathcal{E}$ is non-empty, compact, and convex, its *support function* in each direction $q = [q_1\ q_2]^\intercal \in \mathbb{S}^1 \equiv \{v \in \mathbb{R}^2 : \|v\|_E = 1\}$, defined as

$$h_{\mathcal{E}}(q) \equiv \max_{e\in\mathcal{E}} q^\intercal e,$$

uniquely characterizes $\mathcal{E}$ through the identity (Rockafellar, 1997, Chapter 13)

$$\mathcal{E} = \bigcap_{q\in\mathbb{S}^1} \left\{z \in \mathbb{R}^2 : q^\intercal z \le h_{\mathcal{E}}(q)\right\}. \tag{9}$$

We next provide a closed-form expression for $h_{\mathcal{E}}(q)$.

PROPOSITION 3.1: *Let Assumption 1 hold. Then:*

$$\begin{aligned} h_{\mathcal{E}}(q) &= \mathbb{E}\left[\max\{(\mathcal{M}q)^{\intercal}\boldsymbol{\theta}_0(X), (\mathcal{M}q)^{\intercal}\boldsymbol{\theta}_1(X)\}\right] \\ &= \mathbb{E}\big[(\mathcal{M}q)^{\intercal}\mathbf{L}_0 + (\mathcal{M}q)^{\intercal}(\mathbf{L}_1 - \mathbf{L}_0)\mathbb{1}\{k(\boldsymbol{\theta}(X), \mathcal{M}q) > 0\}\big], \end{aligned} \tag{10}$$

where $k(\boldsymbol{\theta}, \mathcal{M}q) \equiv (\mathcal{M}q)^{\intercal}\big(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)\big)$, $\mathbf{L}_d \equiv [L_d^r \;\; L_d^b]^{\intercal}$, *and* $L_d^g \equiv \ell(d, Y)\mathbb{1}\{G = g\}$.

The support function $h_{\mathcal{E}}(q)$ is our key inferential tool. One can garner the intuition behind its closed-form expression by rewriting $e_g(a) = \mathbb{E}\left[\frac{\theta_0^g(X)}{\mu_g}\right] + \mathbb{E}\left[a(X)\frac{\theta_1^g(X) - \theta_0^g(X)}{\mu_g}\right]$ and

$$h_{\mathcal{E}}(q) = \mathbb{E}\left[q_1\frac{\theta_0^r(X)}{\mu_r} + q_2\frac{\theta_0^b(X)}{\mu_b}\right] + \max_{a \in \mathcal{A}(\mathcal{X})} \mathbb{E}\left[a(X)k\left(\boldsymbol{\theta}(X), \mathcal{M}q\right)\right].$$

The maximum in the above expression is achieved by the algorithm $a^{\texttt{opt}}(X; q) = \mathbb{1}\{k(\boldsymbol{\theta}(X), \mathcal{M}q) > 0\}$, yielding Eq. (10) upon applying the law of iterated expectations.

REMARK 3.1: We allow for *randomized decision rules* and for $\mathcal{A}(\mathcal{X})$ to be unrestricted. If instead the family of algorithms is restricted a priori (e.g., by capacity constraints) so that $a(X) = \Pr(D = 1|X) \in [\underline{a}(X), \bar{a}(X)]$, $0 \leq \underline{a}(X) \leq \bar{a}(X) \leq 1$, with $\underline{a}(\cdot), \bar{a}(\cdot)$ known functions, our analysis continues to apply by replacing $\{\boldsymbol{\theta}_0(X), \boldsymbol{\theta}_1(X)\}$ with $\{\bar{a}(X)\boldsymbol{\theta}_0(X) + (1 - \bar{a}(X))\boldsymbol{\theta}_1(X), \underline{a}(X)\boldsymbol{\theta}_0(X) + (1 - \underline{a}(X))\boldsymbol{\theta}_1(X)\}$. In Appendix B.1, we also show that our results continue to hold if one restricts attention to *threshold rules* of the form $D = \mathbb{1}\{a(X) \geq 0\}$ for unrestricted $a : \mathcal{X} \mapsto \mathbb{R}$ (and in fact $a^{\texttt{opt}}(X; q)$ is a threshold rule), or to *linear threshold rules* where $a(X) = [1; \; X]^{\intercal}\beta$ for some $\beta \in \mathbb{R}^{d_X+1}$, provided the space of algorithms is sufficiently rich (see Assumption B.1).

3.2. *Best Group-Specific Points on the FA-Frontier*

We next define the *support set* of $\mathcal{E}$ in direction $q \in \mathbb{S}^1$:

$$\mathcal{S}_{\mathcal{E}}(q) \equiv \mathcal{E} \cap \left\{z \in \mathbb{R}^2 : q^{\intercal}z = h_{\mathcal{E}}(q)\right\}, \tag{11}$$

i.e., $\mathcal{S}_{\mathcal{E}}(q)$ is the intersection between $\mathcal{E}$ and the hyperplane with normal vector q and constant $h_{\mathcal{E}}(q)$, hence collecting the extreme point(s) of $\mathcal{E}$ in direction q. To derive a closed-form expression for $\mathcal{S}_{\mathcal{E}}(q)$, we impose the following assumption:

ASSUMPTION 2—(Margin Condition): *There exists* $0 < m \leq 1$ *such that for any* $\delta > 0$, $\sup_{q \in \mathbb{S}^1} \mathbb{P}(|k(\boldsymbol{\theta}(X), \mathcal{M}q)| < \delta) \lesssim \delta^m$, *with the probabilities taken with respect to* $\mathbb{P}(X)$.

Assumption 2 is a margin condition that guarantees sufficient smoothness in the distribution of $\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)$ for us to show that $h_{\mathcal{E}}(q)$ is differentiable in $q \in \mathbb{S}^1$ and consequently $\mathcal{S}_{\mathcal{E}}(q)$ includes a single element in each direction q (Schneider, 1993, Corollary 1.7.3). We further show that $\mathcal{S}_{\mathcal{E}}(q)$ equals the gradient of the support function $h_{\mathcal{E}}(\cdot)$ with respect to q. We denote by $\mathcal{S}_{\mathcal{E}}(q)$ both the singleton set and its only element.

PROPOSITION 3.2: *Let Assumptions 1-2 hold. Then,*

$$\nabla_q h_{\mathcal{E}}(q) = \mathbb{E}\big[\mathcal{M}\mathbf{L}_0 + \mathcal{M}(\mathbf{L}_1 - \mathbf{L}_0)\mathbb{1}\{k(\boldsymbol{\theta}(X), \mathcal{M}q) > 0\}\big] = \mathcal{S}_{\mathcal{E}}(q), \tag{12}$$

uniformly in $q \in \mathbb{S}^1$, *where* $\mathcal{S}_{\mathcal{E}}(q)$ *is uniformly continuous in* $q \in \mathbb{S}^1$.

Let $\mathfrak{u}_1 \equiv [-1 \;\; 0]^\intercal$ and $\mathfrak{u}_2 \equiv [0 \;\; -1]^\intercal$. The best group-specific points satisfy:

$$R = \mathcal{S}_{\mathcal{E}}(\mathfrak{u}_1) \quad \text{and} \quad B = \mathcal{S}_{\mathcal{E}}(\mathfrak{u}_2). \tag{13}$$

REMARK 3.2: Assumption 2 allows for discrete covariates (see Proposition B.1), but is violated if X includes discrete covariates *only* (see Appendix B.1), in which case we can use data jittering to satisfy Assumption 2 by adding to one discrete covariate a small amount of smoothly distributed noise, thereby garbling that input. The feasible set constructed with a jittered covariate can be made arbitrarily close to the true feasible set (this result can be proved by adapting arguments in Chandrasekhar et al., 2018, Lemma 8).

REMARK 3.3: The argument in Bontemps et al. (2012, Supplemental Appendix B.2.3) shows that $\mathcal{E}$ has no *kinks* (i.e., no support points such that there exist at least two distinct vectors q and v satisfying $\mathcal{S}_{\mathcal{E}}(q) = \mathcal{S}_{\mathcal{E}}(v)$) if and only if for any $q, v \in \mathbb{S}^1, q \neq v$,

$$\mathbb{P}\big(k(\boldsymbol{\theta}(X), \mathcal{M}q) > 0, k(\boldsymbol{\theta}(X), \mathcal{M}v) < 0\big) > 0. \tag{14}$$

If $(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X))$ admits a positive density function on a ball of positive radius that includes zero, Eq. (14) is satisfied. Assumption B.2 in Appendix B is an example of low level conditions yielding this result. The absence of kinks renders simpler limit distributions for the test statistics that we put forward in Sections 5-7. Nonetheless, Eq. (14) is *not* needed for our results to apply and we provide a full treatment allowing for the presence of kinks.

3.3. *Fairest Point on the FA-Frontier*

Determining the coordinates of the fairest point F is more laborious, as they depend on whether $\mathcal{E}$ lies entirely above, entirely below, or on top of the 45-degree line. Figure 2 illustrates all possible locations of the feasible set $\mathcal{E}$ relative to the 45-degree line. When $\mathcal{E}$ lies entirely on one side of the 45-degree line, as depicted in panels (b) and (d), F is the support set of $\mathcal{E}$, respectively, in directions $\mathfrak{u}_2 - \mathfrak{u}_1 = [1 \;\; -1]^\intercal$ and $\mathfrak{u}_1 - \mathfrak{u}_2 = [-1 \;\; 1]^\intercal$:

$$F = \mathcal{S}_{\mathcal{E}}(\mathfrak{u}_2 - \mathfrak{u}_1) \quad \text{when } \mathcal{E} \text{ lies entirely above the 45-degree line,} \tag{15}$$

$$F = \mathcal{S}_{\mathcal{E}}(\mathfrak{u}_1 - \mathfrak{u}_2) \quad \text{when } \mathcal{E} \text{ lies entirely below the 45-degree line.} \tag{16}$$

Complications arise when $\mathcal{E}$ intersects with the 45-degree line, as depicted in panels (a), (c), and (e) of Figure 2. In this case, the direction at which we can obtain F as the support set of $\mathcal{E}$ is difficult to determine. To circumvent this challenge, we propose a different approach. We focus on the convex set that results when $\mathcal{E}$ intersects the 45-degree line:

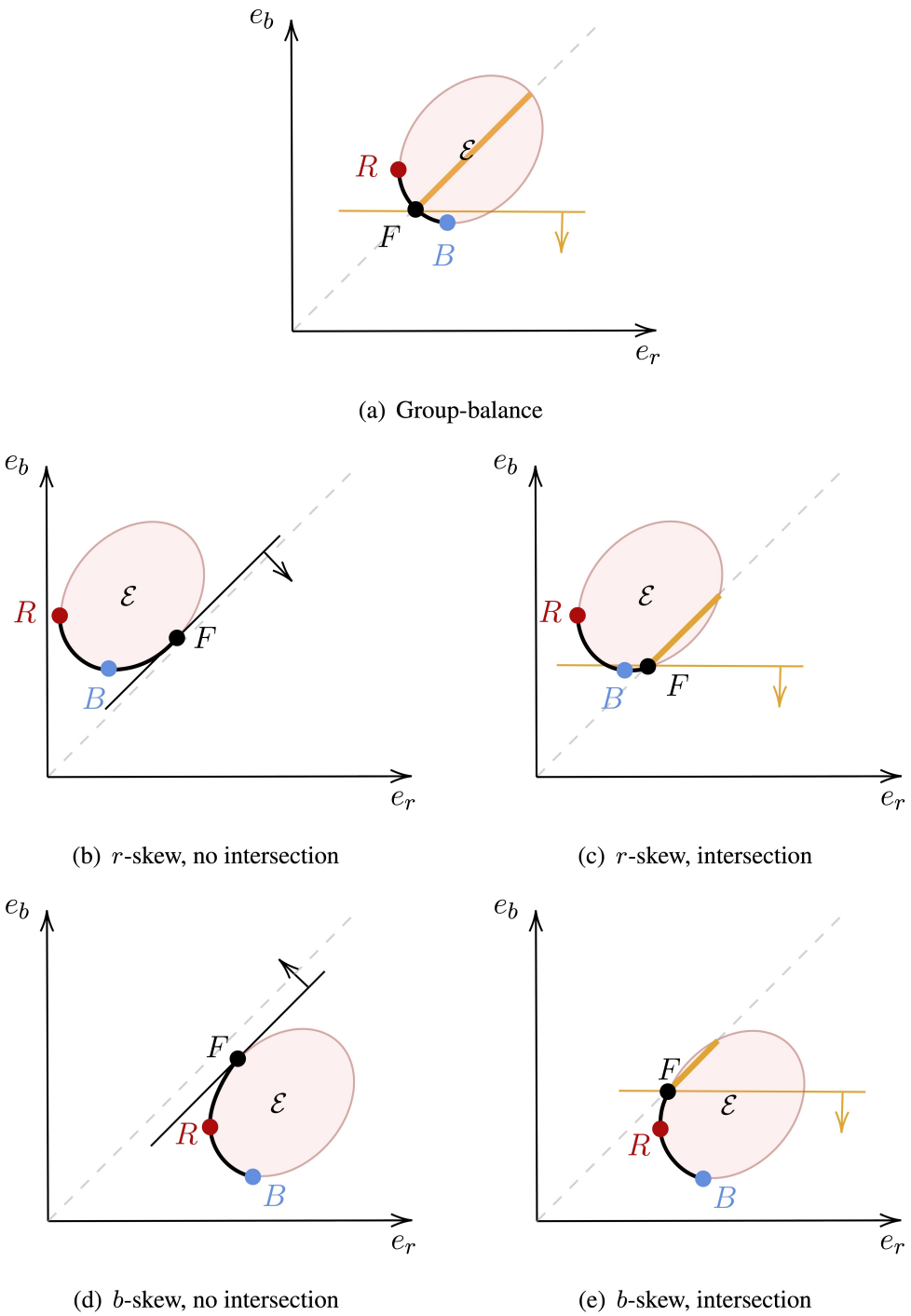
$$\tilde{\mathcal{E}} \equiv \mathcal{E} \cap \mathcal{H}_{45}, \quad \text{where} \quad \mathcal{H}_{45} \equiv \{e \in \mathbb{R}^2 : e_r = e_b\}. \tag{17}$$

The new set $\tilde{\mathcal{E}}$ is depicted in panels (a), (c), and (e) of Figure 2 as an orange line segment. In these cases, F is the support set of $\tilde{\mathcal{E}}$ in direction $\mathfrak{u}_1$ with identical values for its two coordinates. Hence,

$$F = (\mathfrak{u}_1 + \mathfrak{u}_2) h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1) \quad \text{when } \mathcal{E} \text{ intersects with the 45-degree line.} \tag{18}$$

We are left with providing an expression for $h_{\tilde{\mathcal{E}}}(q)$. When $\mathcal{E}$ intersects $\mathcal{H}_{45}$,

$$h_{\tilde{\mathcal{E}}}(q) = \inf_{p_1, p_2 \in \mathbb{R}^2 : p_1 + p_2 = q} h_{\mathcal{E}}(p_1) + h_{\mathcal{H}_{45}}(p_2) = \inf_{c \in \mathbb{R}} h_{\mathcal{E}} \left(q - c \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right), \tag{19}$$



(a) Group-balance

(b) r-skew, no intersection

(c) r-skew, intersection

(d) b-skew, no intersection

(e) b-skew, intersection

FIGURE 2.—All possible locations of the feasible set $\mathcal{E}$ relative to the 45° line, $\mathcal{H}_{45}$. In panels (a), (c), and (e), $\mathcal{E}$ intersects with $\mathcal{H}_{45}$, and the fairest point F is the support set of $\tilde{\mathcal{E}}$ in direction u_1, where $\tilde{\mathcal{E}}$ is the intersection between $\mathcal{E}$ and $\mathcal{H}_{45}$ (depicted as an orange line segment). In panels (b) and (d), $\mathcal{E} \cap \mathcal{H}_{45} = \emptyset$, and F is the support set of $\mathcal{E}$ in direction $u_2 - u_1$ for the r-skewed case in (b) and $u_1 - u_2$ for the b-skewed case in (d).

where the first equality follows from Rockafellar (1997, Corollary 16.4.1), and the second follows from the fact that $h_{\mathcal{H}_{45}}(p_2)$ is bounded from above only along the direction $p_2 = c[1 \ -1]^\intercal$ for any scalar $c \in \mathbb{R}$, in which case $h_{\mathcal{H}_{45}}(p_2) = 0$. Importantly, the last expression in Eq. (19) is always well defined, regardless of whether $\mathcal{E}$ intersects with $\mathcal{H}_{45}$ or not; the infimum equals a bounded scalar in the case of intersection and $-\infty$ otherwise.[3]

3.4. *Support Function-Based Characterization of the FA-Frontier*

We next show that the FA-frontier put forward by LLMO and reproduced in our Eq. (4) can be characterized using the support function of the feasible set $\mathcal{E}$ and that of an auxiliary set that we introduce in this subsection.

Given an algorithm $a^* \in \mathcal{A}(\mathcal{X})$ that induces the risk pair $e^* = [e_r^*, e_b^*]^\intercal \in \mathcal{E}$, let

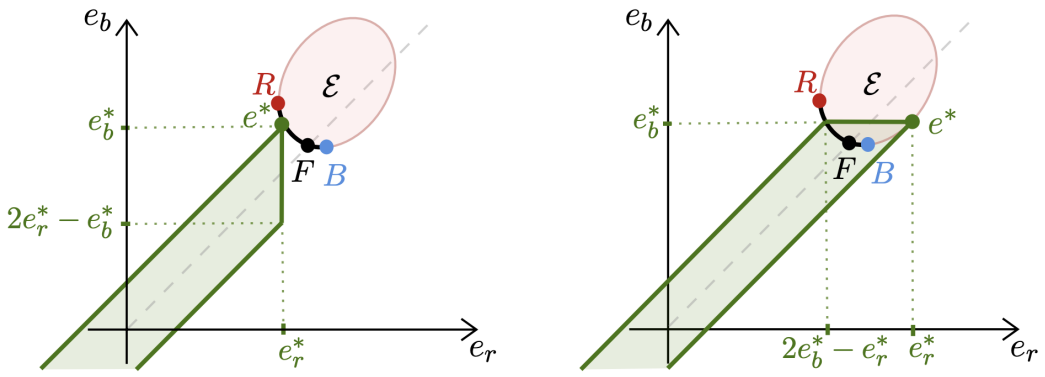
$$\mathcal{C}(e^*) = \left\{ e \in \mathbb{R}^2 : e_r \leq e_r^*, e_b \leq e_b^*, |e_r - e_b| \leq |e_r^* - e_b^*| \right\}, \tag{20}$$

denote the set of risk allocations $e \in \mathbb{R}^2$—whether or not they are feasible—that are both weakly more accurate and weakly fairer than e^*. The set $\mathcal{C}(e^*)$, depicted in the two panels of Figure 3 as the shaded green regions corresponding to two different values of e^*, is a closed and convex subset of $\mathbb{R}^2$. Panel (a) depicts a case where $e^* \in \mathcal{F}$, whereas panel (b) depicts a case where $e^* \notin \mathcal{F}$. The key insight from the figure, which we prove can be used to characterize the FA-frontier using the support function of $\mathcal{E}$ and that of $\mathcal{C}(e^*)$, is that for any $e^* \in \mathcal{F}$, the sets $\mathcal{E}$ and $\mathcal{C}(e^*)$ can be properly separated (see, e.g., Schneider, 1993, p.12, for a definition of proper separation), while in the case where $e^* \notin \mathcal{F}$ they cannot.

PROPOSITION 3.3: *Under Assumptions 1-2, $e^* \in \mathcal{F}$ if and only if there exists a hyperplane that properly separates $\mathcal{C}(e^*)$ and $\mathcal{E}$, i.e., there exists $q \in \mathbb{S}^1$ such that $h_{\mathcal{C}^*}(q) = -h_{\mathcal{E}}(-q)$. Let $\tilde{\mathbb{S}}^1 \equiv \mathbb{S}^1 \setminus \{q \in \mathbb{S}^1 : q_1 + q_2 < 0\}$ and $[\,\cdot\,]_- \equiv -\min\{\cdot, 0\}$. We then have that*

$$\mathcal{F} = \left\{ e^* \in \mathcal{E} : \left[\max_{q \in \tilde{\mathbb{S}}^1} (-h_{\mathcal{C}(e^*)}(q) - h_{\mathcal{E}}(-q)) \right]_- = 0 \right\}. \tag{21}$$

[3] To see this, let $q(c) \equiv q - c[1 \ \ -1]^\intercal$ and note that $h_{\mathcal{E}}(q(c)) = \|q(c)\|_E \cdot h_{\mathcal{E}}\left(\frac{q(c)}{\|q(c)\|_E}\right)$. When $\mathcal{E}$ intersects with $\mathcal{H}_{45}$, $h_{\mathcal{E}}\left(\frac{q(c)}{\|q(c)\|_E}\right)$ is bounded and nonnegative along the sequences $c \to \infty$ and $c \to -\infty$, but when $\mathcal{E}$ and $\mathcal{H}_{45}$ are disjoint, it takes negative value along one of these sequences, yielding $\inf_c h_{\mathcal{E}}(q(c)) = -\infty$.



(a) e^* is above the 45° line and on the frontier

(b) e^* is below the 45° line and off the frontier

FIGURE 3.—The set $\mathcal{C}(e^*)$, which collects all improvements relative to e^*, is the region shaded in green. Its shape depends on whether e^* lies above or below the 45° line. Panel (a) shows an example of the case where e^* lies on the frontier and there exists a hyperplane that properly separates $\mathcal{C}(e^*)$ and $\mathcal{E}$, whereas in panel (b) e^* is not on the frontier and no hyperplane can separate $\mathcal{C}(e^*)$ and $\mathcal{E}$.

Remarkably, this characterization of the FA-frontier does not require knowledge of whether $\mathcal{E}$ intersects the 45^o line, or whether there is group balance or skew. As we show in Sections 5-6, the characterization of the FA-frontier in Eq. (21) is very helpful for estimation and inference, and for testing whether there exists an LDA to a given algorithm.

4. SUPPORT FUNCTION ESTIMATOR AND ITS ASYMPTOTIC DISTRIBUTION

In practice, the policymaker does not have perfect knowledge of $\mathbb{P}$. Hence, $h_{\mathcal{E}}(q)$ can only be estimated from a finite sample. Let a sample of size n, $\{(Y_i, G_i, X_i)\}_{i=1}^n$, drawn independently and identically from $\mathbb{P}$, be available. Recall Eq. (10): $h_{\mathcal{E}}(q) = \mathbb{E}\big[(\mathcal{M}q)^\intercal \mathbf{L}_0 + (\mathcal{M}q)^\intercal(\mathbf{L}_1 - \mathbf{L}_0)\mathbb{1}\{k(\boldsymbol{\theta}(X), \mathcal{M}q) > 0\}\big]$ with $k(\boldsymbol{\theta}, \mathcal{M}q) \equiv (\mathcal{M}q)^\intercal\big(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)\big)$, so that $\boldsymbol{\theta}(X) \equiv [\theta_1^r(X)\ \theta_0^r(X)\ \theta_1^b(X)\ \theta_0^b(X)]^\intercal$ enters the expression for $h_{\mathcal{E}}(q)$ only through $(\theta_1^r(X) - \theta_0^r(X))$ and $(\theta_1^b(X) - \theta_0^b(X))$. We propose estimating $h_{\mathcal{E}}(q)$ by first estimating the finite dimensional parameters $\mathcal{M}$ by sample averages and the nuisance functions $\Delta\boldsymbol{\theta}(X) \equiv [(\theta_1^r(X) - \theta_0^r(X))\ (\theta_1^b(X) - \theta_0^b(X))]^\intercal$ by flexible machine learning methods (allowing the complexity of the parameter space containing the estimator to grow with sample size), and then plugging their estimators, denoted $\widehat{\mathcal{M}}$ and $\widehat{\Delta\boldsymbol{\theta}}(X)$, into the sample analogue of Eq. (10). Following the literature on debiased machine learning (e.g., Newey, 1994,

Chernozhukov et al., 2018, Semenova and Chernozhukov, 2021, Chernozhukov et al., 2022, Ichimura and Newey, 2022), we show that, at the population $\mathcal{M}$, the moment in Eq. (10) is Neyman-orthogonal and hence "insensitive" to the errors in the first-stage estimation of $\Delta\boldsymbol{\theta}$ (Neyman, 1979, 1959). We use sample splitting to relax the otherwise needed Donsker condition that limits the complexity of the relevant parameter space (e.g., Bickel, 1982, Robins et al., 2008, 2017), and we account for the estimation error of $\mathcal{M}$.

Recall that $\mathbf{L}_d \equiv [L_d^r \;\; L_d^b]^\intercal$, with $L_d^g \equiv \ell(d,Y)\mathbb{1}\{G=g\}$, and $\theta_d^g(X) = \mathbb{E}[L_d^g|X]$. Let $\Delta L^g \equiv L_1^g - L_0^g$, so that $\Delta\theta^g(X) = \mathbb{E}[\Delta L^g|X]$. To learn the nuisance parameter $\Delta\theta^g(X)$, the effective label that, given X, we train machine learners to predict is ΔL^g. Let Θ denote the convex nuisance parameter space (a subset of a vector space with $L^2(\mathbb{P})$ norm) to which $\Delta\boldsymbol{\theta} = [\Delta\theta^g(X) \;\; \Delta\theta^b(X)]^\intercal$ belongs and $\Delta\boldsymbol{\vartheta} \equiv [\Delta\vartheta^r(X) \;\; \Delta\vartheta^b(X)]^\intercal$ be a generic element from Θ (to simplify notation, we drop the dependence of $\boldsymbol{\theta}$ on X unless explicitly needed). For the i-th observation and a given 2×2 diagonal matrix $\mathring{\mathcal{M}}$, we note that

$$k\big(\boldsymbol{\vartheta}(X_i), \mathring{\mathcal{M}}q\big) = q^\intercal \mathring{\mathcal{M}} \Delta\boldsymbol{\vartheta}(X_i), \tag{22}$$

and using Eq. (22) to recognize that we estimate $\Delta\boldsymbol{\theta}$ while keeping the notation as close as possible to that in Eq. (10), we define the mapping $\zeta_i(\mathring{\mathcal{M}}q;\,\cdot\,) : \Theta \to \mathbb{R}$ as

$$\zeta_i(\mathring{\mathcal{M}}q;\boldsymbol{\vartheta}) \equiv (\mathring{\mathcal{M}}q)^\intercal \mathbf{L}_{0_i} + (\mathring{\mathcal{M}}q)^\intercal(\mathbf{L}_{1_i} - \mathbf{L}_{0_i}) \cdot \mathbb{1}\big\{k\big(\boldsymbol{\vartheta}(X_i), \mathring{\mathcal{M}}q\big) > 0\big\}. \tag{23}$$

By Proposition 3.1, $h_{\mathcal{E}}(q) = \mathbb{E}[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})]$. We next show that, when $\mathring{\mathcal{M}}$ is fixed at the population $\mathcal{M}$, the score function $\zeta_i(\mathcal{M}q;\boldsymbol{\vartheta})$ is Neyman-orthogonal at $\Delta\boldsymbol{\vartheta} = \Delta\boldsymbol{\theta}$.

PROPOSITION 4.1: *Let Assumptions 1-2 hold and* $\sup_{\Delta\boldsymbol{\vartheta}\in\Theta} \|\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta}\|_{L^2(\mathbb{P})} < \infty$. *Then the map* $\Delta\boldsymbol{\vartheta} \mapsto \mathbb{E}[\zeta_i(\mathcal{M}q;\boldsymbol{\vartheta})]$ *satisfies the Neyman orthogonality condition uniformly in* $q \in \mathbb{S}^1$, *i.e., for any* $\Delta\boldsymbol{\vartheta} \in \Theta$ *and scalar* $t \in (0,1)$,

$$\lim_{t\to 0} \sup_{q\in\mathbb{S}^1} \left| \frac{1}{t}\Big(\mathbb{E}\big[\zeta_i\big(\mathcal{M}q;\boldsymbol{\theta} + t(\boldsymbol{\vartheta}-\boldsymbol{\theta})\big)\big] - \mathbb{E}\big[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})\big]\Big)\right| = 0.$$

Intuitively, Proposition 4.1 shows that, under its maintained assumptions, the first-order mistake in the sign of $k(\boldsymbol{\theta}(X), \mathcal{M}q)$ due to the estimation error in $\Delta\boldsymbol{\theta}(X)$ is negligible.

We next show that, provided $n^{1/4}$-consistent first-stage estimators are available, the estimated support function using sample splitting and cross-fitting, as described in Definition 1 below, converges to a Gaussian process uniformly in $q \in \mathbb{S}^1$. The proof requires showing the residual in estimating the indicator functions is bounded by the quadratic rate of convergence of the nuisance parameter, which we establish under this assumption:

ASSUMPTION 3—(Nuisance Parameter Structure): *There is a known partition of* X,

$$X = (X_1, X_2, X_{[3:d_X]}),$$

where $X_1, X_2 \in \mathbb{R}$ *and* $X_{[3:d_X]} \in \mathbb{R}^{(d_X-2)}$ *are such that* (X_1, X_2) *has a bounded support, the density of* $|k(\boldsymbol{\theta}, \mathcal{M}q)|$ *conditional on* $X_{[3:d_X]}$ *is uniformly bounded in* $q \in \mathbb{S}^1$, *and*

$$\Delta\theta^g(X) = \alpha^g X_1 + \beta^g X_2 + \eta^g(X_{[3:d_X]}),$$

for some $\alpha^g, \beta^g \in \mathbb{R}$ *satisfying* $\alpha^b \cdot \beta^r \neq \alpha^r \cdot \beta^b$ *and* $\eta^g \in H$ *for convex* $H \subseteq \Theta$, $g \in \{r, b\}$.

Assumption 3 requires $\Delta\boldsymbol{\theta}(X)$ to be linearizable in a known set of covariates (X_1, X_2) and that $\alpha^b \cdot \beta^r \neq \alpha^r \cdot \beta^b$. This assures that for each $q \in \mathbb{S}^1$, $k(\boldsymbol{\theta}, \mathcal{M}q)$ depends on at least one of X_1 or X_2, so that we can employ a proof technique similar to that in Semenova (2023, Lemma 4.1) to show that the bias induced by errors in estimating the sign of $k(\boldsymbol{\theta}, \mathcal{M}q)$ is bounded by the desired quadratic L^2-rate of the nuisance estimator, accommodating a menu of flexible machine learners in the first stage. For example, one can adapt the method in Robinson (1988, pp. 935-936) to estimate $\Delta\theta^g$. Example machine learning methods that, under suitable conditions, are $n^{1/4}$-consistent in the L^2-norm include ℓ_1-penalized methods, boosting, regression trees and forests, and neural nets (see, e.g., Chernozhukov et al., 2018, and references therein, for a discussion of machine learners compatible with the rate requirement). Assumption 3 can be eliminated at the price of using a more restrictive class of machine learning methods, by deriving rate bounds that depend on the squared L^∞-rate of the first-stage estimation (e.g., the Lasso, for which an L^∞-rate is established for approximately sparse models, see Belloni et al., 2017, Semenova, 2023). Alternatively, Assumption 3 can be eliminated by smoothing the indicators (e.g., Chen et al., 2023, Park,

2024) at the cost of introducing an additional tuning parameter that controls the degree of smoothing.[4] Importantly, Assumption 3 implies the margin condition in Assumption 2 with $m = 1$ whenever (X_1, X_2) is continuously distributed with a bounded density and $X_{[3:d_X]}$ can be either continuous or discrete; see Assumption B.2 and Proposition B.1.

We next define cross-fitting, adapting Definition 3.2 in Chernozhukov et al. (2018):

DEFINITION 1—Cross-Fitting: *(i) Randomly partition the size-n sample with observations indexed by $i \in [n] \equiv \{1, ..., n\}$ to $K \geq 2$ subsamples, each of size n/K (assumed to be an integer), where K is a fixed integer. (ii) For each partition $k \in [K] \equiv \{1, ..., K\}$ with observations indexed by the set $I_k \subset [n]$, estimate $\Delta\boldsymbol{\theta}$ by $\widehat{\Delta\boldsymbol{\theta}}_k \equiv [(\widehat{\Delta\theta}^r)_k \ (\widehat{\Delta\theta}^b)_k]^\intercal$, where each $(\widehat{\Delta\theta}^g)_k = (\widehat{\alpha}^g)_k X_1 + (\widehat{\beta}^g)_k X_2 + (\widehat{\eta}^g)_k (X_{[3:d_X]})$ is estimated using only observations from $I_k^c \equiv [n] \backslash I_k$. For $i \in I_k$, let $\widehat{\Delta\boldsymbol{\theta}}(X_i) \equiv \widehat{\Delta\boldsymbol{\theta}}_k(X_i)$. (iii) Let $\widehat{\mathcal{M}} = \text{diag}(1/\widehat{\mu}_r, 1/\widehat{\mu}_b)$, with $\widehat{\mu}_g \equiv \frac{1}{n}\sum_{i=1} \mathbb{1}\{G_i = g\}$, and construct the second-stage estimator as*

$$\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}}) = \frac{1}{K} \sum_{k \in [K]} \left(\frac{1}{n/K} \sum_{i \in I_k} \zeta_i(\widehat{\mathcal{M}} q; \widehat{\boldsymbol{\theta}}_k) \right) \equiv \frac{1}{n} \sum_{i=1}^{n} \zeta_i(\widehat{\mathcal{M}} q; \widehat{\boldsymbol{\theta}}). \tag{24}$$

In Eq. (24), to simplify notation and keep it as close as possible to that in Eq. (10) and $h_{\mathcal{E}}(q) = \mathbb{E}[\zeta_i(\mathcal{M} q; \boldsymbol{\theta})]$, we use the shorthand notation $\widehat{\Delta\boldsymbol{\theta}}(X_i) \equiv \widehat{\Delta\boldsymbol{\theta}}_k(X_i)$ for $i \in I_k$, suppress the dependence of $\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})$ on $\widehat{\mathcal{M}}$, and adapt Eqs. (22)-(23) to let

$$\zeta_i(\widehat{\mathcal{M}} q; \widehat{\boldsymbol{\theta}}) = (\widehat{\mathcal{M}} q)^\intercal \mathbf{L}_{0_i} + (\widehat{\mathcal{M}} q)^\intercal (\mathbf{L}_{1_i} - \mathbf{L}_{0_i}) \cdot \mathbb{1}\left\{ k\big(\widehat{\boldsymbol{\theta}}(X_i), \widehat{\mathcal{M}} q\big) > 0 \right\}, \tag{25}$$

$$k\big(\widehat{\boldsymbol{\theta}}(X_i), \widehat{\mathcal{M}} q\big) = q^\intercal \widehat{\mathcal{M}} \widehat{\Delta\boldsymbol{\theta}}(X_i) \tag{26}$$

Our main asymptotic result shows that $\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})$ converges to a Gaussian process uniformly in the direction $q \in \mathbb{S}^1$, where the score $\zeta_i(\mathcal{M} q; \boldsymbol{\vartheta})$ in Eq. (23) evaluated at $\Delta\boldsymbol{\vartheta} = \Delta\boldsymbol{\theta}$ is the influence function that governs the part of the asymptotic distribution of $\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})$ due to $\widehat{\Delta\boldsymbol{\theta}}$, and the the remaining part is attributed to estimating $\mathcal{M}$:

[4] An earlier version of this paper (Liu and Molinari, 2024) obtains $\sqrt{n}$-Gaussianty of the second-stage estimator without imposing Assumption 3, but under the classical Donsker condition that limits Θ's complexity.

THEOREM 4.1: *Let Assumptions 1-2-3 hold and* $\{(Y_i, G_i, X_i)\}_{i=1}^n$ *be a random sample from* $\mathbb{P}$. *Define a shrinking neighborhood around* $\Delta\boldsymbol{\theta}$ *as*

$$\Theta_n \equiv \big\{\Delta\boldsymbol{\vartheta} \in \Theta : \forall g \in \{0,1\}, \Delta\vartheta^g(X) = \tilde{\alpha}^g X_1 + \tilde{\beta}^g X_2 + \tilde{\eta}^g(X_{3:d_X}), \\ \max\{|\tilde{\alpha}^g - \alpha^g|, |\tilde{\beta}^g - \beta^g|, \|\tilde{\eta}^g - \eta^g\|_{L^2(\mathbb{P})}\} = o(n^{-1/4})\big\}.$$

Let $\widehat{\Delta\boldsymbol{\theta}}_k \in \Theta_n$ *with probability approaching 1,* $\forall k \in [K]$. *Then, for* $\widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}})$ *in Eq. (24),*

$$\sqrt{n}\Big(\widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q)\Big) = \mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] + o_p(1) \quad \text{in } \ell^\infty(\mathbb{S}^1),$$

where $\mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})]$ *is a Gaussian process in* $\ell^\infty(\mathbb{S}^1)$ *indexed by*

$$\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta}) \equiv \zeta_i(\mathcal{M}q;\boldsymbol{\theta}) + (\mathcal{M}_i^* q)^\intercal \mathcal{M}^{-1}\mathcal{S}_{\mathcal{E}}(q), \text{ for } \mathcal{M}_i^* \equiv \operatorname{diag}\left(\frac{\mathbb{1}\{G_i = r\}}{-\mu_r^2}, \frac{\mathbb{1}\{G_i = b\}}{-\mu_b^2}\right)$$

with $\zeta_i(\mathcal{M}q;\boldsymbol{\theta})$ *defined in Eq. (23) and the covariance function of* $\mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})]$ *equal to*

$$\Omega(q,\tilde{q}) = \mathbb{E}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})\zeta_i^*(\mathcal{M}\tilde{q};\boldsymbol{\theta})] - \mathbb{E}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})]\mathbb{E}[\zeta_i^*(\mathcal{M}\tilde{q};\boldsymbol{\theta})]. \tag{27}$$

If $Var(\mathbf{L}_d|X)$ *is positive definite, then* $Var(\mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})]) > 0$ *for each* $q \in \mathbb{S}^1$.

5. ESTIMATION AND INFERENCE FOR THE FRONTIER

5.1. *Estimation and Inference for the Feasible Set*

We propose an estimator of the set $\mathcal{E}$ based on $\widehat{h}_{\mathcal{E}}(q)$ and Eq. (9), given by

$$\widehat{\mathcal{E}} \equiv \bigcap_{q \in \mathbb{S}^1} \left\{z \in \mathbb{R}^2 : q^\intercal z \leq \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}})\right\}, \tag{28}$$

which is convex, almost surely compact, and non-empty with probability approaching one if $\mathcal{E}$ has a non-empty interior. As the Hausdorff distance between two non-empty convex and compact sets $A, B \in \mathbb{R}^d$, denoted $\mathbf{d}_H(A,B)$, equals the uniform distance between their support functions (Molchanov and Molinari, 2018, p. 101), when $\widehat{\mathcal{E}}$ is non-empty we have

$$\mathbf{d}_H(\widehat{\mathcal{E}}, \mathcal{E}) = \sup_{q \in \mathbb{S}^1} \left|\widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q;\boldsymbol{\theta})\right|,$$

By Theorem 4.1 and the continuous mapping theorem, $\mathbf{d}_H(\widehat{\mathcal{E}}, \mathcal{E}) \xrightarrow{p} 0$ and $\sqrt{n}\mathbf{d}_H(\widehat{\mathcal{E}}, \mathcal{E}) \xrightarrow{d} \sup_{q\in\mathbb{S}^1} |\mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})]|$. Hence, asymptotically valid tests of hypotheses about $\mathcal{E}$, and confidence sets covering it, can be obtained as in Beresteanu and Molinari (2008, Section 2).

5.2. *Estimation and Inference for the FA-Frontier*

As shown in LLMO (Theorem 1), when $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ is group-balanced, $\mathcal{F}$ is the curve connecting R and B and coincides with the Pareto frontier (panel (a) in Figure 2). However, when $\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ is g-skewed, $\mathcal{F}$ is the curve connecting F with the feasible point that minimizes the risk for group g (panels (b)-(e) in Figure 2). A further challenge is that, while it is simple to express F through $\mathcal{S}_{\mathcal{E}}$ when $\mathcal{E}$ is fully contained in one of the two half-spaces defined by the 45-degree line $\mathcal{H}_{45}$, as shown in Eqs. (15)-(16) (panels (b) and (d) in Figure 2), characterizing F is not as straightforward when g-skew occurs but $\mathcal{E} \cap \mathcal{H}_{45} \neq \emptyset$.

We therefore leverage the characterization of the FA-frontier $\mathcal{F}$ in Proposition 3.3, whereby $\mathcal{F} = \left\{ e \in \mathcal{E} : \ \left[\max_{q\in\tilde{\mathbb{S}}^1}(-h_{\mathcal{C}(e)}(q) - h_{\mathcal{E}}(-q))\right]_- = 0 \right\}$, and the definition of the set $\mathcal{C}(\cdot)$ in Eq. (20) to sidestep these difficulties. We propose to estimate $\mathcal{F}$ using

$$\widehat{\mathcal{F}} = \left\{ e \in \mathbb{B}_C : \ \left[\max_{q\in\mathbb{S}^1}(q^\intercal e - \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}))\right]_+ + \left[\max_{q\in\tilde{\mathbb{S}}^1}(-h_{\mathcal{C}(e)}(q) - \widehat{h}_{\mathcal{E}}(-q;\widehat{\boldsymbol{\theta}}))\right]_- \leq \frac{\kappa_n}{\sqrt{n}} \right\}, \quad (29)$$

where $[\,\cdot\,]_+ \equiv \max\{\cdot, 0\}$, $\kappa_n = o(\sqrt{n})$ is a sequence that diverges to infinity, and $\mathbb{B}_C \equiv \{e \in \mathbb{R}^2 : \|e\|_E \leq C\}$, with $\mathcal{E} \subset \mathbb{B}_C$ by Assumption 1 and $C < \infty$ a constant pinned down by c_1, c_2 defined in Assumption 1. The first maximization problem in Eq. (29) is the sample analog to the requirement that $e \in \mathcal{E}$; the second is the sample analog to the requirement that $\mathcal{E}$ and $\mathcal{C}(e)$ can be properly separated. To implement this estimator, we derive a closed-form expression for the support function of $\mathcal{C}(e)$ in direction $q = [q_1,\ q_2]^\intercal$. Only two points are "active" for evaluating $h_{\mathcal{C}(e)}(q)$: $[\min\{e_r, 2e_b - e_r\},\ e_b]^\intercal$ and $[e_r,\ \min\{e_b, 2e_r - e_b\}]^\intercal$, which correspond to the points e and $[e_r,\ 2e_r - e_b]^\intercal$ (respectively, $[2e_b - e_r,\ e_b]^\intercal$ and e) when e is above (respectively, below) the 45-degree line as shown in Figure 3-panel (a) (respectively, panel (b)). By Proposition 3.3, it is without loss of generality to focus on

$q \in \tilde{\mathbb{S}}^1$. Hence, the support function of $\mathcal{C}(e)$ at any $q \in \tilde{\mathbb{S}}^1$ equals

$$h_{\mathcal{C}(e)}(q) = \max\Big\{ q_1 \min\{e_r, 2e_b - e_r\} + q_2 e_b,\ \ q_1 e_r + q_2 \min\{e_b, 2e_r - e_b\} \Big\}. \tag{30}$$

We next propose a test statistic for the null hypothesis $\mathsf{H}_0 :\ e \in \mathcal{F}$, against the alternative $\mathsf{H}_A :\ e \notin \mathcal{F}$. Our test statistic is given by

$$T_n^{\mathcal{F}}(e) \equiv \sqrt{n} \left(\left[\max_{q \in \mathbb{S}^1} (q^\intercal e - \widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})) \right]_+ + \left[\max_{q \in \tilde{\mathbb{S}}^1} (-h_{\mathcal{C}(e)}(q) - \widehat{h}_{\mathcal{E}}(-q; \widehat{\boldsymbol{\theta}})) \right]_- \right). \tag{31}$$

As shown in the proof of Proposition 5.1, if Eq. (14) is satisfied and consequently $\mathcal{E}$ has no kinks, the large sample distribution of $T_n^{\mathcal{F}}(e)$, denoted $\psi^{\mathcal{F}}(e)$, simplifies to

$$\psi^{\mathcal{F}}(e) = \left| \mathbb{G} \left[\zeta_i^*(\mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta}) \right] \right|, \tag{32}$$

where $q^*_{\mathbb{S}^1}(e) \equiv \arg\max_{q \in \mathbb{S}^1} q^\intercal e - h_{\mathcal{E}}(q)$. The quantiles of this distribution can be estimated by standard methods. We build a confidence set for the elements of $\mathcal{F}$ by test inversion:

$$\mathcal{CS}_n(\mathcal{F}) = \left\{ e \in \mathbb{B}_C : T_n^{\mathcal{F}}(e) \leq c^{\mathcal{F}}_{1-\alpha}(e) \right\}, \tag{33}$$

where for any $\beta \in (0,1)$, $c^{\mathcal{F}}_\beta$ is the β-quantile of $\psi^{\mathcal{F}}$. When Eq. (14) is not assumed and $\mathcal{E}$ has kinks, the expression for $\psi^{\mathcal{F}}(e)$ is more complex and given in Eq. (72). In this case, restrictions that would guarantee uniform continuity and strict increasing properties for $\psi^{\mathcal{F}}(e)$ are harder to verify, and we follow Andrews and Shi (2013, p.625) to replace the confidence set in Eq. (33) with $\mathcal{CS}_n(\mathcal{F}) = \left\{ e \in \mathbb{B}_C : T_n^{\mathcal{F}}(e) \leq c^{\mathcal{F}}_{1-\alpha+\varsigma}(e) + \varsigma \right\}$, for $\varsigma > 0$ an arbitrarily small constant. In this case, the critical value can be approximated through bootstrap methods, as in Procedure 2 in Section 6.2.

We next establish consistency of our estimator $\widehat{\mathcal{F}}$ and validity of the confidence set, building respectively on Chernozhukov et al. (2007) and Fang and Santos (2019).

PROPOSITION 5.1: *Let the assumptions of Theorem 4.1 hold and let $\kappa_n \to \infty$ with $\kappa_n = o(\sqrt{n})$. Then, as $n \to \infty$,*

$$\mathbf{d}_H(\widehat{\mathcal{F}}, \mathcal{F}) \xrightarrow{p} 0 \tag{34}$$

$$\liminf_{n\to\infty} \mathbb{P}\left(e \in \mathcal{CS}_n(\mathcal{F})\right) \geq 1-\alpha \quad \textit{for all } e \in \mathcal{F}. \tag{35}$$

5.3. *Estimation and Inference for the Pareto Frontier*

The Pareto Frontier ($\mathcal{PF}$) is the lower boundary of the feasible set $\mathcal{E}$ connecting points R and B (see Figure 1). Denoting $\mathbb{Q} \equiv \left\{q \in \mathbb{S}^1 : q = [\cos\gamma \;\; \sin\gamma]^\intercal,\ \gamma \in [\pi, 3/2\pi]\right\}$, we can express $\mathcal{PF}$ through the support set $\mathcal{S}_\mathcal{E}(\cdot)$, or equivalently through the support function:

$$\mathcal{PF} \equiv \{\mathcal{S}_\mathcal{E}(q) : q \in \mathbb{Q}\} \tag{36a}$$

$$= \left\{ e \in \mathbb{R}^2 : \left[\max_{q\in\mathbb{S}^1}(q^\intercal e - h_\mathcal{E}(q))\right]_+ = 0,\ \left[\max_{q\in\mathbb{Q}}(q^\intercal e - h_\mathcal{E}(q))\right]_- = 0 \right\}, \tag{36b}$$

where the first condition in Eq. (36b) enforces that $e \in \mathcal{E}$ and the second that it belongs to the supporting hyperplane of $\mathcal{E}$ in a direction $q \in \mathbb{Q}$. Knowing the directions that determine the support points comprising $\mathcal{PF}$ simplifies the expression for it in Eq. (36b) relative to Eq. (21). It also allows for an estimator, put forward in Proposition 5.2 below, based on the support set characterization in Eq. (36a) that is free from the tuning parameter κ_n, which instead is needed for the estimator of $\mathcal{F}$ in Eq. (29). Let $\boldsymbol{\zeta}_{\mathcal{S},i}(\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) \equiv \widehat{\mathcal{M}}\mathbf{L}_{0_i} + \widehat{\mathcal{M}}(\mathbf{L}_{1_i} - \mathbf{L}_{0_i}) \cdot \mathbb{1}\left\{k\big(\widehat{\boldsymbol{\theta}}(X_i), \widehat{\mathcal{M}}q\big) > 0\right\}$, with $k\big(\widehat{\boldsymbol{\theta}}(X_i), \widehat{\mathcal{M}}q\big) = q^\intercal \widehat{\mathcal{M}}\widehat{\Delta\boldsymbol{\theta}}(X_i)$ as in Eq. (26) and[5]

$$\widehat{\mathcal{S}}_\mathcal{E}(q;\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\zeta}_{\mathcal{S},i}(\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}). \tag{37}$$

Proposition 5.2 delivers an estimator for $\mathcal{PF}$ based on $\widehat{\mathcal{S}}_\mathcal{E}(q;\widehat{\boldsymbol{\theta}})$ in Eq. (37) and establishes its Hausdorff-consistency.

PROPOSITION 5.2: *Let the assumptions of Theorem 4.1 hold. Let* $\widehat{\mathcal{PF}} \equiv \left\{\widehat{\mathcal{S}}_\mathcal{E}(q;\widehat{\boldsymbol{\theta}}) : q \in \mathbb{Q}\right\}$. *Then, as* $n \to \infty$,

$$\max_{q\in\mathbb{Q}} \|\widehat{\mathcal{S}}_\mathcal{E}(q;\widehat{\boldsymbol{\theta}}) - \mathcal{S}_\mathcal{E}(q)\|_E \xrightarrow{p} 0, \tag{38}$$

$$\mathbf{d}_H(\widehat{\mathcal{PF}}, \mathcal{PF}) \xrightarrow{p} 0. \tag{39}$$

[5] As in Eq. (24), this expression is a shorthand for $\widehat{\mathcal{S}}_\mathcal{E}(q;\widehat{\boldsymbol{\theta}}) \equiv \frac{1}{K}\sum_{k\in[K]}\left(\frac{1}{n/K}\sum_{i\in I_k}\boldsymbol{\zeta}_{\mathcal{S},i}(\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}_k)\right)$.

We next provide a method to test, for a given $e \in \mathbb{R}^2$,

$$\mathsf{H}_0 : \ e \in \mathcal{PF} \quad \text{against} \quad \mathsf{H}_A : \ e \notin \mathcal{PF}, \tag{40}$$

using the characterization of $\mathcal{PF}$ in Eq. (36b). We first propose a test statistic and derive its asymptotic distribution in Proposition 5.3 below, building on results in Kaido (2016).

REMARK 5.1: We use different characterizations of $\mathcal{PF}$ for estimation and for inference due to difficulties with DML estimation of $\mathcal{S}_{\mathcal{E}}(q)$ that we explain here. Using Eq. (12) we can write the first (second) coordinate of $\mathcal{S}_{\mathcal{E}}(q)$ as $\mathbb{E}\big[(\mathcal{M}v)^{\intercal}\mathbf{L}_0 + (\mathcal{M}v)^{\intercal}(\mathbf{L}_1 - \mathbf{L}_0)\mathbb{1}\{k(\boldsymbol{\theta}(X), \mathcal{M}q) > 0\}\big]$ for $v = [1 \ \ 0]^{\intercal}$ ($v = [0 \ \ 1]^{\intercal}$). As $k(\boldsymbol{\theta}(X), \mathcal{M}v) = \mathbb{E}[(\mathcal{M}v)^{\intercal}(\mathbf{L}_1 - \mathbf{L}_0)|X]$, the proof of Theorem 4.1 shows that, if $v = q$, the first-stage estimation error in the sign of $k(\boldsymbol{\theta}(X), \mathcal{M}q)$ is controlled by the size of the error itself (because in case of sign disagreement, $|k(\boldsymbol{\theta}(X), \mathcal{M}q)| \leq |k(\widehat{\boldsymbol{\theta}}(X), \mathcal{M}q) - k(\boldsymbol{\theta}(X), \mathcal{M}q)|$), with $k(\widehat{\boldsymbol{\theta}}(X), \mathcal{M}q)$ as in Eq. (26). However, when $v \neq q$, which necessarily occurs for at least one coordinate of $\mathcal{S}_{\mathcal{E}}(q)$, sign errors are not controlled for. Consequently, we switch to the moment inequality characterization in Eq. (36b) that involves $h_{\mathcal{E}}(q)$ only. If one uses a parametric estimator for $\Delta\boldsymbol{\theta}(X)$, the asymptotic distribution of $\widehat{\mathcal{S}}_{\mathcal{E}}(\cdot; \widehat{\boldsymbol{\theta}})$ can be obtained and inference is simplified, as a special case of the treatment in Liu and Molinari (2024), who work with a sieve nonparametric estimator of $\Delta\boldsymbol{\theta}(X)$ under Donsker conditions.

PROPOSITION 5.3: *Let* $Q^*_{\mathbb{T}}(e) \equiv \arg\max_{q \in \mathbb{T}} q^{\intercal}e - h_{\mathcal{E}}(q)$, $\mathbb{T} \in \{\mathbb{S}^1, \mathbb{Q}\}$. *Then, under the null in Eq.* (40) *and the assumptions of Theorem 4.1,*

$$T_n^{\mathcal{PF}}(e) \equiv \sqrt{n}\left(\left[\max_{q \in \mathbb{S}^1} q^{\intercal}e - \widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})\right]_+ + \left[\max_{q \in \mathbb{Q}} q^{\intercal}e - \widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})\right]_-\right) \tag{41}$$

$$\xrightarrow{d} \left[\sup_{q \in Q^*_{\mathbb{S}^1}(e)} \mathbb{G}[-\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta})]\right]_+ + \left[\sup_{q \in Q^*_{\mathbb{Q}}(e)} \mathbb{G}[-\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta})]\right]_- . \tag{42}$$

If $Var(\mathbf{L}_d|X)$ *is positive definite for each* $d \in \{0, 1\}$, $X - a.s.$, *the limit law in Eq.* (42) *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{R}_{++}$.

If Eq. (14) is satisfied and the set $\mathcal{E}$ has no kinks (see Remark 3.3), $Q^*_{\mathbb{S}^1}(e)$ and $Q^*_{\mathbb{Q}}(e)$ are singletons that can be consistently estimated through standard methods, and the limit distribution in Eq. (42) coincides with that in Eq. (32). If instead kinks are not ruled out for $q \in \mathbb{Q}$, $Q^*_{\mathbb{S}^1}(e)$ and $Q^*_{\mathbb{Q}}(e)$ may not be singletons. In this case we can consistently estimate these sets (Kaido, 2016, Lemma D.5) as

$$\widehat{Q}^*_{\mathbb{T}}(e) \equiv \left\{ q \in \mathbb{T} : q^\intercal e - \widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}}) \geq \sup_{\tilde{q} \in \mathbb{T}} \tilde{q}^\intercal e - \widehat{h}_{\mathcal{E}}(\tilde{q}; \widehat{\boldsymbol{\theta}}) - \kappa_n / \sqrt{n} \right\}, \tag{43}$$

where $\mathbb{T} \in \{\mathbb{S}^1, \mathbb{Q}\}$ and $\kappa_n = o(\sqrt{n})$ is a sequence that diverges to infinity. We use Procedure 2-Step 1 to obtain a valid bootstrap-based approximation to the Gaussian process in Theorem 4.1. Denote $\mathbb{P}^*$ this bootstrap distribution conditional on $\{(Y_i, G_i, X_i)\}_{i=1}^n$. Let

$$\hat{c}^{\mathcal{PF}}_{1-\alpha}(e) \equiv \inf \left\{ c : \mathbb{P}^* \left(\left[\sup_{q \in Q^*_{\mathbb{S}^1}(e)} \mathbb{G}[-\zeta^*_i(\mathcal{M}q; \boldsymbol{\theta})] \right]_+ + \left[\sup_{q \in Q^*_{\mathbb{Q}}(e)} \mathbb{G}[-\zeta^*_i(\mathcal{M}q; \boldsymbol{\theta})] \right]_- > c \right) = \alpha \right\} \tag{44}$$

When $Var(\mathbf{L}_d | X)$ is positive definite, it follows that if $\alpha \in (0, 0.5)$,

$$\limsup_{n \to \infty} \mathbb{P}\left(T^{\mathcal{PF}}_n(e) > \hat{c}^{\mathcal{PF}}_{1-\alpha}(e) \right) \begin{cases} \leq \alpha & \text{if } e \in \mathcal{PF} \\ = 1 & \text{if } e \notin \mathcal{PF} \end{cases}$$

(Kaido, 2016, Corollary 3.2). A confidence set that covers each $e \in \mathcal{PF}$ with asymptotic probability at least equal to $(1 - \alpha)$ can be obtained by test inversion.

REMARK 5.2: The result in Proposition 5.3 can be adapted to testing whether $e^* \in \mathcal{PF}$ when e^* needs to be estimated, by adjusting the covariance function of the limit Gaussian process in Theorem 4.1 and using the test statistic in Eq. (41). If the limit law in Eq. (42) is not guaranteed to be absolutely continuous on $\mathbb{R}_{++}$, one can use infinitesimal adjustments to the critical value to maintain asymptotic validity, as in Kaido (2016).

5.4. *Algorithms Yielding a Risk Allocation on the Frontier*

An algorithm designer or a regulator may wonder if one can characterize the algorithm yielding a specific point on the FA-frontier $\mathcal{F}$ or the Pareto frontier $\mathcal{PF}$. It turns out that,

using our support function approach, the answer to this question is affirmative, and particularly simple for points in $\mathcal{PF}$.

Suppose we have two data sets: one for training the algorithm and the other for evaluating it. We denote the training sample as $\{(\tilde{Y}_i, \tilde{G}_i, \tilde{X}_i)\}_{i=1}^{n_1}$ and the evaluation sample as $\{(Y_j, G_j, X_j)\}_{j=1}^{n_2}$, with $\frac{n_1}{n_2} \to c$ for some positive constant c and both samples drawn from the same distribution $\mathbb{P}$. Use the training data to estimate $\Delta\boldsymbol{\theta}$ via machine learning and $\mathcal{M}$ by sample averages as described in Section 4 and denote the resulting estimators $\widehat{\Delta\boldsymbol{\theta}}_{n_1}$ and $\widehat{\mathcal{M}}_{n_1}$, where the subscript n_1 indicates that it is based on the training sample. Let

$$\widehat{a}_{n_1}(X_j; q) = \mathbb{1}\big\{k\big(\widehat{\boldsymbol{\theta}}_{n_1}(X_j), \widehat{\mathcal{M}}_{n_1} q\big) > 0\big\}, \tag{45}$$

with $k\big(\widehat{\boldsymbol{\theta}}_{n_1}(X_j), \widehat{\mathcal{M}}_{n_1} q\big)$ as in Eq. (26). Note that $\widehat{a}_{n_1}(X_j; q)$ only takes as input the covariates, X_j, and does not depend on group identity G_j. Consequently, individuals with the same covariates are assigned the same treatment irrespective of their group. Nonetheless, information on group identity contained in the training data is used in the prediction model (to separately estimate $\Delta\theta^g$, $g \in \{r, b\}$), thereby offering a compromise between a utilitarian perspective as in Manski et al. (2023) which advocates for group-aware decisions, and proponents of group-blind decision making.

Then, for any $q \in \mathbb{S}^1$, algorithm $\widehat{a}_{n_1}(\,\cdot\,; q)$ leads to a loss for each observation i in the evaluation sample equal to $\ell(0, Y_j) + (\ell(1, Y_j) - \ell(0, Y_j)) \cdot \mathbb{1}\big\{k\big(\widehat{\boldsymbol{\theta}}_{n_1}(X_j), q\big) > 0\big\}$. Therefore, the average losses for the r group and for the b group in the evaluation sample equal

$$\begin{bmatrix} \frac{\sum_{j:G_j=r} \ell(0,Y_j) + (\ell(1,Y_j) - \ell(0,Y_j)) \cdot \mathbb{1}\big\{k\big(\widehat{\boldsymbol{\theta}}_{n_1}(X_j), q\big) > 0\big\}}{\sum_{j=1}^{n_2} \mathbb{1}\{G_j = r\}} \\ \frac{\sum_{j:G_j=b} \ell(0,Y_j) + (\ell(1,Y_j) - \ell(0,Y_j)) \cdot \mathbb{1}\big\{k\big(\widehat{\boldsymbol{\theta}}_{n_1}(X_j), q\big) > 0\big\}}{\sum_{j=1}^{n_2} \mathbb{1}\{G_j = b\}} \end{bmatrix} = \widehat{\mathcal{S}}_{\mathcal{E}, n_2}(q; \widehat{\boldsymbol{\theta}}_{n_1}) \tag{46}$$

By the same argument as in the proof of Proposition 5.2, it follows that

$$\max_{q \in \mathbb{Q}} \left\| \widehat{\mathcal{S}}_{\mathcal{E}, n_2}(q; \widehat{\boldsymbol{\theta}}_{n_1}) - \mathcal{S}_{\mathcal{E}}(q) \right\| \xrightarrow{p} 0 \quad \text{as } n_1, n_2 \to \infty.$$

Hence, the algorithm in Eq. (45) with $q \in \mathbb{Q}$ gives consistent estimators of points on $\mathcal{PF}$.

Algorithms that return consistent estimators of points on $\mathcal{F}$ (other than those on $\mathcal{PF}$) are harder to obtain, as one needs to determine which direction q corresponds to a point $e \in \mathcal{F}$

and plug that direction in the algorithm in Eq. (45). Suppose Eq. (14) is satisfied and $\mathcal{E}$ has no kinks. Use the same DML construction in Section 4 applied to the training sample to obtain an estimator of $h_{\mathcal{E}}(q)$, denoted $\widehat{h}_{\mathcal{E},n_1}(q;\widehat{\boldsymbol{\theta}}_{n_1})$, as in Eq. (24), and an estimator of $\mathcal{F}$, denoted $\widehat{\mathcal{F}}_{n_1}$, as in Eq. (29). Select $\widehat{e}_{n_1} \in \widehat{\mathcal{F}}_{n_1}$ through a projection method detailed in the proof of Proposition 5.4 so that $\|\widehat{e}_{n_1} - e\| = o_p(1)$ for some $e \in \mathcal{F}$ and let

$$\widehat{q}^*_{n_1}(\widehat{e}_{n_1}) = \arg\max_{q\in\mathbb{S}^1} q^\intercal \widehat{e}_{n_1} - \widehat{h}_{\mathcal{E},n_1}(q;\widehat{\boldsymbol{\theta}}_{n_1}). \tag{47}$$

Let $\widehat{\mathcal{S}}_{\mathcal{E},n_2}(\widehat{q}^*_{n_1}(\widehat{e}_{n_1});\widehat{\boldsymbol{\theta}}_{n_1})$ be as in Eq. (46) but with $\widehat{q}^*_{n_1}(\widehat{e}_{n_1})$ replacing q. The next proposition establishes that $\widehat{\mathcal{S}}_{\mathcal{E},n_2}(\widehat{q}^*_{n_1}(\widehat{e}_{n_1});\widehat{\boldsymbol{\theta}}_{n_1})$ is a consistent estimator of $\mathcal{S}_{\mathcal{E}}(q^*_{\mathbb{S}^1}(e)) \in \mathcal{F}$.

PROPOSITION 5.4: *Let the assumptions of Theorem 4.1 hold. Then, as* $n_1, n_2 \to \infty$,

$$\left\| \widehat{\mathcal{S}}_{\mathcal{E},n_2}(\widehat{q}^*_{n_1}(\widehat{e}_{n_1});\widehat{\boldsymbol{\theta}}_{n_1}) - \mathcal{S}_{\mathcal{E}}(q^*_{\mathbb{S}^1}(e)) \right\| \xrightarrow{p} 0.$$

Regardless of whether one aims at obtaining points in $\mathcal{PF}$ or the entire $\mathcal{F}$, the direction q can be interpreted as the vector of weights that the agent choosing the algorithm puts on each group's risk. In other words, one may think of the agent as evaluating group risks according to the welfare loss function $U(e;q) \equiv q_1 e_r(a) + q_2 e_b(a)$. For example, the more the agent cares about group r, the closer q is to $\mathfrak{u}_1$. We note that $\widehat{a}_{n_1}(X;q)$ is an empirical success rule, and leave its statistical decision theory analysis to future research.

6. HYPOTHESIS TESTING

In this Section we propose hypothesis tests to answer the following policy questions:

(1) Should the policymaker consider banning group identity as an input to the algorithm?

(2) Is there a less discriminatory alternative (LDA) to an existing algorithm?

We show how to express the first policy question in terms of restrictions on $h_{\mathcal{E}}(q)$, and we leverage Proposition 3.3 to do the same for the second policy question. We then establish asymptotic validity of the corresponding testing procedures.

6.1. *How to Test Whether Group Identity Should be Banned*

LLMO (Proposition 6) show that using X only instead of (X, G) as algorithmic input uniformly worsens the frontier if R and B, obtained when only X is used as input, are strictly separated by the 45-degree line.[6] We therefore aim at testing the null hypothesis that R and B lie weakly on the same side of the 45 degree line, i.e., that the difference in the two coordinates of R has the same sign as that of B, against the alternative that they are strictly separated by the 45-degree line:

$$\begin{aligned} \mathsf{H}_0 &: \Big((u_1 - u_2)^\intercal R\Big)\Big((u_1 - u_2)^\intercal B\Big) \geq 0, \\ \mathsf{H}_A &: \Big((u_1 - u_2)^\intercal R\Big)\Big((u_1 - u_2)^\intercal B\Big) < 0. \end{aligned} \tag{48}$$

If the null in Eq. (48) is rejected, the policymaker should not ban group identity G as algorithm's input. A Type-I error amounts to the case where one concludes that there is strict group-balance, while instead weak group-skew holds. As a consequence, one does not ban G as input to the algorithm, thinking that banning G is uniformly welfare-reducing, while instead depending on the preferences of the designer it might not be the case. Another interpretation of this test amounts to determining, based on whether H_0 is rejected or not, if one can justify implementing algorithms that lead to Pareto-dominated risks based on the designer's preference over fairness and accuracy. When H_0 holds true, this justification is possible, as the frontier includes an upward-sloping segment (e.g., Panels (b)-(c) of Figure 1). On the other hand, when H_A holds true, such justification is untenable, as the frontier coincides with the Pareto frontier (e.g., Panel (a) of Figure 1). Hence, a Type-I error can also be interpreted as a case where one concludes that only Pareto-optimal risks should be implemented, while fairness considerations may justify Pareto-dominated risks.

As discussed in Remark 5.1, carrying out inference if we use DML to directly estimate both coordinates of the points R and B is difficult. Yet, using Eqs. (11) and (13), we can

[6] "Uniformly worsening the frontier" means that, under the preference relations defined in Eq. (2), every point on the frontier $\mathcal{F}\big(\mathbb{P}, \mathcal{A}(\mathcal{X})\big)$ is dominated by a point on the frontier $\mathcal{F}\big(\mathbb{P}, \mathcal{A}(\mathcal{X} \times \{r, b\})\big)$.

represent these points through moment equalities and inequalities that involve $h_{\mathcal{E}}(q)$ only:

$$R: \begin{cases} h_{\mathcal{E}}(\mathfrak{u}_1) - \mathfrak{u}_1^{\intercal} R = 0, \\ h_{\mathcal{E}}(q) - q^{\intercal} R \geq 0, \ \forall q \in \mathbb{S}^1, \end{cases} \qquad B: \begin{cases} h_{\mathcal{E}}(\mathfrak{u}_2) - \mathfrak{u}_2^{\intercal} B = 0, \\ h_{\mathcal{E}}(q) - q^{\intercal} B \geq 0, \ \forall q \in \mathbb{S}^1, \end{cases} \tag{49}$$

where in Eq. (49), the equality constraint for R restricts it to have horizontal coordinate equal to that of $\mathcal{S}_{\mathcal{E}}(\mathfrak{u}_1)$, as per Eq. (11), and the continuum of inequality constraints indexed by $q \in \mathbb{S}^1$ restricts R to be an element of $\mathcal{E}$. The moment constraints that define B are interpreted similarly. Because the support set $\mathcal{S}_{\mathcal{E}}(\cdot)$ in any direction is a singleton by Proposition 3.2, the moments in Eq. (49) yield points R and B that coincide with Eq. (13).

We test the null in Eq. (48) at a given significance level $\alpha \in (0,1)$ based on our procedure to test, for given $e \in \mathbb{R}^2$, whether $e \in \mathcal{PF}$:

PROCEDURE 1—Testing Weak Group-Skew:

1. *Build a* $(1-\alpha)$*-level confidence set for* (R,B) *by*

$$\mathcal{CS}_n(R,B) \equiv \left\{(\tilde{R},\tilde{B}) \in \mathbb{B}_C \times \mathbb{B}_C : T_n(\tilde{R},\tilde{B}) \leq \hat{c}_{1-\alpha}(\tilde{R},\tilde{B})\right\}, \tag{50}$$

where for the support function estimator $\widehat{h}_{\mathcal{E}}(\cdot;\widehat{\boldsymbol{\theta}})$ *in Theorem 4.1,* $T_n(\tilde{R},\tilde{B})$ *adapts the test statistic in Eq.* (41)*:*

$$T_n(\tilde{R},\tilde{B}) \equiv \sum_{\substack{(e,\mathfrak{u}_j) \\ \in\{(\tilde{R},\mathfrak{u}_1),(\tilde{B},\mathfrak{u}_2)\}}} \sqrt{n}\left(\left[\max_{q\in\mathbb{S}^1} q^{\intercal}e - \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}})\right]_+ + \left[\mathfrak{u}_j^{\intercal}e - \widehat{h}_{\mathcal{E}}(\mathfrak{u}_j;\widehat{\boldsymbol{\theta}})\right]_-\right).$$

The critical value $\hat{c}_{1-\alpha}(\tilde{R},\tilde{B})$ *is obtained similarly to* $\hat{c}^{\mathcal{PF}}_{1-\alpha}(e)$ *in Eq.* (44)*.*

2. *Reject* H_0 *in Eq.* (48) *if*

$$\varphi_n^{skew} \equiv \mathbb{1}\left\{\sup_{(\tilde{R},\tilde{B})\in\mathcal{CS}_n(R,B)} \left((\mathfrak{u}_1-\mathfrak{u}_2)^{\intercal}\tilde{R}\right)\left((\mathfrak{u}_1-\mathfrak{u}_2)^{\intercal}\tilde{B}\right) < 0\right\} = 1. \tag{51}$$

We note that the test in Eq. (51) may be conservative as it is based on projection.

PROPOSITION 6.1: *Let the assumptions in Theorem 4.1 hold. Then*

$$\limsup_{n\to\infty} \mathbb{E}\left[\varphi_n^{skew}\right] \leq \alpha. \tag{52}$$

6.2. *How to Test for the Existence of an LDA*

Given an algorithm $a^* \in \mathcal{A}(\mathcal{X})$ that induces the risk pair $e^* = (e_r^*, e_b^*) \in \mathcal{E}$, call another algorithm that yields a feasible risk pair $e = (e_r, e_b) \in \mathcal{E}$ an LDA if it is at least as accurate as a^* for both groups and at least as fair, with one of these inequalities strict. It follows from the characterization in Proposition 3.3 and Eq. (20) that no LDA to a^* exists if and only if $\mathcal{E}$ can be properly separated from

$$\mathcal{C}(e^*) \equiv \left\{ e \in \mathbb{R}^2 : e_r \leq e_r^*, e_b \leq e_b^*, |e_r - e_b| \leq |e_r^* - e_b^*| \right\}.$$

Recall that the closed form expression for the support function of $\mathcal{C}^* \equiv \mathcal{C}(e^*)$ in direction $q = [q_1,\ q_2]^\intercal \in \tilde{\mathbb{S}}^1$ is given in Eq. (30). We then test the null hypothesis

$$\mathsf{H}_0 : \max_{q \in \tilde{\mathbb{S}}^1} (-h_{\mathcal{C}^*}(q) - h_{\mathcal{E}}(-q)) = 0 \tag{53}$$

against the alternative that H_0 is false. Rejecting the null in Eq. (53) means that $e^* \notin \mathcal{F}$ and there exists an LDA. We propose estimating $h_{\mathcal{C}^*}(q)$ for $q \in \tilde{\mathbb{S}}^1$ by

$$\widehat{h}_{\mathcal{C}^*}(q) = \max\left\{ q_1 \min\{\widehat{e}_r^*, 2\widehat{e}_b^* - \widehat{e}_r^*\} + q_2 \widehat{e}_b^*,\ \ q_1 \widehat{e}_r^* + q_2 \min\{\widehat{e}_b^*, 2\widehat{e}_r^* - \widehat{e}_b^*\} \right\},$$

where for $g \in \{r, b\}$, e_g^* is estimated by sample means,

$$\widehat{e}_g^* = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i^g}{\widehat{\mu}_g}, \quad \text{where } Z_i^g \equiv \mathbb{1}\{G_i = g\} \big(a^*(X_i)\ell(1, Y_i) + (1 - a^*(X_i))\ell(0, Y_i) \big) \tag{54}$$

and $\widehat{\mu}_g \equiv \frac{1}{n} \sum_{i=1} \mathbb{1}\{G_i = g\}$. We propose the following test statistic:

$$T_n^{\mathrm{LDA}} \equiv \sqrt{n} \left(\left[\max_{q \in \mathbb{S}^1} (q^\intercal \widehat{e}^* - \widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})) \right]_+ + \left[\max_{q \in \tilde{\mathbb{S}}^1} \left(-\widehat{h}_{\mathcal{C}^*}(q) - \widehat{h}_{\mathcal{E}}(-q) \right) \right]_- \right). \tag{55}$$

T_n^{LDA} differs from $T_n^{\mathcal{F}}$ in Eq. (31) only in that $\widehat{e}^*$ is estimated in the former.

PROPOSITION 6.2: *Let the assumptions in Theorem 4.1 hold. Then, for any pre-specified significance level $\alpha \in (0, 1)$, the test below has asymptotically correct size control:*

$$\textit{Reject the null in Eq. (53) if} \quad T_n^{LDA} > c_{1-\alpha+\varsigma}^{LDA} + \varsigma,$$

where $\varsigma > 0$ is an arbitrarily small positive constant and for any $\beta \in (0,1)$ the critical value c_β^{LDA} is the β-quantile of ψ^{LDA}, for ψ^{LDA} a random variable defined in Eq. (84).

When Eq. (14) holds and $Var(\mathbf{L}_d|X)$ is positive definite for $d \in \{0,1\}$, X-a.s., one can take $\varsigma = 0$ in Proposition 6.2 and the expression for ψ^{LDA} simplifies to that in Eq. (85). When kinks might be present and the limit distribution is not guaranteed to be continuous and strictly increasing, we take $\varsigma > 0$ as in Andrews and Shi (2013). The derivation of ψ^{LDA} uses the fact that T_n^{LDA} is formed by compositions of the $\max$ and $\min$ functions in Eq. (30) and the $\max$ function in Eq. (53)—all of which are Hadamard directionally differentiable, as shown in Fang and Santos (2019) and Cárcamo et al. (2020); since compositions preserve directional differentiability (Shapiro, 1990, Proposition 3.6), an extension to the functional Delta method can be applied to Theorem 4.1. However, standard bootstraps are inconsistent due to the lack of full differentiability (see Section 3.2 of Fang and Santos, 2019). As such, we leverage results in Fang and Santos (2019) to approximate the distribution of ψ^{LDA} and its quantiles via a modified multiplier bootstrap procedure detailed below, similar to that in Semenova (2023), where we denote ϕ any generic Hadamard directionally differentiable function, $\widehat{he^*} = \widehat{he^*}(\widehat{\boldsymbol{\theta}}) \equiv [\widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}),\ \widehat{e}_r^*,\ \widehat{e}_b^*]^\intercal$ the vector of estimators, and $he^* \equiv [h_{\mathcal{E}}(q),\ e_r^*,\ e_b^*]^\intercal$ the vector of truths.

PROCEDURE 2—Bootstrap for the Quantiles of $\sqrt{n}\{\phi(\widehat{he^*}) - \phi(he^*)\}$:

1. *Draw $\{W_i\}_{i=1}^n$ i.i.d. from the exponential distribution with mean 1 independent of the sample $\{(Y_i, G_i, X_i)\}_{i=1}^n$ and construct the bootstrap analogue of $\widehat{he^*}$:*

$$\widetilde{he^*} = \widetilde{he^*}(\widehat{\boldsymbol{\theta}}) \equiv [\widetilde{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}),\ \widetilde{e}_r^*,\ \widetilde{e}_b^*]^\intercal, \tag{56}$$

   where $\widetilde{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) \equiv \frac{1}{n}\sum_{i=1}^n \frac{W_i}{\overline{W}} \zeta_i(\widetilde{\mathcal{M}}q;\widehat{\boldsymbol{\theta}})$ for $\overline{W} \equiv \frac{1}{n}\sum_{i=1}^n W_i$, $\widetilde{\mathcal{M}} \equiv \text{diag}(1/\widetilde{\mu}_r, 1/\widetilde{\mu}_b)$, $\widetilde{\mu}_g \equiv \frac{1}{n}\sum_{i=1}^n \frac{W_i}{\overline{W}} \mathbb{1}\{G_i = g\}$, and $\widetilde{e}_g^* \equiv \frac{1}{n}\sum_{i=1}^n \frac{W_i}{\overline{W}} \frac{Z_i^g}{\widetilde{\mu}_g}$.

2. *Numerically approximate $\phi'_{he^*}(\cdot)$, the directional derivative of $\phi(\cdot)$ at he^*, by*

$$\widehat{\phi}'_{he^*}(\ddot{he}) = \frac{1}{s_n}\left(\phi\left(\widehat{he^*} + s_n(\ddot{he})\right) - \phi\left(\widehat{he^*}\right)\right),$$

   where $\ddot{he} \in \ell^\infty(\mathbb{S}^1) \times \mathbb{R}^2$ is a candidate direction at which we evaluate $\phi'_{he^*}(\cdot)$ and s_n is a vanishing sequence of step sizes such that $\sqrt{n}s_n \to \infty$.

3. *Obtain* $\widehat{\phi'}_{he^*}\big(\sqrt{n}\{\widetilde{he^*} - \widehat{he^*}\}\big)$ *and*

$$\widehat{c}_\beta \equiv \inf\left\{c : \mathbb{P}\left(\widehat{\phi'}_{he^*}\big(\sqrt{n}\{\widetilde{he^*} - \widehat{he^*}\}\big) \leq c \;\middle|\; \{(Y_i, G_i, X_i)\}_{i=1}^n\right) \geq \beta\right\},$$

as estimators, respectively, for the limit distribution of $\sqrt{n}\{\phi\big(\widehat{he^*}\big) - \phi\big(he^*\big)\}$ *and its* β*-quantile, denoted as* c_β*.*

The consistency of the bootstrap outlined in Procedure 2 is stated in the following result.

PROPOSITION 6.3: *Under the assumptions of Theorem 4.1,*

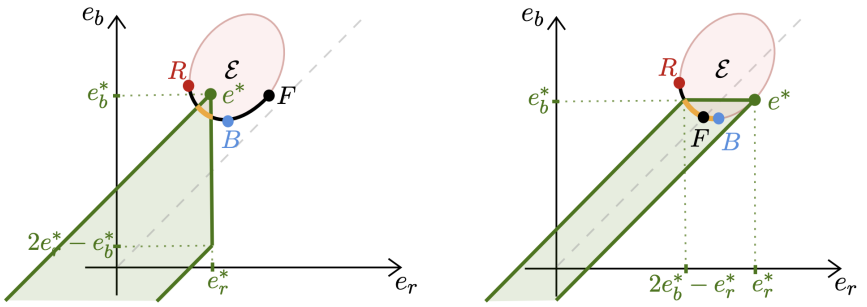
$$\sup_{f \in \mathcal{BL}_1} \left|\mathbb{E}\big[f\big(\widehat{\phi'}_{he^*}\big(\sqrt{n}\{\widetilde{he^*} - \widehat{he^*}\}\big)\big) \,\big|\, \{(Y_i, G_i, X_i)\}_{i=1}^n\big] - \mathbb{E}\big[f\big(\phi'_{he^*}(\mathbb{G}_{he^*})\big)\big]\right| = o_p(1),$$

where $\mathcal{BL}_1$ *is the set of* 1*-Lipschitz functions* $f : \mathbb{R} \to \mathbb{R}$ *such that* $|f|_\infty \leq 1$ *and* $\mathbb{G}_{he^*}$ *is the Gaussian limit process of* $\sqrt{n}(\widehat{he^*} - he^*)$ *given in Eq. (77) in Appendix A. If the cdf of* $\phi'_{he^*}(\mathbb{G}_{he^*})$ *is continuous and increasing at its* β*-quantile, denoted* c_β*, then* $\widehat{c}_\beta = c_\beta + o_p(1)$*.*

The proof of Proposition 6.2 shows that $\psi^{\mathrm{LDA}} = \phi'_{he^*}(\mathbb{G}_{he^*})$ for a particular $\phi'_{he^*}(\cdot)$ that is the composition of the directional derivatives of the $\min$, $\max$, and $\inf$ functions that constitute T_n^{LDA}. The expression of $\phi'_{he^*}(\cdot)$, given in Eq. (84), is complex and hence we recommend the numerical approximation approach in Step 2 of Procedure 2. Under Proposition 6.3, we can estimate c_β^{LDA}, the β-quantile of ψ^{LDA}, by going through the steps in Procedure 2, where we replace $\phi(\cdot)$ by the composition of the above mentioned directionally differentiable functions accordingly. The same bootstrap procedure can be used to consistently estimate the critical values put forward to carry out inference in Section 7. The use of the infinitesimal constant ς in Proposition 6.2 and Procedure 3 accounts for the possibility that the cdf of $\phi'_{he^*}(\mathbb{G}_{he^*})$ may not be continuous and increasing at $c_{1-\alpha}$.

6.2.1. *Algorithms Yielding an LDA*

An algorithm designer or a regulator may wonder if one can characterize the algorithms yielding LDAs to a given algorithm e^*. It turns out that it is possible to do so, by combining the algorithm that we put forward in Eq. (45) with a careful use of our characterization of $\mathcal{F}$



(a) e^* is above the 45° line and in the interior (b) e^* is below the 45° line and on the boundary

FIGURE 4.—The set $\mathcal{F}^*$, marked in orange, is the portion of the FA-frontier yielding risk pairs preferred to e^*.

in Proposition 3.3. We illustrate the idea for the case that Eq. (14) holds and $\mathcal{E}$ has no kinks. For a given risk pair e^* induced by an algorithm $a^* \in \mathcal{A}$ such that $e^* \notin \mathcal{F}$, by definition the set of risk pairs $e \in \mathcal{F}$ such that $e >_{FA} e^*$ are:

$$\mathcal{F}^* \equiv \mathcal{F} \cap \mathcal{C}(e^*) = \left\{ e \in \mathcal{E} : \left[\max_{q \in \tilde{\mathbb{S}}^1} (-h_{\mathcal{C}(e)}(q) - h_{\mathcal{E}}(-q)) \right]_- = 0, \right.$$
$$\left. \left[\max_{q \in \{u_1 - u_2, u_2 - u_1, -u_1, -u_2\}} \left(q^\intercal e - h_{\mathcal{C}(e^*)}(q) \right) \right]_+ = 0 \right\}. \quad (57)$$

The set $\mathcal{F}^*$ is depicted in Figure 4 as the orange portion of $\mathcal{F}$ that intersects with $\mathcal{C}(e^*)$. Using the same notation as in Section 5.4, we can use a training sample $\{(\tilde{Y}_i, \tilde{G}_i, \tilde{X}_i)\}_{i=1}^{n_1}$ and a construction that mimics the procedure to build the estimator of $\mathcal{F}$ in Eq. (29) to obtain a consistent estimator $\widehat{\mathcal{F}}^*_{n_1}$ of $\mathcal{F}^*$ (the consistency of this estimator can be established through the same steps as in the proof of Proposition 5.1). Select $\widehat{e}_{n_1} \in \widehat{\mathcal{F}}_{n_1}$ through a projection method detailed in the proof of Proposition 5.4 and denote by $e \in \mathcal{F}^*$ the point to which it converges. Let $\widehat{q}^*_{n_1}(\widehat{e}_{n_1})$ be the consistent estimator of $q^*_{\mathbb{S}^1}(e)$ defined in Eq. (47). Let $\widehat{\mathcal{S}}_{\mathcal{E},n_2}(\widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1})$ be defined as in Eq. (46) but with $\widehat{q}^*_{n_1}(\widehat{e}_{n_1})$ replacing q. Then $\widehat{\mathcal{S}}_{\mathcal{E},n_2}(\widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1})$ is a consistent estimator of $\mathcal{S}_{\mathcal{E}}(q^*_{\mathbb{S}^1}(e)) \in \mathcal{F}^*$, by the same argument used to establish Proposition 5.4.

7. DISTANCE TO THE FAIREST POINT

In this section, we propose a method to build a confidence interval for the distance between the risk e^* induced by a given algorithm and the fairest point F on the frontier, denoted $\rho(e^*, F)$, where ρ is a Hadamard directionally differentiable distance function (e.g., the Euclidean distance, Manhattan distance, Chebyshev distance, etc.). This can inform a decision maker of the relative merits in promoting equity and achieving business efficiency of different algorithms by comparing the confidence intervals on their distance to F.

Recall $\mathcal{H}_{45} \equiv \{e \in \mathbb{R}^2 : e_r = e_b\}$ denotes the 45-degree line; let $\mathcal{H}_{45}^+ \equiv \{e \in \mathbb{R}^2 : e_r < e_b\}$ and $\mathcal{H}_{45}^- \equiv \{e \in \mathbb{R}^2 : e_r > e_b\}$ denote, respectively, the open halfspace above and below the 45-degree line. As shown in Section 3.3, the coordinates of F depend on whether $\mathcal{E}$ intersects with $\mathcal{H}_{45}$. When $\tilde{\mathcal{E}} \equiv \mathcal{E} \cap \mathcal{H}_{45} \neq \emptyset$, as shown in Eqs. (18)-(19) we have $F = \mathfrak{u} \cdot h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)$, with $h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1) = \inf_{c \in \mathbb{R}} h_{\mathcal{E}}(\mathfrak{u}_1(c))$, $\mathfrak{u}_1(c) \equiv \mathfrak{u}_1 - c[1 \ \ -1]^\intercal$, and $\mathfrak{u} \equiv (\mathfrak{u}_1 + \mathfrak{u}_2)$. Note that $h_{\tilde{\mathcal{E}}}(\cdot)$ is a Hadamard directionally differentiable function of $h_{\mathcal{E}}(\cdot)$, and its composition with the distance function ρ is again directionally differentiable. Hence, we use our DML estimator, Theorem 4.1, and the results in Fang and Santos (2019) to directly obtain the limit distribution of an estimator for $\rho(e^*, F)$. However, when $\mathcal{E} \subset \mathcal{H}_{45}^+$ (respectively, $\mathcal{E} \subset \mathcal{H}_{45}^-$), F is given by the support set of $\mathcal{E}$ in direction $(\mathfrak{u}_2 - \mathfrak{u}_1)$ (respectively, $(\mathfrak{u}_1 - \mathfrak{u}_2)$); see Eqs. (15)-(16). As discussed in Remark 5.1, carrying out inference if we use DML to directly estimate both coordinates of a support point is difficult, and therefore we use Eq. (11) to represent F through moments that involve $h_{\mathcal{E}}(\cdot)$ only; see Eqs. (58)-(59) below.

Observe that $\mathcal{E} \subset \mathcal{H}_{45}^+$ if and only if $F \in \mathcal{H}_{45}^+$, and $\mathcal{E} \subset \mathcal{H}_{45}^-$ if and only if $F \in \mathcal{H}_{45}^-$, as illustrated in Panels (b) and (d) of Figure 2. Hence, we partition the parameter space $\mathbb{B}_C$, to which F belongs Assumption 1, into three sets: $\mathbb{B}_C^+ \equiv \mathbb{B}_C \cap \mathcal{H}_{45}^+$, $\mathbb{B}_C^- \equiv \mathbb{B}_C \cap \mathcal{H}_{45}^-$, and $\mathbb{B}_C^{45} \equiv \mathbb{B}_C \cap \mathcal{H}_{45}$. We then have the following expressions for $\rho(e^*, F)$:

1. If $F \in \mathbb{B}_C^{45}$, $\rho(e^*, F) = \rho(e^*, \mathfrak{u} \cdot h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1))$.
2. If $F \in \mathbb{B}_C^+$, $\rho(e^*, F) = \rho(\tilde{e}, \tilde{F})$ for $(\tilde{e}, \tilde{F})$ satisfying:

$$\begin{cases} \tilde{e} - e^* = 0, \\ h_{\mathcal{E}}\big((\mathfrak{u}_2 - \mathfrak{u}_1)/\sqrt{2}\big) - \big((\mathfrak{u}_2 - \mathfrak{u}_1)/\sqrt{2}\big)^\intercal \tilde{F} = 0, \\ h_{\mathcal{E}}(q) \ - \ q^\intercal \tilde{F} \ \geq 0, \ \forall q \in \mathbb{S}^1. \end{cases} \tag{58}$$

3. If $F \in \mathbb{B}_C^-$, $\rho(e^*, F) = \rho(\tilde{e}, \tilde{F})$ for $(\tilde{e}, \tilde{F})$ satisfying:

$$\begin{cases} \tilde{e} - e^* = 0, \\ h_{\mathcal{E}}\big((\mathfrak{u}_1 - \mathfrak{u}_2)/\sqrt{2}\big) - \big((\mathfrak{u}_1 - \mathfrak{u}_2)/\sqrt{2}\big)^{\intercal} \tilde{F} = 0, \\ h_{\mathcal{E}}(q) \; - \; q^{\intercal} \tilde{F} \; \geq 0, \; \forall q \in \mathbb{S}^1. \end{cases} \tag{59}$$

In each of Eqs. (58)-(59), the first condition pins down $\tilde{e}$ to equal e^*; the second and third conditions restrict, respectively, $\tilde{F}$ to be a point on the supporting hyperplane of $\mathcal{E}$ in the appropriate direction and $\tilde{F} \in \mathcal{E}$, which together restricts $\tilde{F}$ to be the support set of $\mathcal{E}$ in this direction. We propose the following testing procedure:

PROCEDURE 3—Confidence Interval for $\rho(e^*, F)$:

1. *Construct estimators $\widehat{e}^*$ as in Eq. (54) and $\widehat{h}_{\mathcal{E}}\big(\cdot; \widehat{\boldsymbol{\theta}}\big)$ as in Theorem 4.1.*
2. *Emulate the construction in Step 1 of Procedure 1 to obtain two $(1-\alpha)$-level confidence sets for (e^*, F): let $\mathcal{CS}_n^+(e^*, F)$ denote the one for the moments in Eq. (58) (in the analog of Eq. (50), replace $\mathbb{B}_C$ with $\operatorname{cl}\mathbb{B}_C^+$), and $\mathcal{CS}_n^-(e^*, F)$ the one for the moments in Eq. (59) (in the analog of Eq. (50), replace $\mathbb{B}_C$ with $\operatorname{cl}\mathbb{B}_C^-$).*[7]
3. *For a given $(\tilde{e}, \tilde{F}) \in \mathbb{B}_C \times \mathbb{B}_C^{45}$ and $\widehat{h}_{\tilde{\mathcal{E}}}\left(\mathfrak{u}_1; \widehat{\boldsymbol{\theta}}\right) \equiv \inf_{c \in \mathbb{R}} \widehat{h}_{\mathcal{E}}\big(\mathfrak{u}_1(c); \widehat{\boldsymbol{\theta}}\big)$, let:*

$$T_n^{45}(\rho(\tilde{e}, \tilde{F})) \equiv \sqrt{n} \left| \rho\left(\widehat{e}^*, \mathfrak{u} \cdot \widehat{h}_{\tilde{\mathcal{E}}}(\mathfrak{u}_1; \widehat{\boldsymbol{\theta}})\right) - \rho(\tilde{e}, \tilde{F}) \right|. \tag{60}$$

Let ψ^{45} denote the random variable to which $T_n^{45}(\rho(\tilde{e}, \tilde{F}))$ converges in distribution for $\tilde{e} = e^*$ and $\tilde{F} = F_{45} \equiv \mathfrak{u} \cdot h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)$. Let c_β^{45} denote the β-quantile of ψ^{45} and $\varsigma > 0$ an infinitesimal uniformity factor. Use test inversion to construct the confidence set:

$$\mathcal{CS}_n^{45}(\rho(e^*, F)) = \left\{ \rho(\tilde{e}, \tilde{F}) : (\tilde{e}, \tilde{F}) \in \mathbb{B}_C \times \mathbb{B}_C^{45}, T_n^{45} \leq c_{1-\alpha+\varsigma}^{45} + \varsigma \right\}.$$

The expression for ψ^{45} is complex and given in Eq. (91), with a simpler expression provided in Eq. (92) for the case that Eq. (14) is satisfied and $\mathcal{E}$ has no kinks.

4. *Obtain a confidence interval for $\rho(e^*, F)$ as*

$$\mathcal{CS}_n^{\rho(e^*, F)} \equiv \left\{ \rho(\tilde{e}, \tilde{F}) : (\tilde{e}, \tilde{F}) \in \mathcal{CS}_n^+(e^*, F) \bigcup \mathcal{CS}_n^-(e^*, F) \right\} \bigcup \left\{ \mathcal{CS}_n^{45}(\rho(e^*, F)) \right\}.$$

[7] For a given set $\mathbb{B}$, we denote by $\operatorname{cl}\mathbb{B}$ its closure.

Intuitively, this construction inverts a test that jointly assesses the location of $\mathcal{E}$ relative to $\mathcal{H}_{45}$ *and* the value of $\rho(e^*, F)$. For example, if $F \in \mathcal{H}_{45}^-$, then both $\mathcal{CS}_n^+(e^*, F)$ and $\mathcal{CS}_n^{45}(\rho(e^*, F))$ are empty with probability approaching one (recall from Section 3.3 that $\inf_{c\in\mathbb{R}} h_{\mathcal{E}}(\mathfrak{u}_1(c))$ is unbounded when $F \notin \mathcal{H}^{45}$). Our next result shows that Procedure 3 delivers an asymptotically valid confidence interval.

PROPOSITION 7.1: *Let ρ be a Hadamard directionally differentiable distance function and the assumptions in Theorem 4.1 hold. Then the confidence interval constructed following Procedure 3 asymptotically covers the true $\rho(e^*, F)$ with probability at least $1-\alpha$.*

As we show in the proof, ψ^{45} is again a composition of Hadamard directionally differentiable functions that define $T_n^{\rho(\tilde{e},\tilde{F})}$ in Eq. (60). We can therefore employ the same bootstrap method detailed in Procedure 2 to consistently estimate the quantiles c_β^{45} for $\beta \in (0,1)$, where we replace $\phi(\cdot)$ by the composition of the directionally differentiable functions, including $\inf$ and ρ, that define T_n^{45}, whose exact expression we relegate to Appendix A.

8. MONTE CARLO EXPERIMENTS AND EMPIRICAL ILLUSTRATION

8.1. *Monte Carlo Simulations*

We evaluate the finite sample properties of the tests introduced in Sections 6-7 using two distinct data generating processes (DGPs). For both DGPs, the covariates $X \equiv [X_1, \ldots, X_{20}]^\intercal \in \mathbb{R}^{20}$ and group identity G are drawn from the following distributions:

$$X_2 \overset{d}{\sim} Unif(0,1),\ X_3 \overset{d}{\sim} Beta(2,2),\ G \overset{d}{\sim} Bern\,(0.6)\,;$$

$$\text{for } j \in \{1,4,...,20\},\ X_j \overset{d}{\sim} \mathcal{N}(0,1) \text{ truncated to } [-3,3].$$

We consider two different ways of generating the outcome Y:

1. Group-balanced DGP: $Y \,|\, G, X \overset{d}{\sim} Bern\left(\frac{G}{1+e^{-(X_1+X_2+0.5X_3)}} + \frac{(1-G)}{1+e^{-(-X_1-0.5X_2+X_4)}}\right)$;
2. r-skewed DGP: $Y \,|\, G, X \overset{d}{\sim} Bern\left(\frac{G}{1+e^{-2(X_1+X_2+X_3)}} + \frac{(1-G)}{1+e^{-0.7(X_1+0.5X_2+0.6X_4)}}\right)$,

where the group-balanced DGP is such that X is informative about Y in opposite directions for group r ($G=1$) and group b ($G=0$), but its predictive power is similar across groups. In contrast, the r-skewed DGP is such that X is systematically more informative about the

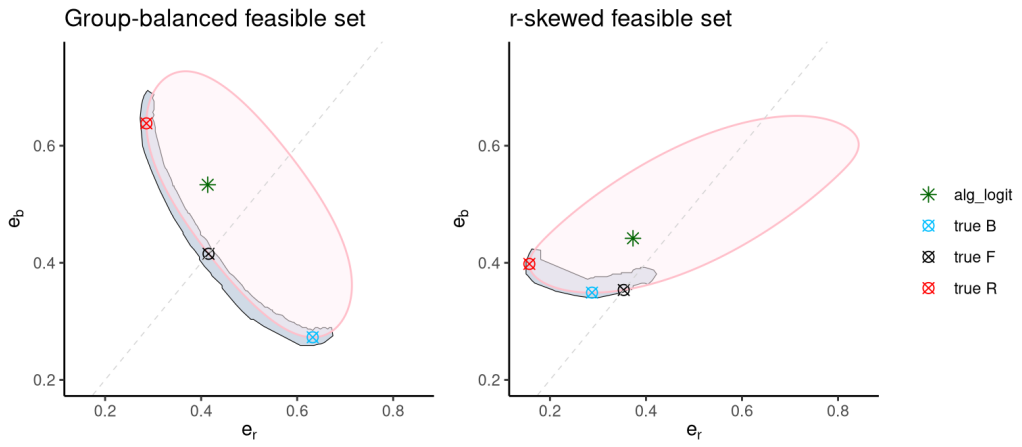


FIGURE 5.—True feasible set $\mathcal{E}$ for group-balanced DGP (left) and r-skewed DGP (right), based on evaluating $\mathcal{S}_{\mathcal{E}}(q)$ in Eq. (12) at true $\boldsymbol{\theta}$ in 500 directions, with expectations approximated by averaging over 10^7 observations drawn from the respective DGP. True R and B obtained similarly, using Eq. (13). True F based on evaluating Eq. (19) at true $\boldsymbol{\theta}$ and minimizing over c via stochastic gradient descent. True risk e^* (green asterisk) induced by the logit algorithm based Eq. (5). Shaded region: 95% confidence set for $\mathcal{F}$ built using $10,000$ observations drawn from the respective DGP and test inversion detailed in Section 5.2 with $\Delta\boldsymbol{\theta}$ estimated by logit lasso.

group-r outcome. We take the loss function to be the classification error, $\ell(d, y) = \mathbb{1}\{d \neq y\}$. We also construct a status quo algorithm a^*, that we fix in the simulations, by training once a logistic regression on a sample of size $10,000$, with $5,000$ observations from the balanced DGP and $5,000$ observations from the r-skewed DGP.[8]

Figure 5 depicts the population feasible set $\mathcal{E}$ corresponding to each DGP (pink region), along with a 95% confidence set for the frontier $\mathcal{F}$ (shaded grey region), constructed using a random sample of $10,000$ observations drawn from the respective DGPs, and the risk e^* induced by the status-quo algorithm a^* (green asterisk). Throughout Section 8.1, we fit the nuisance parameter $\Delta\boldsymbol{\theta}$ using logit lasso and 5-fold sample splitting. To assess the finite sample properties of the LDA test in Section 6.2 and the distance-to-F test in Section 7, we test whether e^* is on the frontier and whether it is at a specific distance from F.

[8]We train the logistic regression on a mixture of group-balanced and r-skewed data so that we test for existence of an LDA to the same algorithm a^* in both DGPs. DGP-specific logistic regressions trained on $10,000$ observations drawn form that DGP for each case yield group risks e^* very close to the ones plotted in Figure 5.

Table I reports the simulation results. Overall, the Monte Carlo exercise suggests good finite sample properties for our proposed tests, especially as the sample size increases. The top panel corresponds to the test for weak group skew (Section 6.1), where the third column reports the frequency with which the 95% confidence set for the vector (R, B) (Eq. 50) based on the available sample fails to cover the true value of (R, B). While the r-skewed DGP exhibits some over-rejection at $n = 1{,}000$, which is a small sample size relative to the complexity of DML estimation with 20 covariates and estimating the support function across directions, this over-rejection quickly disappears as sample size increases. On the other hand, the weak group skew test (fourth column) rejects the null of weak group skew in the balanced DGP and fails to reject it in the r-skewed DGP essentially with probabilities 1 and 0, respectively. This is not surprising because the simulation DGPs are far from the boundary of the null, and because the test is conservative due to the projection step.

The middle panel reports the LDA test results (Section 6.2). The third (fourth) column shows the frequency with which the population point R (B), which by definition belongs to $\mathcal{F}$, is rejected by the test of the null that it belongs to $\mathcal{F}$. The fifth (sixth) column reports the frequency with which $(R+B)/2$ (the risk e^* associated with the logit algorithm a^*), which by construction does not belong to $\mathcal{F}$, is rejected by the test as an element of $\mathcal{F}$. While at small sample size ($n = 1,000$), the test exhibits some over-rejection for the point B, this quickly disappears as sample size grows. At small and medium sample sizes ($n = 1,000$ and $n = 5,000$), the rejection probability for the false null that $(R+B)/2 \in \mathcal{F}$ is low for the r-skewed DGP, while it is high at all sample sizes for the balanced DGP. This is justifiable in light of Figure 5, which shows that in the r-skewed DGP the chord between R and B is close to the frontier and lies largely inside its 95% confidence set. For $n = 10,000$, the test detects the false nulls in columns five and six with substantially higher probability.

The bottom panel reports the results for the distance-to-F test (Section 7). This test is more delicate because its implementation requires solving an optimization problem to estimate F. Here we observe more substantial over-rejection for $n = 1,000$, but the distortion gets markedly reduced as sample size increases.

TABLE I
REJECTION RATES (1000 MONTE CARLO SIMULATIONS, $\alpha = 0.05$)

Test H_0 : *Weak Group Skew*			
n	DGP	$(R,B) \notin \mathcal{CS}_n(R,B)$	H_0 rejected
$1,000$	balance	0.035	0.999
	r-skew	0.12	0
$5,000$	balance	0.012	1
	r-skew	0.021	0
$10,000$	balance	0.01	1
	r-skew	0.012	0

Test H_0 : *There Is No LDA to $\tilde{e}$*					
n	DGP	$\tilde{e} = R$	$\tilde{e} = B$	$\tilde{e} = (R+B)/2$	$\tilde{e} = e^*$
$1,000$	balance	0.046	0.104	0.145	0.397
	r-skew	0.051	0.236	0.073	0.162
$5,000$	balance	0.026	0.037	0.993	1
	r-skew	0.018	0.043	0.059	1
$10,000$	balance	0.026	0.029	1	1
	r-skew	0.007	0.039	0.264	1

Test $\mathsf{H}_0 : \rho(\tilde{e}, F) = \delta$ *for Constant δ Equal to the True Distance*					
n	DGP	$\tilde{e} = R$	$\tilde{e} = B$	$\tilde{e} = (R+B)/2$	$\tilde{e} = e^*$
$1,000$	balance	0.242	0.273	0.251	0.159
	r-skew	0.063	0.048	0.052	0.204
$5,000$	balance	0.081	0.097	0.088	0.051
	r-skew	0.019	0.027	0.017	0.077
$10,000$	balance	0.069	0.078	0.073	0.046
	r-skew	0.012	0.024	0.013	0.034

Population values for the balanced DGP are $R = [0.286, 0.638]^\intercal, B = [0.632, 0.273]^\intercal, F = [0.415, 0.415]^\intercal$, $e^* = [0.414, 0.533]^\intercal$; for the r-skewed DGP are $R = [0.157, 0.398]^\intercal, B = [0.288, 0.349]^\intercal, F = [0.354, 0.354]^\intercal$, $e^* = [0.373, 0.442]^\intercal$. We take ρ to be the squared Euclidean distance.

8.2. *Empirical Illustration*

We revisit the analysis in Obermeyer, Powers, Vogeli, and Mullainathan (2019, OPVM henceforth), who analyze properties of the algorithm used by a research hospital to determine if a patient should be automatically enrolled in a high-risk care management program.

The algorithm used by the research hospital aims at predicting patients' health needs based on total medical expenditures (the label on which the algorithm is trained). It produces a health risk score and automatically enrolls a patient in the high-risk care management program if that patient's risk score exceeds the 97^{th} percentile of all predicted scores. We reassess this algorithm through our testing procedures. To do so, we use the synthetic data made available by Li, Lin, and Obermeyer (2019) at GitLab to replicate all analyses in OPVM. The data include $48,784$ patient observations, of which $5,582$ self report as Black and the others self report as White ($g \in \{\mathsf{bl}, \mathsf{wh}\}$). The data include 149 covariates, such as age, gender, comorbidity and medication variables, costs, and biomarkers, and provide information about each patient's number of active chronic conditions in the subsequent year, viewed as the true measure of health needs of a patient. Following LLMO, we let $\ell(d, Y) = \mathbb{1}\{Y \neq d\}$ be the classification loss and, unless explicitly stated otherwise, $Y_i = \mathbb{1}\{\text{patient } i \text{ has 6 or more chronic conditions}\}$, with the choice of 6 driven by the fact that it is the 97^{th} percentile of active chronic condition numbers across patients in the sample. We use random forests with $5,000$ trees to estimate $\Delta\boldsymbol{\theta}$ as described in Section 4.

8.2.1. *Feasible Set Estimation and Inference for the Frontier*

We report in Figure 6 an estimate of the feasible set $\mathcal{E}$ based on Eq. (28) using $1,000$ directions (top-left panel), along with 100 estimated supporting hyperplanes (top-right panel), where across all panels the horizontal (vertical) axis is the risk for Blacks (Whites), denoted as e_{bl} (e_{wh}). We zoom in to show the estimated FA-frontier $\widehat{\mathcal{F}}$ and its 95% confidence set (bottom-left panel), and zoom in further to show the best risk achievable for the Black patients (bottom-right panel, red point labeled BL), which coincides with and is overlaid by the best risk for the White patients (bottom-right panel, blue point labeled WH).

8.2.2. *Hypothesis Testing*

We next report the weak group skew test results in Figure 7-Panel (a), where both BL and WH are below the 45° degree line and the feasible set is wh-skewed; the test fails to reject the null of weak group skew, suggesting that implementing Pareto-dominated algorithms could be justified based on the designer's preferences over fairness and accuracy. In

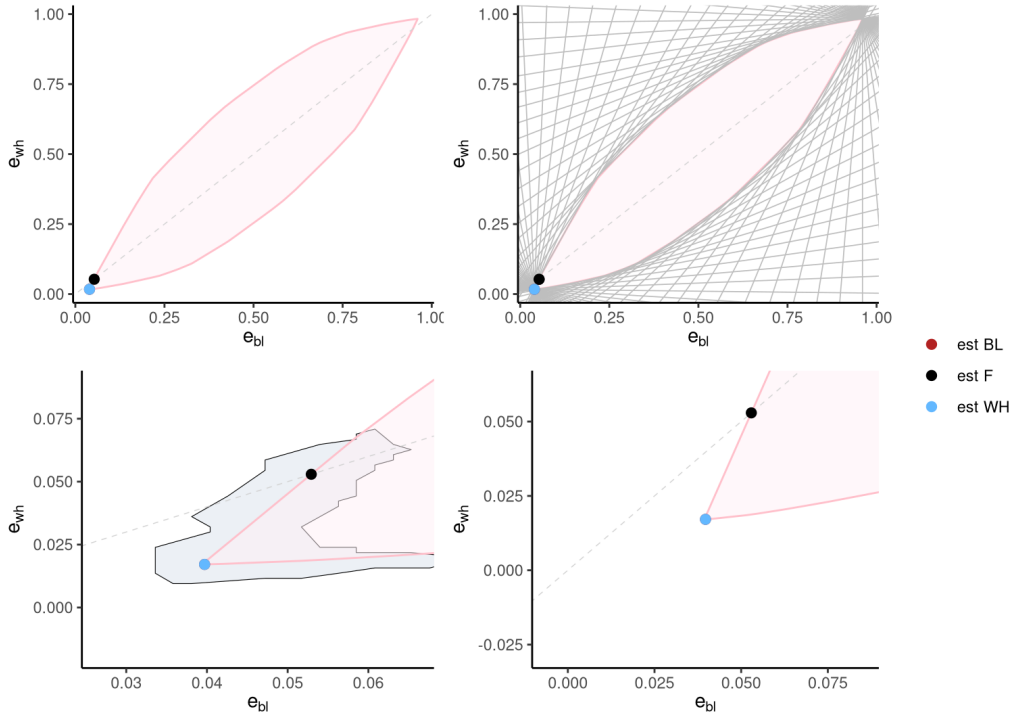
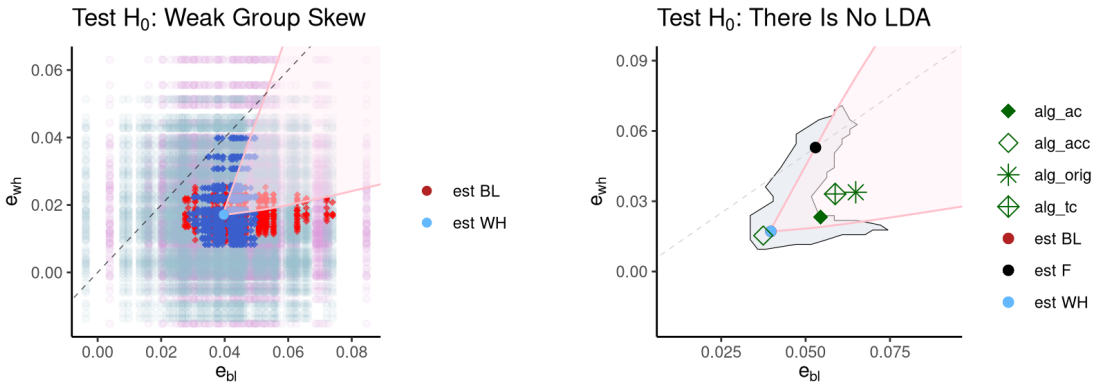


FIGURE 6.—Top-left panel: $\widehat{\mathcal{E}}$; top-right panel: $\widehat{\mathcal{E}}$ along with one hundred supporting hyperplanes; bottom-left panel: zoom-in to $\widehat{\mathcal{F}}$ and the 95% confidence set around this frontier; bottom-right panel: further zoom-in to the best group-specific points BL and WH, and the fairest point F.

Figure 7-Panel (b) we plot $\widehat{\mathcal{F}}$, along with its 95% confidence set. Figure 7-Panel (b) also plots the estimated group risks associated with the original algorithm used by the hospital (a green asterisk) and three alternative algorithms that OPVM experiment with to assess whether different label choices yield decision rules that are more accurate and fairer than the algorithm currently used by the hospital. One of these algorithms is trained to predict total cost (hollow diamond with a cross in Figure 7-Panel (b)), one to predict avoidable costs (filled diamond), and the other one to predict the number of active chronic conditions (hollow diamond). All our tests take into account the finite sample estimation error of the group risks induced by these four algorithms. The top panel of Table II reports the values of the LDA test statistics and the associated critical values that we compute for the four algorithms considered by OPVM. The hypothesis that the original algorithm yields group risks on the frontier is rejected, and so is the same hypothesis for group risks associated



(a) Candidate values for (BL, WH)

(b) 3 experimental algorithms

FIGURE 7.—Panel (a): Plum-colored (respectively, light-blue colored) circles correspond to candidate values for the best point for Blacks BL (Whites WH) sampled from a normal distribution centered at the estimated BL (WH), and red (blue) diamonds correspond to non-rejected values. Panel (b): $\widehat{\mathcal{F}}$ along with its 95% confidence set and the estimated group risks for four algorithms considered by OPVM: the original algorithm used by the hospital (asterisk); one that predicts total cost (hollow diamond with a cross); one that predicts avoidable costs (filled diamond); and one that predicts the number of active chronic conditions (hollow diamond).

with the algorithm trained to predict total costs. On the other hand, we fail to reject that the algorithms trained to predict avoidable costs and the number of active chronic conditions yield group risks on the FA-frontier. Regarding the three algorithms proposed by OPVM, if one were to plot the set $\mathcal{C}^*$ corresponding to the original algorithm, it would be immediate to see that one cannot reject the hypothesis that all other algorithms considered by OPVM improve upon it, both in terms of fairness and accuracy.

The bottom panel of Table II shows that the closest risk to F is the one associated with the algorithm predicting total costs. However, all confidence intervals overlap.

8.2.3. *Performance of Algorithms Constructed to Be on a Constrained Frontier*

Finally, we compare various outcomes associated with the algorithms considered by OPVM with those of decision rules resulting from the algorithms that we propose in Section 5.4. To do so, we randomly split the sample into two halves, and use one half (training data) to estimate the nuisance parameter $\Delta\boldsymbol{\theta}$, and implement Eq. (45) on the other half (evaluation data) for the following choices of q:

TABLE II

RESULTS FOR THE LDA TEST AND CONFIDENCE SETS FOR THE DISTANCE TO F (FOR $\alpha = 0.05$)

Test H_0 : ***There Is No LDA***				
	Original	Total Costs	Avoid. Costs	Act. Chr. Cond.
Estimated Risks	$(0.065, 0.034)$	$(0.059, 0.033)$	$(0.054, 0.023)$	$(0.037, 0.015)$
Test Statistic	3.767	2.753	1.428	0.541
Critical Value	1.885	1.948	1.732	1.642
Conclusion	Rejected	Rejected	Not Rejected	Not Rejected
Distance to $F = (0.049, 0.049)$				
Estimated Distance	0.0005	0.0004	0.0007	0.0013
Confidence Set	$(0.000, 0.001)$	$(0.000, 0.001)$	$(0.000, 0.002)$	$(0.000, 0.003)$

Top panel: LDA test statistics and 0.05-level critical values associated with the original algorithm and the three experimental algorithms (predicting, respectively, total costs; avoidable costs; number of active chronic conditions) analyzed by OPVM. Bottom panel: estimated squared-Euclidean distance to the F point and corresponding confidence set for this distance.

(i) $q = [-1 \quad 0]^\intercal$, yielding an algorithm that (asymptotically) achieves the best point on $\mathcal{F}$ for Black patients, which we call Rawlsian because $e_{\mathsf{bl}} > e_{\mathsf{wh}}$ at BL.

(ii) $q = [0 \quad -1]^\intercal$, yielding an algorithm that (asymptotically) achieves the best point on $\mathcal{F}$ for White patients, which we call Majority as Whites are the majority of the sample.

(iii) $\widehat{q}(\widehat{F})$, yielding an algorithm based on a direction estimated using the entire sample that (asymptotically) achieves F, the fairest point on $\mathcal{F}$, as established in Proposition 5.4, which we call Egalitarian.

(iv) $q = -\frac{\sqrt{2}}{2}[1 \;\; 1]^\intercal$, yielding an algorithm that weighs both groups equally, which we call Utilitarian as it can be expressed as a generalization of the Utilitarian rule.

Throughout, we recognize that to compare algorithms we should enforce a global capacity constraint on the total percentage of patients assigned to the high-risk case management program, so that variation across algorithms is not confounded with a possibly more generous care program. To enforce the capacity constraint, we need to require the algorithms to satisfy $\int a(x) d\mathbb{P}_X \leq \bar{a}$ for some known constant $\bar{a}$. For example, $\bar{a} = 0.03$ when comparing with the algorithm currently used by the hospital which assigns only patients with risk score above the 97^{th} percentile to the high-risk care program. Recall from the discussion

following Proposition 3.1 that without capacity constraints, we have

$$h_{\mathcal{E}}(q) = \mathbb{E}\left[q_1 \frac{\theta_0^r(X)}{\mu_r} + q_2 \frac{\theta_0^b(X)}{\mu_b}\right] + \max_{a \in \mathcal{A}(\mathcal{X})} \mathbb{E}\left[a(X) k\left(\boldsymbol{\theta}(X), \mathcal{M}q\right)\right]. \tag{61}$$

When $\mathcal{A}(\mathcal{X})$ is constrained to only include algorithms such that $\int a(x) d\mathbb{P}_X \leq \bar{a}$, maximization in Eq. (61) is achieved by setting

$$a^{\text{opt}}(X; q) = \mathbb{1}\{k\left(\boldsymbol{\theta}(X), \mathcal{M}q\right) > \max(0, \text{quant}_{k(\boldsymbol{\theta}(X), \mathcal{M}q)}(1 - \bar{a}))\}, \tag{62}$$

for $\text{quant}_{k(\boldsymbol{\theta}(X), \mathcal{M}q)}(\alpha)$ the α-quantile of $k(\boldsymbol{\theta}(X), \mathcal{M}q)$ and $k(\boldsymbol{\theta}(X), \mathcal{M}q)$ as in Eq. (22).[9] In words, for our example with $\bar{a} = 0.03$, high-risk care management is assigned to those patients with positive values of $k(\boldsymbol{\theta}(X), \mathcal{M}q)$ that exceed its 97^{th} percentile.

The results of our first exercise are reported in Figure 8. The top panel replicates Figure 1-(a) in OPVM, and shows that at each percentile of the original algorithm's risk score (the horizontal axis), Black patients have a substantially larger number of active chronic conditions (the vertical axis) than White patients.[10] Even among patients automatically enrolled in the high-risk care management program (those with a risk score above the 97^{th} percentile), Black patients appear to be in worse health than White patients.

We ask whether the algorithms that we propose are able to select for treatment patients that, in fact, exhibit substantially worse health outcomes one year later. The bottom four panels in Figure 8 plot, for each of the algorithms described at the beginning of this section, the number of active chronic conditions for (a) Black patients that our algorithm does not assign to the high-risk care program (dash-dotted, light-purple line); (b) White patients that our algorithm does not assign to the high-risk care program (solid salmon line); (c) Black patients that our algorithm assigns to the high-risk care program (dash-dotted, dark-purple line); (b) White patients that our algorithm assigns to the high-risk care program (solid

[9] In this case, the *constrained* feasible set is $\mathcal{E}_{\bar{a}}^{\subset\circ} \equiv \left\{\left(e_r(a), e_b(a)\right) \in \mathbb{R}^2 : a \in \mathcal{A}(\mathcal{X}) \text{ and } \int a(x) d\mathbb{P}_X \leq \bar{a}\right\}$, a convex subset of $\mathcal{E}$, and the algorithms in Eq. (62) is on its frontier.

[10] The top panel of Figure 8 is not identical to Figure 1-(a) of OPVM because it is plotted using the synthetic data, instead of the real data, and because the code used to generate Figure 1-(a) of OPVM had an error that was later corrected after the publication of OPVM; see the documentation of Li et al. (2019) for more details.

orange line). For comparability with the top panel, on the horizontal axis we continue to report the risk score produced by the algorithm used by the hospital.

The main takeaway from Figure 8 is that our four proposed algorithms, which use group identity to estimate $\Delta\boldsymbol{\theta}$ but not for treatment assignment, are successful at selecting for treatment patients who, one year later, experience substantially worse health outcomes. Moreover, with the Ralwsian, Majority, and Utilitarian algorithms, Black and White patients assigned to treatment have similar numbers of chronic conditions. Only the Egalitarian algorithm shows notable disparities for patients with hospital risk scores below the 70^{th} percentile. We think this result might be due to two factors: estimation of F is challenging and hence the direction $\widehat{q}(\widehat{F})$ might be imprecisely estimated; and the feasible set is wh-skewed and hence the Egalitarian algorithm leads to a Pareto-dominated outcome.

Our last exercise aims at further assessing the extent to which our proposed algorithms may reduce the substantial disparities between Black and White patients in the current program screening practices documented by OPVM. To illustrate the potential for improvement over the hospital's algorithm, OPVM simulate "a counterfactual world with no gap in health conditional on risk" (p.3). They construct an infeasible, *couterfactual* algorithm that uses group identity and patients' ex-post active chronic health conditions to find the sickest Black patient with health risk score just below a threshold (the "inframarginal Black patient") and the healthiest White patient with health risk score just above the same threshold (the "supramarginal White patient"). If the number of chronic health conditions of the inframarginal Black patient is larger than that of the supramarginal White patient, they iteratively swap them until the number of chronic health conditions of the inframarginal Black patient equals that of the supramarginal White patient. OPVM find that at all risk thresholds α above the median, the counterfactual algorithm increases the fraction of Black patients treated. Table III, columns 2-3, show that across the various thresholds for automatic enrollment in the program that we consider, the fraction of Black patients would rise between 6 and 41 percentage points.[11]

[11] In columns 2-3, the fractions reported are based on a denominator that equals the total number of patients with risk scores above a certain percentile of risk scores, e.g., the 55^{th} percentile, with this threshold viewed as

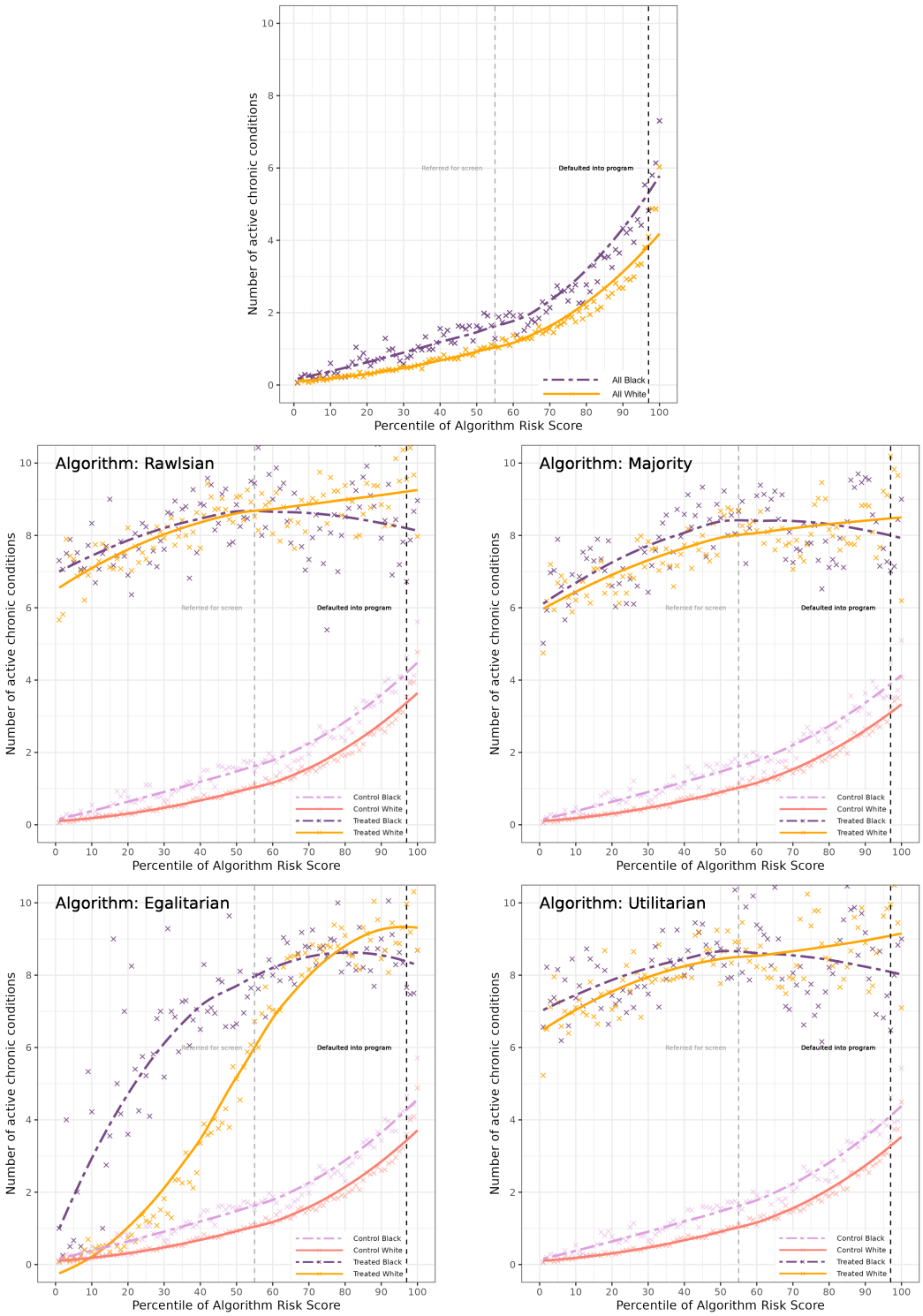


FIGURE 8.—Average number of active chronic conditions within each risk-score percentile bin by treatment group under the alternative algorithms on the FA frontier subject to 3% capacity constraint, averaged across 20 replications of the 50-50 split.

TABLE III

FRACTION OF BLACK PATIENTS TREATED AMONG ALL TREATED

	Algorithms from Obermeyer et al.		Algorithms on the FA-Frontier			
Capacity Threshold	Original	Counterfactual	Rawlsian	Majority	Egalitarian	Utilitarian
55	0.120	0.184	0.164	0.164	0.160	0.164
69	0.128	0.255	0.185	0.184	0.163	0.185
82	0.138	0.327	0.206	0.207	0.160	0.206
89	0.151	0.407	0.218	0.223	0.166	0.219
94	0.167	0.498	0.286	0.262	0.190	0.273
97	0.184	0.592	0.368	0.300	0.253	0.338

The distribution of the number of active chronic conditions is such that the 55^{th} to the 68^{th} percentiles all correspond to 1 active chronic condition, the 69^{th}-81^{st} correspond to 2, the 82^{nd}-88^{th} correspond to 3, the 89^{th}-92^{nd} correspond to 4, the 94^{th}-95^{th} correspond to 5, and the 96^{th}-97^{th} correspond to 6.

We ask how much of this gap could be filled if one were to use our algorithms, which are feasible and do not rely on ex-post knowledge of the number of active chronic conditions nor on using group identity for assignment to the high-risk care program. The results, reported in Table III, columns 4-7,[12] show that each of our four algorithm yields an increase in the fraction of Black patients treated at all capacity thresholds. The Rawlsian and the Utilitarian algorithms yield the largest increases, ranging between 4 and 18 percentage points across the various capacity thresholds. This shows that these two feasible, easy to implement algorithms can close between $26\% - 69\%$ of the gap between the algorithm that the hospital uses and the counterfactual, infeasible algorithm simulated by OPVM.

We conclude by noting that results based on random forests trained using the `grf` package are not guaranteed to reproduce exactly across platforms, even with the same seed (all simulations and estimation are in `R`). This is a known feature of `grf` (see the reference manual). Similarly, tests based on optimization via stochastic gradient descent implemented by

---

the threshold above which the patient is automatically enrolled in the high-risk care program, and a numerator that equals the number of Black patients with risk score above that percentile.

[12]For each algorithm, the fraction of Black patients treated at each capacity threshold is computed as the following ratio. The denominator equals the total number of patients treated under a given algorithm, e.g., the Rawlsian algorithm, subject to the constraint that at each threshold $\bar{a} \in [0.55, 0.97]$ at most $100(1-\bar{a})\%$ of the evaluation sample is treated, and for each threshold $\bar{a}$ the outcome Y equals the indicator of whether the number of active chronic conditions exceeds the $\bar{a}$-quantile of the distribution of the number of active chronic conditions.

the `torch` package are not exactly reproducible, even after seed setting, which affects the F estimator and thus the distance-to-F test (see the repository discussion). To assess sensitivity of our results to these features, we repeat the entire empirical exercise 20 times with different seeds. Appendix C reports results and includes a robustness check using logit lasso for estimating the nuisance function.[13] While the results exhibit some nontrivial variation across seeds (although the qualitative results are unchanged), we view this as expected, considering that the nonparametric estimation step involves 149 covariates against a total sample size of $48,784$ and a minority group of size $5,582$.

9. CONCLUSION

We provide a consistent nonparametric estimator for a theoretical fairness-accuracy frontier proposed by LLMO and algorithms that attain points on this frontier. We obtain the estimator through judicious use of the separating hyperplane theorem and the support function of the (convex) feasible set of expected losses associated with all possible algorithms, a portion of whose boundary coincides with the FA-frontier. We provide a DML estimator of the support function and show it converges to a tight Gaussian process as sample size increases. We formulate important policy-relevant hypotheses that have received much attention in the fairness literature as restrictions on the support function and construct valid test statistics. We provide an estimator for the distance between a given algorithm and the fairest point on the frontier. We carry out a Monte Carlo exercise that illustrates the good finite sample properties of our method, and demonstrate its practical relevance by revisiting the empirical analysis in OPVM. Our results show that the algorithm that a research hospital employs to screen patients for high-risk care is not on the frontier. Our proposed algorithms substantially improve over this status quo in terms of both fairness and accuracy.

REFERENCES

ANDREWS, DONALD W. K. (1994): "Chapter 37 Empirical process methods in econometrics," Elsevier, vol. 4 of *Handbook of Econometrics*, 2247–2294. [57]

[13] For the simulations in Section 8.1, we expect cross-platform variability to be negligible, as results are averaged over 1000 Monte Carlo replications.

ANDREWS, DONALD W. K. AND XIAOXIA SHI (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–666. [21, 30]

ANGWIN, JULIA, JEFF LARSON, SURYA MATTU, AND LAUREN KIRCHNER (2016): "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks." *ProPublica*, 23, 77–91. [2]

ARNOLD, DAVID, WILL DOBBIE, AND PETER HULL (2021): "Measuring racial discrimination in algorithms," in *AEA Papers and Proceedings*, vol. 111, 49–54. [2]

AUERBACH, ERIC, ANNIE LIANG, MAX TABORD-MEEHAN, AND KYOHEI OKUMURA (2024): "Testing the Fairness-Improvability of Algorithms," *arXiv preprint arXiv:2405.04816*. [5]

BAROCAS, SOLON, MORITZ HARDT, AND ARVIND NARAYANAN (2023): *Fairness and Machine Learning: Limitations and Opportunities*, MIT Press, http://www.fairmlbook.org. [4]

BELLONI, ALEXANDRE, VICTOR CHERNOZHUKOV, IVAN FERNANDEZ-VAL, AND CHRISTIAN HANSEN (2017): "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 85, 233–298. [17]

BERESTEANU, ARIE AND FRANCESCA MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814. [4, 20, 58, 61]

BERK, RICHARD, HODA HEIDARI, SHAHIN JABBARI, MATTHEW JOSEPH, MICHAEL KEARNS, JAMIE MORGENSTERN, SETH NEEL, AND AARON ROTH (2017): "A convex framework for fair regression," *arXiv preprint arXiv:1706.02409*. [5]

BERK, RICHARD, HODA HEIDARI, SHAHIN JABBARI, MICHAEL KEARNS, AND AARON ROTH (2021): "Fairness in Criminal Justice Risk Assessments: The State of the Art," *Sociological Methods & Research*, 50, 3–44. [2]

BICKEL, PETER J (1982): "On adaptive estimation," *The Annals of Statistics*, 647–671. [16]

BLATTNER, LAURA AND JANN SPIESS (2022): "Machine Learning Explainability and Fairness: Insights from Consumer Lending," *FinRegLab Empirical White Paper*. [3]

BONTEMPS, CHRISTIAN, THIERRY MAGNAC, AND ERIC MAURIN (2012): "Set identified linear models," *Econometrica*, 80, 1129–1155. [4, 11]

CÁRCAMO, JAVIER, ANTONIO CUEVAS, AND LUIS-ALBERTO RODRÍGUEZ (2020): "Directional differentiability for supremum-type functionals: Statistical applications," *Bernoulli*, 26, 2143 – 2175. [30, 64, 69]

CHANDRASEKHAR, ARUN, VICTOR CHERNOZHUKOV, FRANCESCA MOLINARI, AND PAUL SCHRIMPF (2018): "Best linear approximations to set identified functions: with an application to the gender wage gap," . [4, 11, 52]

CHEN, QIZHAO, MORGANE AUSTERN, AND VASILIS SYRGKANIS (2023): "Inference on Optimal Dynamic Policies via Softmax Approximation," *arXiv preprint arXiv:2303.04416*. [17]

CHERNOZHUKOV, VICTOR, DENIS CHETVERIKOV, MERT DEMIRER, ESTHER DUFLO, CHRISTIAN HANSEN, WHITNEY NEWEY, AND JAMES ROBINS (2018): "Double/debiased machine learning for treatment and structural parameters," . [16, 17, 18, 56]

CHERNOZHUKOV, VICTOR, JUAN CARLOS ESCANCIANO, HIDEHIKO ICHIMURA, WHITNEY K NEWEY, AND JAMES M ROBINS (2022): "Locally robust semiparametric estimation," *Econometrica*, 90, 1501–1535. [16]

CHERNOZHUKOV, VICTOR, HAN HONG, AND ELIE TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [21, 58]

CHOULDECHOVA, ALEXANDRA AND AARON ROTH (2018): "The Frontiers of Fairness in Machine Learning," *arXiv preprint arXiv:1810.08810*. [4]

CORBETT-DAVIES, SAM, JOHANN D. GAEBLER, HAMED NILFOROSHAN, RAVI SHROFF, AND SHARAD GOEL (2024): "The measure and mismeasure of fairness," *Journal of Machine Learning Research*, 24. [4]

COWGILL, BO AND CATHERINE E. TUCKER (2020): "Algorithmic Fairness and Economics," available at http://dx.doi.org/10.2139/ssrn.3361280. [2]

DWORK, CYNTHIA, MORITZ HARDT, TONIANN PITASSI, OMER REINGOLD, AND RICHARD ZEMEL (2012): "Fairness through Awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, New York, NY, USA: Association for Computing Machinery, 214–226. [5]

FANG, ZHENG AND ANDRES SANTOS (2019): "Inference on directionally differentiable functions," *The Review of Economic Studies*, 86, 377–412, with Online Appendix at academic.oup.com/restud/article/86/1/377/5094886#supplementary-data. [4, 21, 30, 33, 64, 65, 66, 69]

ICHIMURA, HIDEHIKO AND WHITNEY K NEWEY (2022): "The influence function of semiparametric estimators," *Quantitative Economics*, 13, 29–61. [16]

KAIDO, HIROAKI (2016): "A dual approach to inference for partially identified econometric models," *Journal of Econometrics*, 192, 269 – 290. [4, 23, 24, 60, 61, 63, 64]

KAIDO, HIROAKI AND ANDRES SANTOS (2014): "Asymptotically efficient estimation of models defined by convex moment inequalities," *Econometrica*, 82, 387–413. [4]

KALLUS, NATHAN, XIAOJIE MAO, AND ANGELA ZHOU (2022): "Assessing Algorithmic Fairness with Unobserved Protected Class Using Data Combination," *Management Science*, 68, 1959–1981. [5]

KLEINBERG, JON, JENS LUDWIG, SENDHIL MULLAINATHAN, AND ASHESH RAMBACHAN (2018a): "Algorithmic fairness," in *AEA Papers and Proceedings*, vol. 108, 22–27. [5]

KLEINBERG, JON, JENS LUDWIG, SENDHIL MULLAINATHAN, AND CASS R SUNSTEIN (2018b): "Discrimination in the Age of Algorithms," *Journal of Legal Analysis*, 10, 113–174. [3]

LI, ZOEY, KATIE LIN, AND ZIAD OBERMEYER (2019): "labsysmed/dissecting-bias: Public Release Synthetic Data and Code for "Dissecting racial bias in an algorithm used to manage the health of populations" by Obermeyer, Powers, Vogeli, and Mullainathan (2019)," Version f7a3b4d86e0dd3c71b1bd28a6706dc5fd85548ad. Accessed at https://gitlab.com/labsysmed/dissecting-bias on May 16, 2025. [39, 43]

LIANG, ANNIE, JAY LU, XIAOSHENG MU, AND KYOHEI OKUMURA (2024): "Algorithm Design: A Fairness-Accuracy Frontier," *arXiv preprint arXiv:2112.09975v5*. [1, 2, 5, 7, 8, 14, 20, 27, 39, 47]

LIU, YIQI AND FRANCESCA MOLINARI (2024): "Inference for an Algorithmic Fariness-Accuracy Frontier," *arXiv preprint arXiv:2402.08879v1*. [18, 23]

MANSKI, CHARLES F., JOHN MULLAHY, AND ATHEENDAR S. VENKATARAMANI (2023): "Using measures of race to make clinical predictions: Decision making, patient health, and fairness," *Proceedings of the National Academy of Sciences*, 120, e2303370120. [25]

MOLCHANOV, I. (2017): *Theory of Random Sets*, London: Springer, 2 ed. [61]

MOLCHANOV, ILYA AND FRANCESCA MOLINARI (2018): *Random Sets in Econometrics*, Econometric Society Monograph Series, Cambridge University Press, Cambridge UK. [9, 19, 51, 71, 72]

MOLINARI, FRANCESCA (2020): "Microeconometrics with Partial Identification," in *Handbook of Econometrics*, Volume 7A, ed. by Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, Amsterdam: Elsevier, 355–486. [4]

NEWEY, WHITNEY K (1994): "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, 1349–1382. [15]

NEYMAN, JERZY (1959): "Optimal asymptotic tests of composite hypotheses," *Probability and statsitics*, 213–234. [16]

——— (1979): "C (α) tests and their use," *Sankhyā: The Indian Journal of Statistics, Series A*, 1–21. [16]

OBERMEYER, ZIAD, BRIAN POWERS, CHRISTINE VOGELI, AND SENDHIL MULLAINATHAN (2019): "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, 366, 447–453. [2, 4, 6, 38, 39, 40, 41, 42, 43, 44, 46, 47, 74, 77]

PARK, GYUNGBAE (2024): "Debiased Machine Learning when Nuisance Parameters Appear in Indicator Functions," *arXiv preprint arXiv:2403.15934*. [17]

RAMBACHAN, ASHESH, JON KLEINBERG, JENS LUDWIG, AND SENDHIL MULLAINATHAN (2020a): "An economic perspective on algorithmic fairness," in *AEA Papers and Proceedings*, vol. 110, 91–95. [5]

RAMBACHAN, ASHESH, JON KLEINBERG, SENDHIL MULLAINATHAN, AND JENS LUDWIG (2020b): "An economic approach to regulating algorithms," Tech. rep., National Bureau of Economic Research. [5]

RAMBACHAN, ASHESH AND JONATHAN ROTH (2020): "Bias In, Bias Out? Evaluating the Folk Wisdom," in *1st Symposium on Foundations of Responsible Computing (FORC 2020)*, ed. by Aaron Roth, vol. 156 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 6:1–6:15. [5]

ROBINS, JAMES, LINGLING LI, ERIC TCHETGEN TCHETGEN, AAD VAN DER VAART, ET AL. (2008): "Higher order influence functions and minimax estimation of nonlinear functionals," in *Probability and statistics: essays in honor of David A. Freedman*, Institute of Mathematical Statistics, vol. 2, 335–422. [16]

ROBINS, JAMES M, LINGLING LI, RAJARSHI MUKHERJEE, ERIC TCHETGEN TCHETGEN, AND AAD VAN DER VAART (2017): "Minimax estimation of a functional on a structured high-dimensional model," . [16]

ROBINSON, PETER M (1988): "Root-N-consistent semiparametric regression," *Econometrica: Journal of the Econometric Society*, 931–954. [17]

ROCKAFELLAR, R. T. (1997): *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Princeton University Press. [9, 14, 51]

SCHNEIDER, ROLF (1993): *Convex Bodies: The Brunn-Minkowski Theory*, Encyclopedia of Mathematics and its Applications, Cambridge University Press. [11, 14, 53, 68]

SEMENOVA, VIRA (2023): "Debiased machine learning of set-identified linear models," *Journal of Econometrics*, 235, 1725–1746. [4, 17, 30, 56, 72]

SEMENOVA, VIRA AND VICTOR CHERNOZHUKOV (2021): "Debiased machine learning of conditional average treatment effects and other causal functions," *The Econometrics Journal*, 24, 264–289. [16]

SHAPIRO, ALEXANDER (1990): "On concepts of directional differentiability," *Journal of Optimization Theory and Applications*, 66, 477–487. [30, 64, 69]

VAN DER VAART, AAD W. AND JON A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, Springer New York. [57, 66]

VIVIANO, DAVIDE AND JELENA BRADIC (2023): "Fair policy targeting," *Journal of the American Statistical Association*, 1–14. [5]

APPENDIX A: PROOFS OF MAIN RESULTS

A.1. *Proofs for Section 3*

PROOF OF PROPOSITION 3.1: As shown in the discussion leading to Eq. (8), $\mathcal{E} = \{\mathbb{E}[\mathcal{M}\vartheta(X)] : \vartheta(X) \in \Lambda(X)\} = \mathbf{E}[\mathcal{M}\Lambda(X)]$, where $\Lambda(X) \equiv \text{conv}(\{\boldsymbol{\theta}_0(X), \boldsymbol{\theta}_1(X)\})$, with $\boldsymbol{\theta}_d(X)$ defined in (6) for $d \in \{0,1\}$. Hence, $\mathcal{M}\Lambda(X)$ is a random compact interval (Molchanov and Molinari, 2018, Example 1.11), and $\mathbf{E}[\mathcal{M}\Lambda(X)]$ is its *Aumann expectation* (Molchanov and Molinari, 2018, Def. 3.1), which is well defined because $\Lambda(X)$ is an integrable random convex set owing to $\mathbb{E}[|\theta_d^g(X)|] \leq \mathbb{E}[|\theta_d^g(X)|^2]^{1/2} \leq \mathbb{E}\big[(L_d^g)^2\big]^{1/2} < \sqrt{c_2} < \infty$ for any $d \in \{0,1\}, g \in \{r,b\}$ by Assumption 1. It follows that $h_{\mathcal{E}}(q) = h_{\mathbf{E}[\mathcal{M}\Lambda(X)]}(q) = \mathbb{E}[h_{\mathcal{M}\Lambda(X)}(q)]$ (Molchanov and Molinari, 2018, Theorem 3.11). Observe that

$$h_{\mathcal{M}\Lambda(X)}(q) \equiv \max_{\vartheta(X)\in\Lambda(X)} (\mathcal{M}q)^\intercal \vartheta(X) = \max\{(\mathcal{M}q)^\intercal \boldsymbol{\theta}_0(X), (\mathcal{M}q)^\intercal \boldsymbol{\theta}_1(X)\}, \tag{63}$$

where the last equality in Eq. (63) is well-known in the literature (e.g., Rockafellar, 1997, p. 105). Taking the expectation with respect to $\mathbb{P}(X)$ yields the first line of Eq. (10), which

we can re-write as

$$h_{\mathcal{E}}(q) = \mathbb{E}\Big[(\mathcal{M}q)^{\intercal}\boldsymbol{\theta}_0(X) + (\mathcal{M}q)^{\intercal}\big(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)\big)\mathbb{1}\big\{(\mathcal{M}q)^{\intercal}\big(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)\big) > 0\big\}\Big].$$

The law of iterated expectations yields the expression in the second line of Eq. (10). *Q.E.D.*

PROOF OF PROPOSITION 3.2: The following proof closely follows the argument from Chandrasekhar et al. (2018, Lemma 3). Take any $\|\delta\|_E \to 0$,

$$\begin{aligned}
&\frac{1}{\|\delta\|_E}\Bigg(\mathbb{E}\big[\big(\mathcal{M}(q+\delta)\big)^{\intercal}\boldsymbol{\theta}_0 + k\big(\boldsymbol{\theta}, \mathcal{M}(q+\delta)\big)\mathbb{1}\big\{k\big(\boldsymbol{\theta}, \mathcal{M}(q+\delta)\big) > 0\big\}\big] \\
&\qquad\qquad - \mathbb{E}\big[(\mathcal{M}q)^{\intercal}\boldsymbol{\theta}_0 + k(\boldsymbol{\theta}, \mathcal{M}q)\mathbb{1}\{k(\boldsymbol{\theta}, \mathcal{M}q) > 0\}\big]\Bigg) \\
=&\frac{\delta^{\intercal}}{\|\delta\|_E}\mathbb{E}\big[\mathcal{M}\boldsymbol{\theta}_0 + \mathcal{M}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)\mathbb{1}\{k(\boldsymbol{\theta}, \mathcal{M}q) > 0\}\big] + \frac{1}{\|\delta\|_E}\mathbb{E}[R(q,\delta)],
\end{aligned}$$

where

$$\begin{aligned}
R(q,\delta) \equiv &\big(\mathcal{M}(q+\delta)\big)^{\intercal}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)\mathbb{1}\big\{k(\boldsymbol{\theta}, \mathcal{M}q) \le 0 < k\big(\boldsymbol{\theta}, \mathcal{M}(q+\delta)\big)\big\} \\
&- \big(\mathcal{M}(q+\delta)\big)^{\intercal}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)\mathbb{1}\big\{k(\boldsymbol{\theta}, \mathcal{M}q) > 0 \ge k\big(\boldsymbol{\theta}, \mathcal{M}(q+\delta)\big)\big\} \\
\implies \sup_{q\in\mathbb{S}^1}\mathbb{E}[|R(q,\delta)|] &\lesssim \|\delta\|_E \cdot \big\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\big\|_{L^2(\mathbb{P})} \cdot \sup_{q\in\mathbb{S}^1}\Big\|\mathbb{1}\big\{|k(\boldsymbol{\theta}, \mathcal{M}q)| < |k(\boldsymbol{\theta}, \mathcal{M}\delta)|\big\}\Big\|_{L^2(\mathbb{P})} \\
&\lesssim \|\delta\|_E \cdot \sup_{q\in\mathbb{S}^1}\mathbb{P}\big(|k(\boldsymbol{\theta}, \mathcal{M}q)| < |k(\boldsymbol{\theta}, \mathcal{M}\delta)|\big)^{1/2} \lesssim \|\delta\|_E\big(\|\delta\|_E^{m/2} + \|\delta\|_E^{1/2}\big)^{1/2}
\end{aligned}$$

where the first inequality follows from Hölder's inequality and that, for each X, conditional on the event $|k(\boldsymbol{\theta}, \mathcal{M}q)| < |k(\boldsymbol{\theta}, \mathcal{M}\delta)|$, we have $|\big(\mathcal{M}(q+\delta)\big)^{\intercal}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)| \le 2|\mathcal{M}\delta^{\intercal}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)| \le 2\|\mathcal{M}\delta\|_E\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\|_E \lesssim \|\delta\|_E\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\|_E$; the second inequality follows from $\big\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\big\|_{L^2(\mathbb{P})} < \infty$ by Assumption 1. The last inequality follows from

$$\begin{aligned}
\sup_{q\in\mathbb{S}^1}\mathbb{P}\big(|k(\boldsymbol{\theta}, \mathcal{M}q)| < |k(\boldsymbol{\theta}, \mathcal{M}\delta)|\big) &\le \sup_{q\in\mathbb{S}^1}\mathbb{P}\big(|k(\boldsymbol{\theta}, \mathcal{M}q)| < \|\delta\|_E^{1/2}\big) + \mathbb{P}\big(|k(\boldsymbol{\theta}, \mathcal{M}\delta)| \ge \|\delta\|_E^{1/2}\big) \\
&\lesssim \|\delta\|_E^{m/2} + \|\delta\|_E^{1/2}, \qquad (64)
\end{aligned}$$

where the last line follows because $\sup_{q\in\mathbb{S}^1}\mathbb{P}\big(|k(\boldsymbol{\theta}, \mathcal{M}q)|\big) \lesssim \|\delta\|_E^{m/2}$ by Assumption 2 and $\mathbb{P}\big(|k(\boldsymbol{\theta}, \mathcal{M}\delta)| \ge \|\delta\|_E^{1/2}\big) \le \frac{\|\mathcal{M}\delta\|_E\mathbb{E}[\|\boldsymbol{\theta}\|_E]}{\|\delta\|_E^{1/2}} \lesssim \|\delta\|_E^{1/2}$ by Markov's inequality and

Assumption 1. Hence, $\sup_{q\in\mathbb{S}^1}\frac{1}{\|\delta\|_E}\mathbb{E}[R(q,\delta)] \le \sup_{q\in\mathbb{S}^1}\frac{1}{\|\delta\|_E}\mathbb{E}[|R(q,\delta)|] \lesssim (\|\delta\|_E^{m/2} + \|\delta\|_E^{1/2})^{1/2} \to 0$ and Eq. (12) follows from applying the law of iterated expectations.

Next, the claim that $\nabla_q h_{\mathcal{E}}(q) = \mathcal{S}_{\mathcal{E}}(q)$ follows from Schneider (1993, Corollary 1.7.3). Finally, uniform continuity of $\mathcal{S}_{\mathcal{E}}(q)$ follows from continuity over compact $\mathbb{S}^1$. *Q.E.D.*

PROOF OF PROPOSITION 3.3: By definition of $\mathcal{F}$ and $\mathcal{C}(e^*)$, $e^* \in \mathcal{F}$ if and only if $\mathcal{C}(e^*) \cap \mathcal{E} = \{e^*\}$. Suppose $e^* \in \mathcal{F}$. Then $\text{relint}(\mathcal{C}(e^*)) \cap \text{relint}(\mathcal{E}) = \emptyset$. Since both $\mathcal{C}(e^*)$ and $\mathcal{E}$ are nonempty convex sets, by Schneider (1993, Theorem 1.3.8) $\mathcal{C}(e^*)$ and $\mathcal{E}$ are properly separated. Hence, there exists $q \in \mathbb{S}^1$ and $z \in \mathbb{R}$ such that

$$\forall\, \tilde{e} \in \mathcal{C}(e^*),\ \ \tilde{e}^\intercal q \le z \quad \text{and} \quad \forall\, e \in \mathcal{E},\ \ e^\intercal(-q) \le -z.$$

Since $e^* \in \mathcal{C}(e^*) \cap \mathcal{E}$, we have $e^{*\intercal} q = z$ and $h_{\mathcal{C}(e^*)}(q) = -h_{\mathcal{E}}(-q) = z$. For the other direction, suppose there exists $q \in \mathbb{S}^1$ such that $h_{\mathcal{C}(e^*)}(q) = -h_{\mathcal{E}}(-q)$. By definition, e^* is feasible and $e^* \in \mathcal{C}(e^*) \cap \mathcal{E}$. If there exists $e' \in \mathcal{C}(e^*) \cap \mathcal{E}$ but $e' \neq e^*$, then $q^\intercal e' = q^\intercal e^* = -h_{\mathcal{E}}(-q)$. This means that the support set of $\mathcal{E}$ in direction $-q$ is not a singleton, contradicting Assumption 2, under which $\mathcal{E}$ has a smooth boundary. To obtain the characterization of $\mathcal{F}$ in Eq. (21), for $\mathbb{B}^1 \equiv \{q : \|q\|_E \le 1\}$, note that the optimization problem $\max_{q\in\mathbb{B}^1}(-h_{\mathcal{C}(e^*)}(q) - h_{\mathcal{E}}(-q))$ is dual to the primal problem $\min_{\tilde{e}\in\mathcal{C}(e^*),e\in\mathcal{E}}\|\tilde{e} - e\|_E$, which measures the distance between $\mathcal{C}(e^*)$ and $\mathcal{E}$. When $e^* \in \mathcal{E}$, this distance is zero, and weak duality along with $\mathbb{S}^1 \subset \mathbb{B}^1$ imply $\max_{q\in\mathbb{S}^1}(-h_{\mathcal{C}(e^*)}(q) - h_{\mathcal{E}}(-q)) \le 0$. Hence, when $e^* \notin \mathcal{F}$ we must have $\sup_{q\in\mathbb{S}^1}(-h_{\mathcal{C}(e^*)}(q) - h_{\mathcal{E}}(-q)) < 0$. Since $h_{\mathcal{C}(e^*)}(q) = \sup_{e\in\mathcal{C}(e^*)} q^\intercal e$ is unbounded for any q such that $q_1 + q_2 < 0$, it is without loss of generality to focus on $q \in \tilde{\mathbb{S}}^1$ and to use the criterion $\left[\max_{q\in\tilde{\mathbb{S}}^1}(-h_{\mathcal{C}(e^*)}(q) - h_{\mathcal{E}}(-q))\right]_- = 0$. *Q.E.D.*

A.2. *Proofs for Section 4*

In the proofs that follow, recall

$$\zeta_i(\mathring{\mathcal{M}}q;\boldsymbol{\vartheta}) \equiv (\mathring{\mathcal{M}}q)^\intercal \mathbf{L}_{0_i} + (\mathring{\mathcal{M}}q)^\intercal(\mathbf{L}_{1_i} - \mathbf{L}_{0_i}) \cdot \mathbb{1}\{k(\boldsymbol{\vartheta}(X_i), \mathring{\mathcal{M}}q) > 0\}.$$

defined in Eq. (23), where $k\big(\boldsymbol{\vartheta}(X_i), \mathring{\mathcal{M}}q\big) = q^\intercal \mathring{\mathcal{M}} \Delta\boldsymbol{\vartheta}(X_i)$ is given in Eq. (22) for generic $\Delta\boldsymbol{\vartheta}(X_i) \in \Theta$ and $\mathring{\mathcal{M}}$. For $\widehat{\Delta\boldsymbol{\theta}}$ and $\widehat{\mathcal{M}}$ estimated as per Definition 1, recall

$$\zeta_i(\widehat{\mathcal{M}}q; \widehat{\boldsymbol{\theta}}) = (\widehat{\mathcal{M}}q)^\intercal \mathbf{L}_{0_i} + (\widehat{\mathcal{M}}q)^\intercal (\mathbf{L}_{1_i} - \mathbf{L}_{0_i}) \cdot \mathbb{1}\big\{k\big(\widehat{\boldsymbol{\theta}}(X_i), \widehat{\mathcal{M}}q\big) > 0\big\}$$

for $k\big(\widehat{\boldsymbol{\theta}}(X_i), \widehat{\mathcal{M}}q\big) = q^\intercal \widehat{\mathcal{M}} \widehat{\Delta\boldsymbol{\theta}}(X_i)$ as in Eqs. (25)-(26).

PROOF OF PROPOSITION 4.1: Apply the law of iterated expectations,

$$\begin{aligned}
&\mathbb{E}\big[\zeta_i\big(\mathcal{M}q; \boldsymbol{\theta} + t(\boldsymbol{\vartheta} - \boldsymbol{\theta})\big)\big] - \mathbb{E}\big[\zeta_i(\mathcal{M}q; \boldsymbol{\theta})\big] \\
=&\mathbb{E}\left[\left(\mathbb{1}\Big\{(\mathcal{M}q)^\intercal(\Delta\boldsymbol{\theta} + t(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta})) \geq 0\Big\} - \mathbb{1}\Big\{(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta} \geq 0\Big\}\right)(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta}\right]. \qquad (65)
\end{aligned}$$

In what follows, we show Eq. (65) is bounded in absolute value by $t(t^{m/2} + t^{1/2})^{1/2}$ for m in Assumption 2, uniformly in $q \in \mathbb{S}^1$. It then follows that, for all $q \in \mathbb{S}^1$,

$$\lim_{t \to 0} \frac{1}{t} \left| \big(\mathbb{E}\big[\zeta_i\big(\mathcal{M}q; \boldsymbol{\theta} + t(\boldsymbol{\vartheta} - \boldsymbol{\theta})\big)\big] - \mathbb{E}\big[\zeta_i(\mathcal{M}q; \boldsymbol{\theta})\big]\big) \right| \lesssim \lim_{t \to 0} \frac{1}{t} \cdot t(t^{m/2} + t^{1/2})^{1/2} = 0.$$

The term in Eq. (65) is non-zero if and only if the indicator involving $\Delta\boldsymbol{\theta} + t(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta})$ equals 1 and that involving $\Delta\boldsymbol{\theta}$ equals 0, or vice versa; this happens on a subset of events

$$\Big\{(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta} \leq 0 < (\mathcal{M}q)^\intercal(\Delta\boldsymbol{\theta} + t(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta}))\Big\} \bigcup \Big\{(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta} > 0 \geq (\mathcal{M}q)^\intercal(\Delta\boldsymbol{\theta} + t(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta}))\Big\},$$

which implies the event $\big\{|(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta}| < t|(\mathcal{M}q)^\intercal(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta})|\big\}$. It then follows that

$$\begin{aligned}
|\text{Eq. (65)}| &\leq \mathbb{E}\left[\mathbb{1}\Big\{|(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta}| < t|(\mathcal{M}q)^\intercal(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta})|\Big\}\big|(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta}\big|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\Big\{|(\mathcal{M}q)^\intercal \Delta\boldsymbol{\theta}| < t|(\mathcal{M}q)^\intercal(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta})|\Big\} \cdot t\big|(\mathcal{M}q)^\intercal(\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta})\big|\right] \\
&\lesssim t(t^{m/2} + t^{1/2})^{1/2},
\end{aligned}$$

where the last line follows from Hölder's inequality, Assumption 2, and $\sup_{\Delta\boldsymbol{\vartheta} \in \Theta} \|\Delta\boldsymbol{\vartheta} - \Delta\boldsymbol{\theta}\|_{L^2(\mathbb{P})} < \infty$, using a similar argument to that of Eq. (64). *Q.E.D.*

PROOF OF THEOREM 4.1: **Part 1.** We begin by showing that for any fixed $q \in \mathbb{S}^1$,

$$\sqrt{n}\left(\widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q;\boldsymbol{\theta})\right) = \mathbb{G}_n[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] + o_p(1), \tag{66}$$

where for a generic measurable function $t \in \mathcal{T}$ of random variable O_i, $\mathbb{G}_n[t(O_i)] \equiv n^{-1/2}\sum_{i=1}^n \big(t(O_i) - \mathbb{E}[t(O_i)]\big)$ denotes the empirical process indexed by the function class $\mathcal{T}$. For 2×2 matrices $\check{\mathcal{M}}$ and $\mathring{\mathcal{M}}$, let

$$\boldsymbol{\xi}_{\mathcal{S},i}(\check{\mathcal{M}};\mathring{\mathcal{M}}q;\boldsymbol{\vartheta}) \equiv \check{\mathcal{M}}\mathbf{L}_0 + \check{\mathcal{M}}(\mathbf{L}_1 - \mathbf{L}_0)\mathbb{1}\{k(\boldsymbol{\vartheta}, \mathring{\mathcal{M}}q) > 0\}. \tag{67}$$

To show Eq. (66), fix $q \in \mathbb{S}^1$ and decompose

$$\sqrt{n}\left(\widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q)\right) = \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^\intercal\big(\widehat{\mathcal{M}}\mathcal{M}^{-1}\big)\big(\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathbb{E}[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})]\big)\right)}_{\equiv A}$$
$$+ \underbrace{\sqrt{n}q^\intercal\big(\widehat{\mathcal{M}}\mathcal{M}^{-1} - \boldsymbol{I}\big)\mathbb{E}[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})]}_{\equiv B},$$

where $\boldsymbol{I}$ is the 2×2 identity matrix, and

$$\begin{aligned} A = {} & \mathbb{G}_n[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})] + o_p(1) \\ & + \underbrace{\Bigg\{\sqrt{n}\Bigg(\frac{1}{n}\sum_{i=1}^n (\mathcal{M}q)^\intercal\big(\mathbf{L}_{0_i} + (\mathbf{L}_{1_i} - \mathbf{L}_{0_i})\mathbb{1}\{k(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{M}}q) > 0\}\big) \\ & \qquad - \mathbb{E}\left[(\mathcal{M}q)^\intercal\big(\mathbf{L}_{0_i} + (\mathbf{L}_{1_i} - \mathbf{L}_{0_i})\mathbb{1}\{k(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{M}}q) > 0\}\big)\right]\Bigg) \\ & \qquad - \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \zeta_i(\mathcal{M}q;\boldsymbol{\theta}) - \mathbb{E}[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})]\right)\Bigg\}}_{\equiv R_1} \qquad (68) \\ & + \underbrace{\sqrt{n}\left(\mathbb{E}\left[(\mathcal{M}q)^\intercal\big(\mathbf{L}_{0_i} + (\mathbf{L}_{1_i} - \mathbf{L}_{0_i})\mathbb{1}\{k(\widehat{\boldsymbol{\theta}}, \widehat{\mathcal{M}}q) > 0\}\big)\right] - \mathbb{E}[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})]\right)}_{\equiv R_2}, \qquad (69) \end{aligned}$$

where the equality follows from $\|\widehat{\mathcal{M}}\mathcal{M} - \boldsymbol{I}\|_{\max} = O_p(n^{-1/2})$ under the theorem's assumptions and adding and subtracting terms and R_1 is an empirical process term indexed by

$\Delta\zeta_i\big(q;\widehat{\boldsymbol{\theta}},\widehat{\mathcal{M}}\big) \equiv (\mathcal{M}q)^\intercal(\mathbf{L}_{1_i}-\mathbf{L}_{0_i})\mathbb{1}\big\{k\big(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{M}}q\big)>0\big\} - (\mathcal{M}q)^\intercal(\mathbf{L}_{1_i}-\mathbf{L}_{0_i})\mathbb{1}\big\{k\big(\boldsymbol{\theta},\mathcal{M}q\big)>0\big\}$. Recall that $k(\widehat{\boldsymbol{\theta}},\widehat{\mathcal{M}}q) = q^\intercal\widehat{\mathcal{M}}\widehat{\Delta\boldsymbol{\theta}}$, where $\widehat{\Delta\boldsymbol{\theta}}(X_i) = \widehat{\Delta\boldsymbol{\theta}}_k(X_i)$ for $i \in I_k$. Fixing any $k \in [K]$ and conditional on the I_k^c sample, $\Delta\widehat{\boldsymbol{\theta}}_k$ is non-stochastic and

$$
\begin{aligned}
\mathbb{E}[R_1^2 \mid I_k^c] &\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \mathbb{E}\left[\Delta\zeta_i\big(q;\widehat{\boldsymbol{\theta}},\widehat{\mathcal{M}}\big)^2\right] \\
&= \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \mathbb{E}\left[\big((\mathcal{M}q)^\intercal(\mathbf{L}_{1_i}-\mathbf{L}_{0_i})\big)^2\left|\mathbb{1}\{q^\intercal\widehat{\mathcal{M}}\Delta\boldsymbol{\vartheta}>0\} - \mathbb{1}\{q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}>0\}\right|\right] \\
&\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \mathbb{E}\left[\big((\mathcal{M}q)^\intercal(\mathbf{L}_{1_i}-\mathbf{L}_{0_i})\big)^2\mathbb{1}\big\{|q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}| < |q^\intercal(\widehat{\mathcal{M}}\Delta\boldsymbol{\vartheta}-\mathcal{M}\Delta\boldsymbol{\theta})|\big\}\right] \\
&\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \mathbb{P}\Big(|q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}| < |q^\intercal(\widehat{\mathcal{M}}\Delta\boldsymbol{\vartheta}-\mathcal{M}\Delta\boldsymbol{\theta})|\Big) = o_p(1), \qquad (70)
\end{aligned}
$$

where the first inequality follows from Chernozhukov et al. (2018, Proof of Theorem 3.1 and Lemma 6.1), the third inequality follows by Assumption 1, and the last equality follows from the definition of Θ_n, $\|\widehat{\mathcal{M}}-\mathcal{M}\|_{\max} = O_p(n^{-1/2})$, and a similar argument to that of Eq. (64). Hence $|R_1| = o_p(1)$. In addition, by Proposition 4.1, R_2 encapsulates higher-order Gateaux derivatives and can be bounded by

$$
|R_2| \lesssim \sqrt{n}\underbrace{\mathbb{E}\left[\mathbb{1}\left\{|q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}| < |q^\intercal(\widehat{\mathcal{M}}\widehat{\Delta\boldsymbol{\theta}}-\mathcal{M}\Delta\boldsymbol{\theta})|\right\}|q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}|\right]}_{R_3},
$$

and we show $|R_3| = o_p(n^{-1/2})$ by leveraging the proof technique from Semenova (2023, Lemma 4.1) under Assumption 3. It then follows that $|R_2| = o_p(1)$.

Under Assumption 3, for $g \in \{r,b\}$ and the k-th fold,

$$
\begin{aligned}
&\left|\frac{(\widehat{\Delta\theta}^g)_k(X)}{\widehat{\mu}_g} - \frac{\Delta\theta^g(X)}{\mu_g}\right| \\
\lesssim &\left|\frac{(\widehat{\alpha}^g)_k}{\widehat{\mu}_g} - \frac{\alpha^g}{\mu_g}\right| + \left|\frac{(\widehat{\beta}^g)_k}{\widehat{\mu}_g} - \frac{\beta^g}{\mu_g}\right| + \left|\frac{(\widehat{\eta}^g)_k(X_{[3:d_X]})}{\widehat{\mu}_g} - \frac{\eta^g(X_{[3:d_X]})}{\mu_g}\right| \equiv \delta_k^g(X_{[3:d_X]}),
\end{aligned}
$$

where $\lesssim$ follows from the assumption that (X_1,X_2) has bounded support. Let $\delta_k(X_{[3:d_X]}) \equiv \sum_{g\in\{r,b\}}\delta_k^g(X_{[3:d_X]})$. Denote the distribution of $|q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}(X)|$ conditional on $X_{[3:d_X]}$ by $\mathbb{P}_{\left(|q^\intercal\mathcal{M}\Delta\boldsymbol{\theta}|\,\middle|\,X_{[3:d_X]}\right)}$. Conditional on X and the sample I_k^c, $|q^\intercal(\widehat{\mathcal{M}}\widehat{\Delta\boldsymbol{\theta}}_k - \mathcal{M}\Delta\boldsymbol{\theta})| \leq$

$\|\widehat{\mathcal{M}}\widehat{\Delta\boldsymbol{\theta}}_k - \mathcal{M}\Delta\boldsymbol{\theta}\|_E \lesssim \delta_k(X_{[3:d_X]})$, and

$$|R_3| \lesssim \mathbb{E}_{X_{[3:d_X]}}\left[\int_{-\delta_k(X_{[3:d_X]})}^{\delta_k(X_{[3:d_X]})} \delta_k(X_{[3:d_X]})\mathbb{P}\left(|q^\intercal \mathcal{M}\Delta\boldsymbol{\theta}| \,\middle|\, X_{[3:d_X]}\right)\right] \lesssim \mathbb{E}_{X_{[3:d_X]}}\left[\delta_k(X_{[3:d_X]})^2\right] = o_p(n^{-1/2}),$$

where the second inequality follows from Assumption 3 that $\mathbb{P}_{\left(|q^\intercal \mathcal{M}\Delta\boldsymbol{\theta}| \,\middle|\, X_{[3:d_X]}\right)}$ is bounded, and the last equality follows from $\widehat{\Delta\boldsymbol{\theta}}_k \in \Theta_n$ as $n \to \infty$, $\|\widehat{\mathcal{M}} - \mathcal{M}\|_{\max} = O_p(n^{-1/2})$, and repeated application of triangle inequality. Therefore,

$$A = \mathbb{G}_n[\zeta_i(\mathcal{M}q; \boldsymbol{\theta})] + o_p(1).$$

In addition,

$$B = \sqrt{n}\big((\widehat{\mathcal{M}} - \mathcal{M})q\big)^\intercal \mathcal{M}^{-1}\mathcal{S}_{\mathcal{E}}(q) = \mathbb{G}_n[(\mathcal{M}_i^* q)^\intercal \mathcal{M}^{-1}\mathcal{S}_{\mathcal{E}}(q)] + o_p(1)$$

for $\mathcal{M}_i^* \equiv \operatorname{diag}\left(\frac{\mathbb{1}\{G_i=r\}}{-\mu_r^2}, \frac{\mathbb{1}\{G_i=b\}}{-\mu_b^2}\right)$ by the Delta method. We hence conclude

$$\sqrt{n}\left(\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q; \boldsymbol{\theta})\right) = \mathbb{G}_n[\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta})] + o_p(1)$$

for $\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta}) \equiv \zeta_i(\mathcal{M}q; \boldsymbol{\theta}) + (\mathcal{M}_i^* q)^\intercal \mathcal{M}^{-1}\mathcal{S}_{\mathcal{E}}(q)$.

Part 2. Since the class of functions over the random variables $(\mathbf{L}_{1_i}, \mathbf{L}_{0_i}, X_i)$,

$$\mathcal{Z} \equiv \left\{\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta}) : q \in \mathbb{S}^1\right\},$$

is composed of linear functions (linear in $q \in \mathbb{S}^1$) and their indicators, $\mathcal{Z}$ is VC-subgraph (see, for example, Andrews, 1994). An envelope function of $\mathcal{Z}$ is

$$\sum_{d\in\{0,1\}, g\in\{r,b\}} (|L_{d_i}^g| + C)/c_1. \tag{71}$$

As C and c_1 are constant and $L_{d_i}^g$ is square-integrable, Eq. (71) is square-integrable under $\mathbb{P}$. By van der Vaart and Wellner (1996, Theorem 2.5.2), $\mathcal{Z}$ is $\mathbb{P}$-Donsker, and hence

$$\sqrt{n}\left(\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q)\right) = \mathbb{G}[\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta})] + o_p(1) \quad \text{in } \ell^\infty(\mathbb{S}^1).$$

Part 3. The fact that $Var\left(\mathbb{G}[\zeta_i(q;\boldsymbol{\theta})]\right) > 0$ for each $q \in \mathbb{S}^1$ follows using the variance decomposition formula and the same argument as in Beresteanu and Molinari (2008, proof of Theorem 4.3-(ii), p.808). *Q.E.D.*

A.3. *Proofs for Section 5*

PROOF OF PROPOSITION 5.1: Recalling the definition of the set $Q^*_{\mathbb{T}}(e),\ \mathbb{T} \in \{\mathbb{S}^1, \tilde{\mathbb{S}}^1\}$ in Proposition 5.3, by the same argument as in the proof of Proposition 5.3 below,

$$\left[\sqrt{n}\max_{q\in\mathbb{S}^1}(q^\intercal e - \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}))\right]_+ \xrightarrow{d} \left[\sup_{q\in Q^*_{\mathbb{S}^1}(e)} \mathbb{G}[-\zeta^*_i(\mathcal{M}q;\boldsymbol{\theta})]\right]_+ .$$

Applying the argument in the proof of Proposition 6.2 to the second maximization problem in the definition of $T_n^{\mathcal{F}}$ in Eq. (31), we have

$$\psi^{\mathcal{F}}(e) = \left[\sup_{q\in Q^*_{\mathbb{S}^1}(e)} \mathbb{G}[-\zeta^*_i(\mathcal{M}q;\boldsymbol{\theta})]\right]_+ + \left[\sup_{q\in Q^*_{\tilde{\mathbb{S}}^1}(e)} \mathbb{G}[-\zeta^*_i(-\mathcal{M}q;\boldsymbol{\theta})]\right]_- . \tag{72}$$

When Eq. (14) holds, $\mathcal{E}$ has no kinks or flat faces (by Assumption 2); hence under the null:

$$Q^*_{\mathbb{S}^1}(e) = \arg\max_{q\in\mathbb{S}^1}(q^\intercal e - h_{\mathcal{E}}(q)) = \left\{q^*_{\mathbb{S}^1}(e)\right\},\ Q^*_{\tilde{\mathbb{S}}^1}(e) = \arg\max_{q\in\tilde{\mathbb{S}}^1}\left(-h_{\mathcal{C}(e)}(q) - h_{\mathcal{E}}(-q)\right) \equiv \{q^*_{\tilde{\mathbb{S}}^1}(e)\}.$$

Under the null $e \in \mathcal{F}$, $(q^*_{\mathbb{S}^1}(e))^\intercal e - h_{\mathcal{E}}(q^*_{\mathbb{S}^1}(e)) = 0$ and $e = \mathcal{S}_{\mathcal{E}}(q^*_{\mathbb{S}^1}(e))$; since $e \in \mathcal{F}$, $-(q^*_{\tilde{\mathbb{S}}^1}(e))^\intercal e - h_{\mathcal{E}}(-q^*_{\tilde{\mathbb{S}}^1}(e)) = 0$, so that $e = \mathcal{S}_{\mathcal{E}}(-q^*_{\tilde{\mathbb{S}}^1}(e))$. By Eq. (14) kinks are absent, and hence $q^*_{\mathbb{S}^1}(e) = -q^*_{\tilde{\mathbb{S}}^1}(e)$. We therefore obtain the expression in Eq. (32), as

$$\psi^{\mathcal{F}}(e) = \left[\mathbb{G}[-\zeta^*_i(\mathcal{M}q^*_{\mathbb{S}^1}(e);\boldsymbol{\theta})]\right]_+ + \mathbb{G}\left[-\zeta^*_i(\mathcal{M}q^*_{\mathbb{S}^1}(e);\boldsymbol{\theta})\right]_- = \left|\mathbb{G}\left[\zeta^*_i(\mathcal{M}q^*_{\mathbb{S}^1}(e);\boldsymbol{\theta})\right]\right| .$$

Eq. (35) follows by standard arguments.

We establish Eq. (34) by verifying Condition C.1 in Chernozhukov et al. (2007), under which Eq. (34) follows from their Theorem 3.1-(1). By definition, the parameter space $\mathbb{B}_C$ is a compact (and convex) set. Our criterion function is

$$\mathsf{f}(e) \equiv \left[\max_{q\in\mathbb{S}^1}(q^\intercal e - h_{\mathcal{E}}(q))\right]_+ + \left[\max_{q\in\tilde{\mathbb{S}}^1}(-h_{\mathcal{C}(e)}(q) - h_{\mathcal{E}}(-q))\right]_- ,$$

and the population set is $\mathcal{F} = \{e \in \mathbb{B}_C : \ \mathsf{f}(e) = 0\}$. The criterion function $\mathsf{f}(e)$ is lower-semicontinuous by Berge's Maximum Theorem and composition with a continuous function. The sample criterion function

$$\widehat{\mathsf{f}}(e) \equiv \left[\max_{q \in \mathbb{S}^1}(q^\intercal e - \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}))\right]_+ + \left[\max_{q \in \tilde{\mathbb{S}}^1}(-h_{\mathcal{C}(e)}(q) - \widehat{h}_{\mathcal{E}}(-q;\widehat{\boldsymbol{\theta}}))\right]_-$$

takes values in $\mathbb{R}_+$ and is jointly measurable in the parameter $e \in \mathcal{E}$ and the data by standard arguments. Finally, by Proposition 6.2, $\sup_{e \in \mathbb{B}_C}\left(\mathsf{f}(e) - \widehat{\mathsf{f}}(e)\right)_+ = O_p(1/\sqrt{n})$ and $\sup_{e \in \mathcal{F}} \widehat{\mathsf{f}}(e) = O_p(1/\sqrt{n})$. *Q.E.D.*

PROOF OF PROPOSITION 5.2: Recall that by Proposition 3.2, $\mathcal{S}_{\mathcal{E}}(q) = \mathbb{E}\big[\boldsymbol{\zeta}_{\mathcal{S}}(\mathcal{M}q;\boldsymbol{\theta})\big]$ for $\boldsymbol{\zeta}_{\mathcal{S}}(\mathcal{M}q;\boldsymbol{\theta}) \equiv \mathcal{M}\mathbf{L}_0 + \mathcal{M}(\mathbf{L}_1 - \mathbf{L}_0)\mathbb{1}\{k(\boldsymbol{\theta}, \mathcal{M}q) > 0\}$. Using the notation $\boldsymbol{\xi}_{\mathcal{S},i}(\check{\mathcal{M}};\mathring{\mathcal{M}}q;\vartheta)$ defined in Eq. (67), we have:

$$\begin{aligned}
\left\|\widehat{\mathcal{S}}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) - \mathcal{S}_{\mathcal{E}}(q)\right\| &= \left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_{\mathcal{S},i}(\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathcal{S}_{\mathcal{E}}(q)\right\| \\
\leq \left\|\widehat{\mathcal{M}}\mathcal{M}^{-1}\right\| &\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})\right]\right\| \\
&+ \left\|\left(\widehat{\mathcal{M}}\mathcal{M}^{-1} - \boldsymbol{I}\right)\mathbb{E}[\boldsymbol{\zeta}_{\mathcal{S},i}(\mathcal{M}q;\boldsymbol{\theta})]\right\| \\
\leq &\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})\right]\right\| && \text{(73a)} \\
&+ \left\|\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}})\right]\right)\right. \\
&\qquad \left. - \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})\right]\right)\right\| && \text{(73b)} \\
&+ \left\|\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\widehat{\mathcal{M}}q;\widehat{\boldsymbol{\theta}})\right] - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})\right]\right\| + o_p(1), && \text{(73c)}
\end{aligned}$$

where the $o_p(1)$ term in line (73c) follows as $\|\widehat{\mathcal{M}}\mathcal{M}^{-1} - \boldsymbol{I}\|_{\max} = O_p(n^{-1/2})$. The term in Eq. (73a) equals $o_p(1)$ by the Law of Large Numbers. Using that $\mathbb{E}\left[\mathcal{M}(\mathbf{L}_1 - \mathbf{L}_0)|X\right] =$

$\mathcal{M}(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)) = [-\mathfrak{u}_1^\intercal \mathcal{M}\Delta\boldsymbol{\theta}(X),\ -\mathfrak{u}_2^\intercal \mathcal{M}\Delta\boldsymbol{\theta}(X)]^\intercal$ and omitting dependence on X to shorten notation, we have that the first term of Eq. (73c) is upper bounded by

$$
\begin{aligned}
&\sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \left\| \mathbb{E}\Big[\mathcal{M}(\mathbf{L}_{1_i} - \mathbf{L}_{0_i}) \left(\mathbb{1}\big\{ q^\intercal \widehat{\mathcal{M}} \Delta\boldsymbol{\vartheta} > 0 \big\} - \mathbb{1}\big\{ q^\intercal \mathcal{M} \Delta\boldsymbol{\theta} > 0 \big\} \right) \Big] \right\| \\
&= \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \left\| \mathbb{E}\Big[[-\mathfrak{u}_1^\intercal \mathcal{M}\Delta\boldsymbol{\theta},\ -\mathfrak{u}_2^\intercal \mathcal{M}\Delta\boldsymbol{\theta}]^\intercal \left(\mathbb{1}\big\{ q^\intercal \widehat{\mathcal{M}} \Delta\boldsymbol{\vartheta} > 0 \big\} - \mathbb{1}\big\{ q^\intercal \mathcal{M} \Delta\boldsymbol{\theta} > 0 \big\} \right) \Big] \right\| \\
&\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \max_{v\in\{-\mathfrak{u}_1,-\mathfrak{u}_2\}} \left| \mathbb{E}\Big[v^\intercal \mathcal{M}\Delta\boldsymbol{\theta} \left(\mathbb{1}\big\{ q^\intercal \widehat{\mathcal{M}} \Delta\boldsymbol{\vartheta} > 0 \big\} - \mathbb{1}\big\{ q^\intercal \mathcal{M} \Delta\boldsymbol{\theta} > 0 \big\} \right) \Big] \right| \\
&\leq \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \max_{v\in\{-\mathfrak{u}_1,-\mathfrak{u}_2\}} \mathbb{E}\Big[\big| v^\intercal \mathcal{M}\Delta\boldsymbol{\theta} \big| \mathbb{1}\big\{ \big| q^\intercal \mathcal{M}\Delta\boldsymbol{\theta} \big| < \big| q^\intercal \widehat{\mathcal{M}} \Delta\boldsymbol{\vartheta} - q^\intercal \mathcal{M}\Delta\boldsymbol{\theta} \big| \big\} \Big] \\
&\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \mathbb{E}\left[\mathbb{1}\big\{ \big| q^\intercal \mathcal{M}\Delta\boldsymbol{\theta} \big| < \big| q^\intercal \widehat{\mathcal{M}} \Delta\boldsymbol{\vartheta} - q^\intercal \mathcal{M}\Delta\boldsymbol{\theta} \big| \big\} \right]^{1/2} = o_p(1)
\end{aligned}
$$

where the last inequality follows by Cauchy-Schwartz and Assumption 1, and the last equality follows from Assumption 2, the fact that $\Delta\boldsymbol{\vartheta} \in \Theta_n$, and that under the proposition's assumptions $\|\widehat{\mathcal{M}}\mathcal{M}^{-1} - \boldsymbol{I}\|_{\max} = O_p(n^{-1/2})$; one can then show Eq. (73c) is $o_p(1)$ by the same argument as that used to establish Eq. (70) in the proof of Theorem 4.1.

We conclude by establishing Eq. (39). Using the definition of Hausdorff distance,

$$
\mathbf{d}_H(\widehat{\mathcal{PF}}, \mathcal{PF}) = \max \left\{ \sup_{\hat{e}\in\widehat{\mathcal{PF}}} \inf_{e\in\mathcal{PF}} \|\hat{e} - e\|,\ \sup_{e\in\mathcal{PF}} \inf_{\hat{e}\in\widehat{\mathcal{PF}}} \|\hat{e} - e\| \right\}
$$

Suppose by contradiction that there is $e^* \in \mathcal{PF}$ such that for all $n \geq 1$, $\inf_{\hat{e}\in\widehat{\mathcal{PF}}} \|\hat{e} - e^*\| > c$ for some constant $c > 0$. By definition, there exists $q^* \in \mathbb{Q}$ such that $\mathcal{S}_{\mathcal{E}}(q^*) = e^*$. By Eq. (38), $\|\widehat{\mathcal{S}}_{\mathcal{E}}(q^*; \widehat{\boldsymbol{\theta}}) - e^*\| = o_p(1)$, and by definition $\widehat{\mathcal{S}}_{\mathcal{E}}(q^*; \widehat{\boldsymbol{\theta}}) \in \widehat{\mathcal{PF}}$, yielding a contradiction. The same argument holds by swapping the role of $\widehat{\mathcal{PF}}$ and $\mathcal{PF}$. *Q.E.D.*

PROOF OF PROPOSITION 5.3: Our proof follows arguments in Kaido (2016, Theorem 3.4). We first observe that for $e \in \mathcal{PF}$, $\max_{q\in\mathbb{S}^1}(q^\intercal e - h_{\mathcal{E}}(q)) = 0$ and $\max_{q\in\mathbb{Q}}(q^\intercal e - h_{\mathcal{E}}(q)) = 0$. Hence, for $\mathbb{T} \in \{\mathbb{S}^1, \mathbb{Q}\}$ and $\phi_{e,\mathbb{T}}(f) \equiv \sup_{q\in\mathbb{T}}(q^\intercal e - f(q))$, we can write

$$
\begin{aligned}
T_n^{\mathcal{PF}}(e) = \max \left\{ \sqrt{n} \left(\phi_{e,\mathbb{S}^1}\big(\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})\big) - \phi_{e,\mathbb{S}^1}\big(h_{\mathcal{E}}(q)\big) \right), 0 \right\} \\
- \min \left\{ \sqrt{n} \left(\phi_{e,\mathbb{Q}}\big(\widehat{h}_{\mathcal{E}}(q; \widehat{\boldsymbol{\theta}})\big) - \phi_{e,\mathbb{Q}}\big(h_{\mathcal{E}}(q)\big) \right), 0 \right\}.
\end{aligned}
$$

Kaido (2016, Lemma D.3) establishes that $\phi_{e,\mathbb{T}}$ is Hadamard directionally differentiable at $h_{\mathcal{E}}(\cdot)$ with Hadamard directional derivative equal to $\phi'_{e,\mathbb{T}}(y) = \sup_{q \in Q^*_{\mathbb{T}}(e)} -y(q)$. By Theorem 4.1, the assumptions in Kaido (2016, Lemma D.4) are satisfied, and the result follows by the Continuous Mapping Theorem and as argued in Kaido (2016, proof of Theorem 3.4). Absolute continuity of the limit law in Eq. (42) follows using the same argument as in Beresteanu and Molinari (2008, proof of Theorem 4.3-(ii), p.808). *Q.E.D.*

PROOF OF PROPOSITION 5.4: Take a point e^0 in the set $\{\tilde{e} : \|\tilde{e}\| = 2C \text{ and } \tilde{e}_r + \tilde{e}_b \leq 0\}$. Select $\widehat{e}_{n_1}$ as the metric projection of e^0 onto $\widehat{\mathcal{F}}$ and let e be the metric projection of e^0 onto $\mathcal{F}$. By the consistency result in Proposition 5.1 and by Molchanov (2017, proof of Theorem 1.7.19), $\|\widehat{e}_{n_1} - e\| \xrightarrow{p} 0$ as $n_1 \to \infty$. Hence, given that by Theorem 4.1 the support function estimator converges to the population support function uniformly in $q \in \mathbb{S}^1$, and given that $\widehat{q}^*_{n_1}(\widehat{e}_{n_1})$ is an extremum estimator and all the conditions for its consistency are satisfied, $\|\widehat{q}^*_{n_1}(\widehat{e}_{n_1}) - q^*_{\mathbb{S}^1}(e)\| \xrightarrow{p} 0$ as $n_1 \to \infty$. Next, using the same argument as in the proof of Proposition 5.2 leading to Eqs. (73a)-(73c), and $\boldsymbol{\xi}_{\mathcal{S},i}(\check{\mathcal{M}}; \mathring{\mathcal{M}} q; \boldsymbol{\vartheta})$ defined in Eq. (67),

$$
\begin{aligned}
&\left\| \widehat{\mathcal{S}}_{\mathcal{E}}(\widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1}) - \mathcal{S}_{\mathcal{E}}(q^*_{\mathbb{S}^1}(e)) \right\| \\
&\qquad = \left\| \frac{1}{n_2} \sum_{i=1}^{n_2} \boldsymbol{\zeta}_{\mathcal{S},i}(\widehat{\mathcal{M}} \widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1}) - \mathbb{E}[\boldsymbol{\zeta}_{\mathcal{S},i}(\mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta})] \right\| \\
&\qquad \leq o_p(1) + \left\| \frac{1}{n_2} \sum_{i=1}^{n_2} \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta}) \right] \right\| \qquad \text{(74a)} \\
&\quad + \left\| \left(\frac{1}{n_2} \sum_{i=1}^{n_2} \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \widehat{\mathcal{M}} \widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \widehat{\mathcal{M}} \widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1}) \right] \right) \right. \\
&\qquad \left. - \left(\frac{1}{n_2} \sum_{i=1}^{n_2} \boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta}) - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta}) \right] \right) \right\| \qquad \text{(74b)} \\
&\qquad + \left\| \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \widehat{\mathcal{M}} \widehat{q}^*_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1}) \right] - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \mathcal{M} q^*_{\mathbb{S}^1}(e); \boldsymbol{\theta}) \right] \right\|. \qquad \text{(74c)}
\end{aligned}
$$

By the same argument as in the proof of Proposition 5.2, the terms in Eqs. (74a)-(74b) are $o_p(1)$. We next show that the same holds for the term in Eq. (74c). Take any $\delta > 0$,

$$
\begin{aligned}
&\left\| \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \widehat{\mathcal{M}}\widehat{q}^{*}_{n_1}(\widehat{e}_{n_1}); \widehat{\boldsymbol{\theta}}_{n_1})\right] - \mathbb{E}\left[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M}; \mathcal{M}q^{*}_{\mathbb{S}^1}(e); \boldsymbol{\theta})\right] \right\| \\
&\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \max_{v\in\{-\mathfrak{u}_1,\mathfrak{u}_2\}} \mathbb{E}\Big[\big|v^{\intercal}\mathcal{M}\Delta\boldsymbol{\theta}\big| \mathbb{1}\big\{\big|q^{*}_{\mathbb{S}^1}(e)^{\intercal}\mathcal{M}\Delta\boldsymbol{\theta}\big| < \big|\widehat{q}^{*}_{n_1}(\widehat{e}_{n_1})^{\intercal}\widehat{\mathcal{M}}\Delta\boldsymbol{\vartheta} - q^{*}_{\mathbb{S}^1}(e)^{\intercal}\mathcal{M}\Delta\boldsymbol{\theta}\big|\big\}\Big] \\
&\lesssim \sup_{\Delta\boldsymbol{\vartheta}\in\Theta_n} \mathbb{E}\Big[\mathbb{1}\big\{\big|q^{*}_{\mathbb{S}^1}(e)^{\intercal}\mathcal{M}\Delta\boldsymbol{\theta}\big| < \big|\widehat{q}^{*}_{n_1}(\widehat{e}_{n_1})^{\intercal}\widehat{\mathcal{M}}\Delta\boldsymbol{\vartheta} - q^{*}_{\mathbb{S}^1}(e)^{\intercal}\mathcal{M}\Delta\boldsymbol{\theta}\big|\big\}\Big]^{1/2} \\
&\leq \Big(\mathbb{E}\Big[\mathbb{1}\big\{\big|q^{*}_{\mathbb{S}^1}(e)^{\intercal}\mathcal{M}\Delta\boldsymbol{\theta}\big| < \delta\big\}\Big] + \mathbb{E}\Big[\mathbb{1}\big\{\|\widehat{\mathcal{M}}\widehat{q}^{*}_{n_1}(\widehat{e}_{n_1}) - \mathcal{M}q^{*}_{\mathbb{S}^1}(e)\| > \delta\big\}\Big]\Big)^{1/2} = o_p(1),
\end{aligned}
$$

using Assumption 2, $\|\widehat{\mathcal{M}}\mathcal{M}^{-1} - \boldsymbol{I}\|_{\max} = O_p(n^{-1/2})$, Markov inequality, and that $\delta > 0$ is arbitrary. *Q.E.D.*

A.4. *Proofs for Section 6*

PROOF OF PROPOSITION 6.1: By the same argument as in the proof of Proposition 5.3 and the discussion in Section 5.3,

$$
\liminf_{n\to\infty} \mathbb{P}\big\{(R,B) \in \mathcal{CS}_n(R,B)\big\} \geq 1-\alpha. \tag{75}
$$

To obtain the result in Eq. (52), observe that under the null in Eq. (48), by Eq. (75),

$$
\begin{aligned}
\mathbb{E}\left[\varphi^{\mathtt{skew}}_n\right] &= 1 - \mathbb{P}\left\{ \sup_{(\tilde{R},\tilde{B})\in\mathcal{CS}_n(R,B)} \big((\mathfrak{u}_1-\mathfrak{u}_2)^{\intercal}\tilde{R}\big)\big((\mathfrak{u}_1-\mathfrak{u}_2)^{\intercal}\tilde{B}\big) \geq 0 \right\} \\
&\leq 1 - \mathbb{P}\big\{(R,B) \in \mathcal{CS}_n(R,B)\big\} \leq \alpha.
\end{aligned}
$$

Q.E.D.

PROOF OF PROPOSITION 6.2: For $g \in \{r,b\}$, recall Z^g_i defined in Eq. (54) and note that

$$
\begin{aligned}
\sqrt{n}\left(\widehat{e}^{*}_g - e^{*}_g\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Z^g_i}{\widehat{\mu}_g} - \frac{\mathbb{E}[Z^g_i]}{\mu_g}\right) \\
&= \frac{1}{\mu_g}\mathbb{G}_n[Z^g_i] - \frac{\mathbb{E}[Z^g_i]}{\mu_g^2}\mathbb{G}_n[\mathbb{1}\{G_i = g\}] + o_p(1) = \mathbb{G}_n[Z^{g,*}_i] + o_p(1),
\end{aligned}
$$

where

$$Z_i^{g,*} \equiv \frac{Z_i^g}{\mu_g} - \frac{\mathbb{E}[Z_i^g]}{\mu_g^2}\mathbb{1}\{G_i = g\}. \tag{76}$$

Since $\max_{d\in\{0,1\},g\in\{r,b\}} \mathbb{E}[(L_d^g)^2] < \infty$, $Z_i^{g,*}$ has finite first and second moments. By the Lindeberg–Lévy central limit theorem, we have $\sqrt{n}\left(\widehat{e}_g^* - e_g^*\right) = \mathbb{G}[Z_i^{g,*}] + o_p(1)$, where $\mathbb{G}[Z_i^{g,*}]$ is a mean-zero Gaussian random variable with variance $\mathbb{E}\left[(Z_i^{g,*})^2\right]$.

By Theorem 4.1 and the Cramér–Wold theorem, we have that jointly,

$$\sqrt{n}\begin{bmatrix} \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}) - h_{\mathcal{E}}(q) \\ \widehat{e}_r^* - e_r^* \\ \widehat{e}_b^* - e_b^* \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] \\ \mathbb{G}[Z_i^{r,*}] \\ \mathbb{G}[Z_i^{b,*}] \end{bmatrix} \equiv \mathbb{G}_{he^*} \quad \text{in } \ell^\infty(\mathbb{B}^1), \tag{77}$$

Next, we analyze the two parts of the test statistic in Eq. (55), beginning with the first part. By the same argument as in the proof of Proposition 5.3, under the null that $e^* \in \mathcal{F}$ we have $\max_{q\in\mathbb{S}^1}(q^\intercal e^* - h_{\mathcal{E}}(q)) = 0$. Hence, for $\phi_{\mathbb{S}^1}\left(e, f(\cdot)\right) \equiv \sup_{q\in\mathbb{S}^1}\left(q^\intercal e - f(q)\right)$, we can write $\sqrt{n}\left[\max_{q\in\mathbb{S}^1}(q^\intercal \widehat{e}^* - \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}}))\right]_+ = \max\left\{\sqrt{n}\left(\phi_{\mathbb{S}^1}\left(\widehat{e}^*, \widehat{h}_{\mathcal{E}}(q;\widehat{\boldsymbol{\theta}})\right) - \phi_{\mathbb{S}^1}\left(e^*, h_{\mathcal{E}}(q)\right)\right), 0\right\}$. Kaido (2016, Lemma D.3) shows that $\phi_{\mathbb{S}^1}$ is Hadamard directionally differentiable at $\left(e^*, h_{\mathcal{E}}(\cdot)\right)$ with derivative $\phi'_{\mathbb{S}^1,(e^*,h_{\mathcal{E}}(\cdot))}\left(e, f(\cdot)\right) = \sup_{q\in Q^*_{\mathbb{S}^1}(e^*)}\left(q^\intercal e - f(q)\right)$.

For the second part of the test statistic in Eq. (55), note that for any $s, t \in \mathbb{R}$, $\min\{s, 2t - s\} = (2t - s) - \max\{2(t - s), 0\}$ and $\min\{t, 2s - t\} = t - \max\{2(t - s), 0\}$, so we can plug in $t = e_r^*$ and $s = e_b^*$ to rewrite Eq. (30) as

$$h_{\mathcal{C}^*}(q) = \max\Big\{q_1(e_r^* - \max\{2(e_r^* - e_b^*), 0\}) + q_2 e_b^*,\ \ q_1 e_r^* + q_2\big(2e_r^* - e_b^* - \max\{2(e_r^* - e_b^*), 0\}\big)\Big\}$$

$$= \max\Big\{2q_2\big(e_r^* - e_b^*\big) + (q_1 - q_2)\max\{2(e_r^* - e_b^*), 0\},\ \ 0\Big\} + q_1 e_r^* + q_2 e_b^* - q_1 \max\{2(e_r^* - e_b^*), 0\}.$$

It follows that we can write

$$\begin{aligned} \mathfrak{h}_1\big(h_{\mathcal{E}}(\cdot), e_r^*, e_b^*;\ q\big) &\equiv -h_{\mathcal{C}^*}(q) - h_{\mathcal{E}}(-q) \\ &= -\Bigg(\max\Big\{2q_2\big(e_r^* - e_b^*\big) + (q_1 - q_2)\max\{2(e_r^* - e_b^*), 0\},\ \ 0\Big\} \\ &\qquad + q_1 e_r^* + q_2 e_b^* - q_1 \max\{2(e_r^* - e_b^*), 0\}\Bigg) - h_{\mathcal{E}}(-q). \end{aligned} \tag{78}$$

Hence, the second part of the test statistic in Eq. (55) is the composition of two mappings applied to $\{\widehat{h}_{\mathcal{E}}(\cdot;\widehat{\boldsymbol{\theta}}), \widehat{e}_r^*, \widehat{e}_b^*\}$: $\mathfrak{h}_1(\,\cdot\,;q)$ in Eq. (78) and $\mathfrak{h}_2(\cdot) \equiv \max_q(\cdot)$. Each of these mappings is Hadamard directionally differentiable at $\{h_{\mathcal{E}}(\cdot), e_r^*, e_b^*\}$ tangentially to $\ell^\infty(\tilde{\mathbb{S}}^1) \times \mathbb{R}^2$ (Fang and Santos, 2019, Example 2.1; Cárcamo et al., 2020, Theorem 2.1). By Shapiro (1990, Proposition 3.6),

$$\max_{q\in\tilde{\mathbb{S}}^1}(-h_{\mathcal{C}^*}(q) - h_{\mathcal{E}}(-q)) = \mathfrak{h}_2 \circ \mathfrak{h}_1\big(h_{\mathcal{E}}(\cdot), e_r^*, e_b^*; q\big) \equiv \mathfrak{h}\big(h_{\mathcal{E}}\big(\cdot\big), e_r^*, e_b^*\big) \equiv \mathfrak{h} \tag{79}$$

is directionally differentiable at $(h_{\mathcal{E}}(\cdot), e_r^*, e_b^*)$ tangentially to $\ell^\infty(\tilde{\mathbb{S}}^1) \times \mathbb{R}^2$, with

$$\mathfrak{h}'(\cdot) = \mathfrak{h}'_{2,\mathfrak{h}_1^*(q)} \circ \mathfrak{h}'_{1,s^*}(\,\cdot\,;q), \tag{80}$$

where $\mathfrak{h}_1^*(q) \equiv \mathfrak{h}_1\big(h_{\mathcal{E}}(\cdot), e_r^*, e_b^*\,; q\big)$ and $s^* \equiv (q_1 - q_2)\max\{2(e_r^* - e_b^*), 0\} + 2q_2\big(e_r^* - e_b^*\big)$; for any $s_r, s_b, s \in \mathbb{R}$, $s_h \in \ell^\infty(\tilde{\mathbb{S}}^1)$ (the space of bounded functions over the compact set $\tilde{\mathbb{S}}^1$) and continuous $f \in \ell^\infty(\tilde{\mathbb{S}}^1)$,

$$\begin{aligned}\mathfrak{h}'_{1,s^*}\big(s_h(\cdot), s_r, s_b\,; q\big) = &-\phi'_{1,s^*}(2q_2[s_r - s_b] + 2(q_1 - q_2)\phi'_{1,e_r^*-e_b^*}[s_r - s_b]) \\ &- q_1 s_r - q_2 s_b + 2q_1\phi'_{1,e_r^*-e_b^*}[s_r - s_b] - s_h(-q),\end{aligned} \tag{81}$$

$$\mathfrak{h}'_{2,\mathfrak{h}_1^*(q)}(f) = \max_{\left\{q\in\tilde{\mathbb{S}}^1:\mathfrak{h}_1^*(q)=\mathfrak{h}\right\}} f(q). \tag{82}$$

In Eq. (81), for any $t \in \mathbb{R}$

$$\phi'_{1,s^*}(t) \equiv \begin{cases} t, & \text{if } s^* > 0 \\ \max\{t, 0\}, & \text{if } s^* = 0 \\ 0, & \text{if } s^* < 0 \end{cases} \tag{83}$$

is the Hadamard directional derivative of $\phi_1(s) \equiv \max\{s, 0\}$ at s^* (Fang and Santos, 2019, Example 2.1). Eq. (82) results from Cárcamo et al. (2020, Corollary 2.3), as $\mathfrak{h}_1(\cdot\,;q)$ is a continuous function over compact support. By Shapiro (1990, Proposition 3.6), T_n^{LDA} is Hadamard directionally differentiable at $(h_{\mathcal{E}}(\cdot), e_r^*, e_b^*)$ tangentially to $\ell^\infty(\tilde{\mathbb{S}}^1) \times \mathbb{R}^2$, with directional derivative given by the sum of the two directional derivatives derived above. By Kaido (2016, Theorem 3.4), Fang and Santos (2019, Theorem 2.1), and Cárcamo et al. (2020, Theorem 2.2), for $\mathbb{G}[\mathbf{Z}_i^*] \equiv \left[\mathbb{G}[Z_i^{r,*}] \quad \mathbb{G}[Z_i^{b,*}]\right]^\intercal$, ψ^{LDA} takes the form

$$\psi^{\text{LDA}} = \left[\sup_{q \in Q^*_{\mathbb{S}^1}(e^*)} q^\intercal \mathbb{G}[\mathbf{Z}_i^*] - \mathbb{G}[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] \right]_+ + \left[\mathfrak{h}'_{2,\mathfrak{h}_1^*(q)} \left\{ \mathfrak{h}'_{1,s^*}\big(\mathbb{G}[\zeta_i^*(-\mathcal{M}(\,\cdot\,);\boldsymbol{\theta})], \mathbb{G}[Z_i^{r,*}], \mathbb{G}[Z_i^{b,*}]\,; q\big) \right\} \right]_- . \quad (84)$$

If $c^{\text{LDA}}_{1-\alpha+\varsigma} + \varsigma$ is a continuity point of the distribution of ψ^{LDA},

$$\lim_{n\to\infty} \mathbb{P}\big(T_n^{\text{LDA}} > c^{\text{LDA}}_{1-\alpha+\varsigma} + \varsigma\big) = \mathbb{P}\big(\psi^{\text{LDA}} > c^{\text{LDA}}_{1-\alpha+\varsigma} + \varsigma\big) \leq \alpha.$$

For ς small enough, if $c^{\text{LDA}}_{1-\alpha+\varsigma} + \varsigma$ is a discontinuity point of ψ^{LDA}, then ψ^{LDA} is continuous at $c^{\text{LDA}}_{1-\alpha}$, and since $\mathbb{P}(T_n^{\text{LDA}} > c^{\text{LDA}}_{1-\alpha+\varsigma} + \varsigma) \leq \mathbb{P}(T_n^{\text{LDA}} > c^{\text{LDA}}_{1-\alpha})$, the result follows.

If Eq. (14) is satisfied, $\mathcal{E}$ has no kinks or flat faces (by Assumption 2). Under the null, since $e^* \in \mathcal{F}$, $\left\{ q \in \tilde{\mathbb{S}}^1 : \mathfrak{h}_1^*(q) = \mathfrak{h} \right\} = \arg\max_{q\in\tilde{\mathbb{S}}^1} \left(-h_{\mathcal{C}^*}(q) - h_{\mathcal{E}}(-q)\right) = \{q^*_{\tilde{\mathbb{S}}^1}\}$ and $-(q^*_{\tilde{\mathbb{S}}^1})^\intercal e^* - h_{\mathcal{E}}(-q^*_{\tilde{\mathbb{S}}^1}) = 0$, so that $e^* = \mathcal{S}_{\mathcal{E}}(-q^*_{\tilde{\mathbb{S}}^1})$. It then follows that under Eq. (14),

$$\psi^{\text{LDA}} = \left[q^{*\intercal}_{\mathbb{S}^1} \mathbb{G}[\mathbf{Z}_i^*] - \mathbb{G}[\zeta_i^*(\mathcal{M}q^*_{\mathbb{S}^1};\boldsymbol{\theta})] \right]_+ + \left[\mathfrak{h}'_{1,s^*}\big(\mathbb{G}[\zeta_i^*(-\mathcal{M}(\cdot);\boldsymbol{\theta})], \mathbb{G}[Z_i^{r,*}], \mathbb{G}[Z_i^{b,*}]\,; q^*_{\mathbb{S}^1}\big) \right]_- . \quad (85)$$

Q.E.D.

PROOF OF PROPOSITION 6.3: We verify the assumptions in Theorem 3.2 of Fang and Santos (2019), from which it follows that Proposition 6.3 holds and that $\widehat{c}_\beta = c_\beta + o_p(1)$ (Fang and Santos, 2019, Online Appendix, Eq. (S.13), p. 4) when c_β is a point at which the cdf of $\phi'_{he^*}(\mathbb{G}_{he^*})$ is continuous and increasing.

Assumptions 1, 3(i), 3(iii), and 3(iv) of Fang and Santos (2019) hold by construction; their Assumption 2 holds by Eq. (77), and Assumption 4 holds by Lemma S.3.8 of Fang and Santos (2019, Online Appendix). Lastly, to show Assumption 3(ii) holds, let $\mathcal{O}_n \equiv \{(Y_i, G_i, X_i)\}_{i=1}^n$. In the next paragraph, we establish that

$$\sup_{f\in\mathcal{BL}_1} \left| \mathbb{E}\left[f\left(\sqrt{n}\{\widehat{he^*}(\widehat{\boldsymbol{\theta}}) - \widehat{he^*}(\widehat{\boldsymbol{\theta}})\} \right) \middle| \mathcal{O}_n \right] - \mathbb{E}[f(\mathbb{G}_{he^*})] \right|$$

$$= \sup_{f \in \mathcal{BL}_1} \left| \mathbb{E}\left[f\left(\begin{bmatrix} \mathbb{G}_n[(W_i - 1)\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] \\ \mathbb{G}_n[(W_i - 1)Z_i^{r,*}] \\ \mathbb{G}_n[(W_i - 1)Z_i^{b,*}] \end{bmatrix} \right) \middle| \mathcal{O}_n \right] - \mathbb{E}[f(\mathbb{G}_{he^*})] \right| + o_p(1), \qquad (86)$$

where $\mathbb{G}_{he^*}$ is defined in Eq. (77), and by Theorem 3.6.13 of van der Vaart and Wellner (1996), Eq. (86) $= o_p(1)$, and therefore Assumption 3(ii) of Fang and Santos (2019) holds.

To show that the equality in Eq. (86) holds, we let $\widetilde{\mu}_g^W \equiv \overline{W}\widetilde{\mu}_g$, $\widetilde{\mathcal{M}}^W \equiv \mathrm{diag}(1/\widetilde{\mu}_r^W, 1/\widetilde{\mu}_b^W)$. Recall $\boldsymbol{\xi}_{\mathcal{S},i}(\check{\mathcal{M}};\mathring{\mathcal{M}}q;\widehat{\boldsymbol{\theta}})$ defined in Eq. (67). Decompose the bootstrapped process by

$$\begin{aligned}
&\sqrt{n}\{\widetilde{he^*}(\widehat{\boldsymbol{\theta}}) - \widehat{he^*}(\widehat{\boldsymbol{\theta}})\} = \sqrt{n}\{\widetilde{he^*}(\widehat{\boldsymbol{\theta}}) - he^*\} - \sqrt{n}\{\widehat{he^*}(\widehat{\boldsymbol{\theta}}) - he^*\} \\
&= \sqrt{n} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n W_i q^\intercal \boldsymbol{\xi}_{\mathcal{S},i}(\widetilde{\mathcal{M}}^W;\widetilde{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathbb{E}[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})] \\ \frac{1}{n}\sum_{i=1}^n W_i \frac{Z_i^r}{\widetilde{\mu}_r^W} - \frac{\mathbb{E}[Z_i^r]}{\mu_r} \\ \frac{1}{n}\sum_{i=1}^n W_i \frac{Z_i^b}{\widetilde{\mu}_b^W} - \frac{\mathbb{E}[Z_i^b]}{\mu_b} \end{bmatrix} - \begin{bmatrix} \mathbb{G}_n[\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] \\ \mathbb{G}_n[Z_i^{r,*}] \\ \mathbb{G}_n[Z_i^{b,*}] \end{bmatrix} + o_p(1),
\end{aligned}$$

where the second equality follows from Theorem 4.1 and the proof of Proposition 6.2. Next, observe that

$$\begin{aligned}
&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n W_i q^\intercal \boldsymbol{\xi}_{\mathcal{S},i}(\widetilde{\mathcal{M}}^W;\widetilde{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathbb{E}[\zeta_i(\mathcal{M}q;\boldsymbol{\theta})] \right) \\
&= \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^\intercal(\widetilde{\mathcal{M}}^W\mathcal{M}^{-1})\big(W_i\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\widetilde{\mathcal{M}}q;\widehat{\boldsymbol{\theta}}) - \mathbb{E}[W_i\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})]\big) \right)}_{\equiv \widetilde{A}} \\
&\qquad + \underbrace{\sqrt{n}q^\intercal\left(\widetilde{\mathcal{M}}^W\mathcal{M}^{-1} - \boldsymbol{I}\right)\mathbb{E}[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})]}_{\equiv \widetilde{B}} \\
&= \mathbb{G}_n[W_i\zeta_i^*(\mathcal{M}q;\boldsymbol{\theta})] + o_p(1),
\end{aligned}$$

where $\mathbb{E}[W_i\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})] = \mathbb{E}[\boldsymbol{\xi}_{\mathcal{S},i}(\mathcal{M};\mathcal{M}q;\boldsymbol{\theta})]$ by independence of W_i and $\mathbb{E}[W_i] = 1$, $\widetilde{A}$ (resp., $\widetilde{B}$) is the bootstrapped analogue of A (resp., B) in the proof of Theorem 4.1, and the last equality follows from a similar argument used in the proof of Theorem 4.1.

In addition, for $g \in \{r, b\}$, by the Delta method,

$$\begin{aligned}
&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} W_i \frac{Z_i^g}{\widetilde{\mu}_g^W} - \frac{\mathbb{E}[Z_i^g]}{\mu_g}\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} W_i \frac{Z_i^g}{\mu_g} - \frac{\mathbb{E}[Z_i^g]}{\mu_g}\right) + \sqrt{n}\left(\frac{1}{\widetilde{\mu}_g^W} - \frac{1}{\mu_g}\right)\left(\frac{1}{n}\sum_{i=1}^{n} W_i Z_i^g\right) \\
&= \mathbb{G}_n[W_i Z_i]\frac{1}{\mu_g} - \mathbb{G}_n[W_i \mathbb{1}\{G_i = g\}]\frac{\mathbb{E}[Z_i^g]}{\mu_g^2} + o_p(1) = \mathbb{G}_n[W_i Z_i^{g,*}] + o_p(1).
\end{aligned}$$

Therefore,

$$\sqrt{n}\big\{\widehat{he^*}(\widehat{\boldsymbol{\theta}}) - \widehat{he^*}(\widehat{\boldsymbol{\theta}})\big\} = \begin{bmatrix} \mathbb{G}_n[(W_i - 1)\zeta_i^*(\mathcal{M}q; \boldsymbol{\theta})] \\ \mathbb{G}_n[(W_i - 1)Z_i^{r,*}] \\ \mathbb{G}_n[(W_i - 1)Z_i^{b,*}] \end{bmatrix} + o_p(1),$$

yielding Eq. (86).

Q.E.D.

A.5. *Proofs for Section 7*

PROOF OF PROPOSITION 7.1: Consider the null hypothesis $\mathsf{H}_0 : \rho(e^*, F) = \delta$ against $\mathsf{H}_A : \rho(e^*, F) \neq \delta$ for some $\delta > 0$, and view our confidence interval as the result of inverting this hypothesis test. Let $\varphi_n^{\texttt{dist}} \equiv \mathbb{1}\left\{\inf_{\rho(\tilde{e},\tilde{F}) \in \mathcal{CS}_n^{\rho(e^*,F)}} \left|\rho(\tilde{e}, \tilde{F}) - \delta\right| > 0\right\}$ and partition the parameter space of the location of $\mathcal{E}$ relative to $\mathcal{H}_{45}$ such that

$$\varphi_n^{\texttt{dist}} = \varphi_n^{\texttt{dist}} \mathbb{1}\{\mathcal{E} \cap \mathcal{H}_{45}^- = \emptyset\} \tag{87}$$

$$+ \varphi_n^{\texttt{dist}} \mathbb{1}\{\mathcal{E} \cap \mathcal{H}_{45}^+ = \emptyset\} \tag{88}$$

$$+ \varphi_n^{\texttt{dist}} \mathbb{1}\{\mathcal{E} \cap \mathcal{H}_{45}^+ \neq \emptyset, \mathcal{E} \cap \mathcal{H}_{45}^- \neq \emptyset\}. \tag{89}$$

Note that whenever $\mathcal{CS}_n^+(e^*, F) \neq \emptyset$,

$$\inf_{\rho(\tilde{e},\tilde{F}) \in \mathcal{CS}_n^{\rho(e^*,F)}} |\rho(\tilde{e}, \tilde{F}) - \delta| \leq \inf_{(\tilde{e},\tilde{F}) \in \mathcal{CS}_n^+(e^*,F)} |\rho(\tilde{e}, \tilde{F}) - \delta|,$$

$$\Rightarrow \mathbb{E}[\varphi_n^{\texttt{dist}}] \leq \mathbb{E}\left[\mathbb{1}\left\{\inf_{(\tilde{e},\tilde{F})\in\mathcal{CS}_n^+(e^*,F)} |\rho(\tilde{e},\tilde{F}) - \delta| > 0\right\}\right], \quad (90)$$

and similarly when $(\tilde{e},\tilde{F}) \in \mathcal{CS}_n^+(e^*,F)$ is replaced with $(\tilde{e},\tilde{F}) \in \mathcal{CS}_n^-(e^*,F)$ or $\rho(\tilde{e},\tilde{F}) \in \mathcal{CS}_n^{45}(\rho(e^*,F))$ (and under the case where these sets are non-empty).

When $\mathcal{E} \cap \mathcal{H}_{45}^- = \emptyset$, we have $F \in \mathcal{H}_{45}^+ \cup \mathcal{H}_{45}$ and $\lim_{n\to\infty} \mathbb{P}\big((e^*,F) \in \mathcal{CS}_n^+(e^*,F)\big) \geq 1-\alpha$ by a similar argument to the proof of Proposition 6.1. It then follows that under the null $\rho(e^*,F) = \delta$, $\mathbb{E}[(87)] \leq \alpha \mathbb{1}\{\mathcal{E} \cap \mathcal{H}_{45}^- = \emptyset\}$ as $n \to \infty$, using Eq. (90) and the fact that $\mathbb{P}\left(\inf_{(\tilde{e},\tilde{F})\in\mathcal{CS}_n^+(e^*,F)} \left|\rho(\tilde{e},\tilde{F}) - \delta\right| > 0\right) \leq 1 - \mathbb{P}\big((e^*,F) \in \mathcal{CS}_n^+(e^*,F)\big)$. Similarly, $\mathbb{E}[(88)] \leq \alpha \mathbb{1}\{\mathcal{E} \cap \mathcal{H}_{45}^+ = \emptyset\}$ as $n \to \infty$.

We complete the proof by showing $\mathbb{E}[(89)] \leq \alpha \mathbb{1}\{\mathcal{E} \cap \mathcal{H}_{45}^+ \neq \emptyset, \mathcal{E} \cap \mathcal{H}_{45}^- \neq \emptyset\}$ if $\rho(e^*,F) = \delta$. In the event $\mathcal{E} \cap \mathcal{H}_{45}^+ \neq \emptyset$ and $\mathcal{E} \cap \mathcal{H}_{45}^- \neq \emptyset$, $F \in \mathcal{H}_{45}$ and—since $\mathcal{E}$ is *not* tangent to $\mathcal{H}_{45}$ in this case—the direction in which F is the support set of $\mathcal{E}$ does *not* live in the span $\{c[1 \ -1]^\intercal : c \in \mathbb{R}\}$. Hence, $c^* \equiv \arg\inf_{c\in\mathbb{R}} h_{\mathcal{E}}(\mathfrak{u}_1(c))$ for $\mathfrak{u}_1(c) \equiv \mathfrak{u}_1 - c[1 \ -1]^\intercal$ is bounded, since otherwise it would require $\mathcal{E}$ to be tangent to $\mathcal{H}_{45}$. In addition, as explained in Footnote 3, $h_{\mathcal{E}}(\mathfrak{u}_1(c))$ is bounded from below when $F \in \mathcal{H}_{45}$. We can therefore restrict attention to $c \in [-c_3, c_3] \equiv \mathcal{C}_3 \subset \mathbb{R}$ for some bounded constant $c_3 > 0$ in the characterization of F in Eq. (19) so that $h_{\mathcal{E}}(\mathfrak{u}_1(\cdot))$ is a bounded function over compact support $\mathcal{C}_3$. Under the maintained assumptions, Eq. (77) holds and implies that

$$\sqrt{n}\begin{bmatrix} \widehat{h}_{\mathcal{E}}\big(\mathfrak{u}_1(c);\widehat{\boldsymbol{\theta}}\big) - h_{\mathcal{E}}\big(\mathfrak{u}_1(c)\big) \\ \widehat{e}_r^* - e_r^* \\ \widehat{e}_b^* - e_b^* \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \|\mathfrak{u}_1(c)\|_E \mathbb{G}[\zeta_i^*(\mathcal{M}\mathfrak{u}_1(c)/\|\mathfrak{u}_1(c)\|_E;\boldsymbol{\theta})] \\ \mathbb{G}[Z_i^{r,*}] \\ \mathbb{G}[Z_i^{b,*}] \end{bmatrix} \text{ in } \ell^\infty(\mathcal{C}_3),$$

where we use the property of support functions that for any constant $\tilde{c} > 0$, $h_{\mathcal{E}}(\tilde{c} \cdot q) = \tilde{c} \cdot h_{\mathcal{E}}(q)$ (see, Schneider, 1993, p.45). Recall $F_{45} \equiv \mathfrak{u} \cdot h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)$ and let $\widehat{F}_{45}$ be its estimator with $h_{\mathcal{E}}(\cdot)$ replaced by $\widehat{h}_{\mathcal{E}}(\cdot;\widehat{\boldsymbol{\theta}})$ in the expression of $h_{\tilde{\mathcal{E}}}(\cdot)$ given in Eq. (19). Observe that $\rho(e^*,F_{45})$ is a composition of two Hadamard directionally differentiable functions: let $\mathfrak{h}_4 : \ell^\infty(\mathcal{C}_3) \to \mathbb{R}$ be the $\inf_{c\in\mathcal{C}_3}(\cdot)$ function,

$$\rho(e^*,F_{45}) = \rho\Big(e^*,\ \mathfrak{u} \cdot \mathfrak{h}_4\big\{h_{\mathcal{E}}\big(\mathfrak{u}_1(c)\big)\big\}\Big),$$

where by Cárcamo et al. (2020, Corollary 2.3), $\mathfrak{h}_4$ is directionally differentiable at $h_{\mathcal{E}}\big(\mathfrak{u}_1(\cdot)\big)$ tangentially to $\ell^\infty(\mathcal{C}_3)$, with

$$\mathfrak{h}'_{4,h_{\mathcal{E}}(\mathfrak{u}_1(\cdot))}(f) = \inf_{\left\{c\in\mathcal{C}_3 : h_{\mathcal{E}}\left(\mathfrak{u}_1(c)\right)=h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)\right\}} f(c), \quad \text{for continuous } f \in \ell^\infty(\mathcal{C}_3).$$

Denote $\rho'_{(e^*,F_{45})} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ the directional derivative of ρ at (e^*, F_{45}). Then by Shapiro (1990, Proposition 3.6), Fang and Santos (2019, Theorem 2.1), and Cárcamo et al. (2020, Theorem 2.2),

$$\sqrt{n}\left[\rho(\widehat{e}^*, \widehat{F}_{45}) - \rho(e^*, F_{45})\right] \xrightarrow{d} \psi^{\rho(e^*,F_{45})},$$

where, for $\mathbb{G}[\mathbf{Z}_i^*] \equiv \left[\mathbb{G}[Z_i^{r,*}] \quad \mathbb{G}[Z_i^{b,*}]\right]^\intercal$,

$$\psi^{\rho(e^*,F_{45})} \equiv \rho'_{(e^*,F_{45})}\left(\mathbb{G}[\mathbf{Z}_i^*], \mathfrak{u}\cdot \mathfrak{h}'_{4,h_{\mathcal{E}}(\mathfrak{u}_1(\cdot))}\big(\|\mathfrak{u}_1(c)\|_E \mathbb{G}[\zeta_i^*(\mathcal{M}\mathfrak{u}_1(c)/\|\mathfrak{u}_1(c)\|_E;\boldsymbol{\theta})]\big)\right),$$

so by the continuous mapping theorem,

$$\psi^{45} = \left|\psi^{\rho(e^*,F_{45})}\right|, \tag{91}$$

Next, if Eq. (14) holds and $\mathcal{E}$ has no kinks, we show that the set $\left\{c \in \mathcal{C}_3 : h_{\mathcal{E}}\big(\mathfrak{u}_1(c)\big) = h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)\right\}$ is a singleton and the expression of $\psi^{\rho(e^*,F_{45})}$ simplifies. Recall $c^* \equiv \arg\inf_{c\in\mathcal{C}_3} h_{\mathcal{E}}\big(\mathfrak{u}_1(c)\big)$. By contradiction, assume there exists $\tilde{c} \neq c^*$ and $h_{\mathcal{E}}\big(\mathfrak{u}_1(\tilde{c})\big) = h_{\mathcal{E}}\big(\mathfrak{u}_1(c^*)\big) = h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)$. This implies that the two linear equations $(-1-\tilde{c})e_r + \tilde{c}e_b = h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)$ and $(-1-c^*)e_r + c^*e_b = h_{\tilde{\mathcal{E}}}(\mathfrak{u}_1)$ intersect at some point $(e_r^\star, e_b^\star)$. Replacing this value in the equations, we find $\tilde{c}(e_b^\star - e_r^\star) = c^*(e_b^\star - e_r^\star)$. If $c^* = 0$, for $\tilde{c}$ not to equal c^* it must be the case that $\tilde{c} \neq 0$, in which case $e_b^\star = e_r^\star$, implying that $(e_r^\star, e_b^\star) = R = F$ and in turn by Eq. (14) it must be the case that $\tilde{c} = c^*$. Similarly, if $c^* \neq 0$, then either $e_b^\star = e_r^\star$, which implies $(e_r^\star, e_b^\star) = F$ and hence $\tilde{c} = c^*$, or $\tilde{c}/c^* = 1$ and the claim follows. Let $\tilde{\mathfrak{u}}_1(c^*) \equiv \frac{\mathfrak{u}_1(c^*)}{\|\mathfrak{u}_1(c^*)\|_E}$, we get:

$$\psi^{45} = \left|\rho'_{(e^*,F_{45})}\left(\mathbb{G}[\mathbf{Z}_i^*], \mathfrak{u}\cdot\|\mathfrak{u}_1(c^*)\|_E \cdot \mathbb{G}\left[\zeta_i^*\left(\mathcal{M}\tilde{\mathfrak{u}}_1(c^*);\boldsymbol{\theta}\right)\right]\right)\right|. \tag{92}$$

If $F \in \mathcal{H}_{45}$, $\lim_{n\to\infty} \mathbb{P}\left(\mathcal{CS}_n^{45}(\rho(e^*, F)) = \emptyset\right) = 0$. Under the null $\rho(e^*, F) = \delta$, and using again Eq. (90), if $c_{1-\alpha+\varsigma}^{\rho(\tilde{e},\tilde{F})} + \varsigma$ is a continuity point of the distribution of $\psi^{\rho(\tilde{e},\tilde{F})}$,

$$\lim_{n\to\infty} \mathbb{E}[\varphi_n^{\texttt{dist}}] \leq \lim_{n\to\infty} \mathbb{P}\big(T_n^{\rho(\tilde{e},\tilde{F})} > c_{1-\alpha+\varsigma}^{\rho(\tilde{e},\tilde{F})} + \varsigma\big) \leq \alpha$$

and if it is a discontinuity point for an infinitesimal ς, then $c_{1-\alpha}^{\rho(\tilde{e},\tilde{F})}$ is a continuity point and

$$\lim_{n\to\infty} \mathbb{E}[\varphi_n^{\texttt{dist}}] \leq \lim_{n\to\infty} \mathbb{P}\big(T_n^{\rho(\tilde{e},\tilde{F})} > c_{1-\alpha}^{\rho(\tilde{e},\tilde{F})}\big) = \alpha.$$

Therefore, under the null $\rho(e^*, F) = \delta$, we conclude

$$\lim_{n\to\infty} \mathbb{E}[\varphi_n^{\texttt{dist}}] \leq \alpha.$$

Test inversion yields the coverage result. *Q.E.D.*

APPENDIX B: AUXILIARY RESULTS

B.1. *Threshold Rules*

In the paper, we allow $\mathcal{A}(\mathcal{X})$, the set of all algorithms that map from the input space $\mathcal{X}$ to $[0, 1]$, to be completely unrestricted. This includes randomized rules where the event $D = 1$ occurs with probability $a(X)$. Here instead we consider *threshold rules*. Let $\mathcal{A}^{\texttt{th}}(\mathcal{X})$ denote a space of algorithms $\{a : \mathcal{X} \mapsto \mathbb{R}\}$ such that $a \in \mathcal{A}^{\texttt{th}}(\mathcal{X})$ induces the decision rule

$$D_a = \mathbb{1}\{a(X) \geq 0\}.$$

One could alternatively pick a constant κ a priori and use a decision rule of the form $\mathbb{1}\{a(X) \geq \kappa\}$, but to ease notation we absorb the threshold κ in a. We maintain the following richness assumption on $\mathcal{A}^{\texttt{th}}(\mathcal{X})$:

ASSUMPTION B.1: *The set of algorithms $\mathcal{A}^{th}(\mathcal{X})$ is (i) convex; and (ii) sufficiently rich, in the sense that, X-a.s.,*

$$\exists a' \in \mathcal{A}^{th}(\mathcal{X}) : \ \mathbb{1}\{a'(X) \geq 0\} = 1,$$
$$\exists a'' \in \mathcal{A}^{th}(\mathcal{X}) : \ \mathbb{1}\{a''(X) \geq 0\} = 0.$$

REMARK B.1—Linear Threshold Rules: Assumption B.1 is satisfied by linear threshold rules with $\mathcal{A}^{\texttt{th}}(\mathcal{X}) = \{[1;X]^\intercal\beta : \beta \in \mathbb{R}^{d_X+1}\}$ and $D_\beta = \mathbb{1}\{[1;X]^\intercal\beta \geq 0\}$.

When using threshold rules, similar to Eq. (5), the group risks can be expressed as

$$\begin{aligned} e_g(D_a) &\equiv \frac{1}{\mu_g}\mathbb{E}_X\left[D_a\theta_1^g(X) + (1-D_a)\theta_0^g(X)\right] \\ &= \frac{1}{\mu_g}\mathbb{E}\left[L_0^g + (L_1^g - L_0^g)\mathbb{1}\{a(X)\geq 0\}\right]. \end{aligned} \tag{93}$$

Compare Eq. (93) with Eq. (5): it follows immediately that threshold rules given by $D_{k(\boldsymbol{\theta}(X),\mathcal{M}q)} = \mathbb{1}\{k(\boldsymbol{\theta}(X),\mathcal{M}q)\geq 0\}$ yield the extreme points of the set $\mathcal{E}$ in Eq. (8). As we show next, under Assumption B.1, the feasible set associated with threshold rules, denoted $\mathcal{E}^{\texttt{th}}$, is convex. To see this, note that

$$\mathcal{E}^{\texttt{th}} = \{\mathbb{E}[\mathcal{M}\vartheta(X)] : \vartheta(X)\in\{\boldsymbol{\theta}_0(X),\boldsymbol{\theta}_1(X)\}\} \equiv \mathbf{E}\left[\mathcal{M}\tilde{\Lambda}(X)\right], \tag{94}$$

with $\boldsymbol{\theta}_d(X)$ defined in Eq. (6) and $\tilde{\Lambda}(X) \equiv \{\boldsymbol{\theta}_0(X),\boldsymbol{\theta}_1(X)\}$. Relative to Eq. (8), the fundamental difference here is that $\{\boldsymbol{\theta}_0(X),\boldsymbol{\theta}_1(X)\}$ is a two-point set instead of an interval. Nonetheless, the set on the right-hand-side of Eq. (94) is by definition (Molchanov and Molinari, 2018, Def. 3.1) the *Aumann expectation* of the two-point set $\tilde{\Lambda}(X)$. Next, observe that under Assumption 2 the probability space on which (Y,G,X) are defined is non-atomic (non-atomicity follows as long as one of the variables in X has a continuous distribution, and if all variables in X had a discrete distribution, Assumption 2 would fail).[14] As both $\boldsymbol{\theta}_0(X)$ and $\boldsymbol{\theta}_1(X)$ are absolutely integrable, all conditions required for Theorem 3.4 in Molchanov and Molinari (2018) are satisfied, yielding:

$$\mathbf{E}\left[\mathcal{M}\tilde{\Lambda}(X)\right] = \mathbf{E}\left[\mathcal{M}\,\text{conv}\left(\{\boldsymbol{\theta}_0(X),\boldsymbol{\theta}_1(X)\}\right)\right] = \mathbf{E}\left[\mathcal{M}\Lambda(X)\right]. \tag{95}$$

[14] To see this, let X have countable support, take $x\in\mathcal{X}$ such that $\mathbb{P}(X=x)=\varsigma>0$. Then for $q = \frac{[\{\theta_1^b(x)-\theta_0^b(x)\}/\mu_b \quad \{\theta_0^r(x)-\theta_1^r(x)\}/\mu_r]^\intercal}{\|\mathcal{M}\{\boldsymbol{\theta}_1(x)-\boldsymbol{\theta}_0(x)\}\|}$, $\mathbb{P}(|q_1\{\theta_1^r(x)-\theta_0^r(x)\}/\mu_r + q_2\{\theta_1^b(x)-\theta_0^b(x)\}/\mu_b| = 0) \geq \varsigma > 0$, and hence $\mathbb{P}(|q_1\{\theta_1^r(x)-\theta_0^r(x)\}/\mu_r + q_2\{\theta_1^b(x)-\theta_0^b(x)\}/\mu_b| < \delta) \geq \varsigma > 0$ for any $\delta > 0$.

Theorem 3.11 in Molchanov and Molinari (2018) also applies, and $h_{\mathcal{E}}(q) = h_{\mathbf{E}[\mathcal{M}\Lambda(X)]}(q) = h_{\mathbf{E}[\mathcal{M}\tilde{\Lambda}(X)]}(q) = \mathbb{E}[h_{\mathcal{M}\Lambda(X)}(q)]$. Hence, under Assumption B.1, the feasible set associated with threshold rules is convex and the support function fully characterizes it.

REMARK B.2—Richness of Threshold Rules: The result in Eq. (95) shows that threshold rules corresponding to a rich algorithm space can replicate any unconstrained algorithm. Inspecting further the linear case is instructive. If one allows for any $\beta \in \mathbb{R}^{d_X+1}$, Assumption B.1 is satisfied. By the same argument given above, the feasible set associated with linear threshold rules, denoted $\mathcal{E}^{\texttt{lin}}$, is convex. Indeed, for each β one can write

$$e_g(\beta) \equiv \frac{1}{\mu_g}\mathbb{E}\left[L_0^g \middle| [1;X]^\intercal\beta < 0\right]\mathbb{P}\left([1;X]^\intercal\beta < 0\right) + \frac{1}{\mu_g}\mathbb{E}\left[L_1^g \middle| [1;X]^\intercal\beta \geq 0\right]\mathbb{P}\left([1;X]^\intercal\beta \geq 0\right).$$

Let $\mathcal{X}_\beta^+ \equiv \{X \in \mathcal{X} : [1;X]^\intercal\beta \geq 0\}$ and $\mathcal{X}_\beta^- \equiv \{X \in \mathcal{X} : [1;X]^\intercal\beta < 0\}$ denote the two sets in which a linear threshold rule with parameter β partitions $\mathcal{X}$. As at least one component of X has continuous distribution and β has support $\mathbb{R}^{d_X+1}$, each realization of X can be allocated either in $\mathcal{X}_\beta^+$ or in $\mathcal{X}_\beta^-$ for some β. Hence, convexification occurs. The support function of $\mathcal{E}^{\texttt{lin}}$ can be expressed as

$$h_{\mathcal{E}^{\texttt{lin}}}(q) = \max_{\beta \in \mathbb{R}^{d_X+1}} q^\intercal e_g(\beta) = \max_{\beta \in \mathbb{R}^{d_X+1}} (\mathcal{M}q)^\intercal \mathbb{E}\left[\mathbf{L}_0 \mathbb{1}([1;X]^\intercal\beta < 0) + \mathbf{L}_1 \mathbb{1}([1;X]^\intercal\beta \geq 0)\right].$$

B.2. *Sufficient Conditions Yielding Strict Convexity and No Kinks*

Assumption 2 plays multiple roles in our analysis. It assures that $\mathcal{E}$ is strictly convex, hence its support set in any direction $q \in \mathbb{S}^1$ is a singleton, and it assures Neyman orthogonality of the moment condition defining $h_{\mathcal{E}}(\cdot)$. Semenova (2023, Section 3.1) provides sufficient conditions for this assumption, based on joint Gaussianity of $\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X)$. A mild strengthening of Assumption 3 is sufficient both for Assumption 2 to hold with $m = 1$ and for Eq. (14) to be satisfied, guaranteeing the absence of kinks in $\mathcal{E}$, as we show next.

ASSUMPTION B.2: *(i) The distribution of $(X_1, X_2)|X_{[3:d_X]}$ is continuous with a bounded density and $\mathbb{E}\left[\left|\eta^g(X_{[3:d_X]})\right|\right] < \infty$ for $g \in \{r, b\}$; (ii) for a set $\tilde{\mathcal{X}}_{[3:d_X]}$ of realizations of $X_{[3:d_X]}$ with positive probability, the density of $(X_1, X_2)|X_{[3:d_X]}$ is positive on a ball of radius $c > 0$ that includes $\mathbf{0}$ and the image of $\tilde{\mathcal{X}}_{[3:d_X]}$ under $\eta^g(\cdot)$ includes 0.*

PROPOSITION B.1: *If Assumptions 3 and B.2(i) hold, Assumption 2 is implied. If Assumption B.2(ii) also holds, Eq. (14) holds and the set $\mathcal{E}$ has no kinks.*

PROOF: Take $\delta > 0$, and note that $\sup_{q\in\mathbb{S}^1} \mathbb{P}(|k(\boldsymbol{\theta}(X), \mathcal{M}q)| < \delta)$ can be written as

$$\begin{aligned}
&\sup_{q\in\mathbb{S}^1} \mathbb{E}_{X_{[3:d_X]}}\left[\int_0^\delta \left|q_1\Delta\theta^r(X)/\mu_r + q_2\Delta\theta^b(X)/\mu_b\right| d\mathbb{P}_{(X_1,X_2)|X_{[3:d_X]}}\right] \\
\lesssim &\max_{g\in\{r,b\}} \mathbb{E}_{X_{[3:d_X]}}\left[\int_0^\delta \left(|\alpha^g| + |\beta^g| + |\eta^g(X_{[3:d_X]})|\right) d\mathbb{P}_{(X_1,X_2)|X_{[3:d_X]}}\right] \\
\lesssim &\max_{g\in\{r,b\}} \mathbb{E}_{X_{[3:d_X]}}\left[\delta\left(|\alpha^g| + |\beta^g| + |\eta^g(X_{[3:d_X]})|\right)\right] \lesssim \delta,
\end{aligned}$$

where the first inequality follows from $\sup_{q\in\mathbb{S}^1} \|q\|_E = 1$, $\mu_g \in (0,1)$, and that (X_1, X_2) has bounded support. The second equality follows from continuity and boundedness of $d\mathbb{P}_{(X_1,X_2)|X_{[3:d_X]}}$. The last inequality follows from $\mathbb{E}\left[\left|\eta^g(X_{[3:d_X]})\right|\right] < \infty$.

For the second result, observe that $\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X) = \Delta\boldsymbol{\theta}(X)$ equals:

$$\underbrace{\begin{bmatrix}\alpha^r & \beta^r \\ \alpha^b & \beta^b\end{bmatrix}}_{\equiv A_1}\begin{bmatrix}X_1 \\ X_2\end{bmatrix} + \underbrace{\begin{bmatrix}\eta^r(X_{[3:d_X]}) \\ \eta^b(X_{[3:d_X]})\end{bmatrix}}_{\equiv A_2},$$

where the matrix A_1 is invertible by $\alpha^b\beta^r \neq \alpha^r\beta^b$. Under Assumption B.2(ii), on the set $\tilde{\mathcal{X}}_{[3:d_X]}$, $A_1[X_1\ X_2]^\intercal + A_2$ realizes in a set containing 0. Hence, Eq. (14) holds by applying the law of iterated expectations. *Q.E.D.*

APPENDIX C: VARIABILITY OF THE EMPIRICAL RESULTS TO NUISANCE PARAMETER ESTIMATION

C.1. *Estimating the Nuisance Parameters Using Lasso*

In this subsection, we report the analogs of Figures 6-7-8 and Tables II-III using multinational logit lasso from the `glmnet` package to estimate the nuisance parameter $\Delta\boldsymbol{\theta}$. That is, the only difference between the results reported in this subsection and those in Section 8.2 lies in the choice of what machine learner is used to estimate nuisance parameters. Respectively, these are Figures C.1-C.2-C.3 and Tables C.I-C.II.

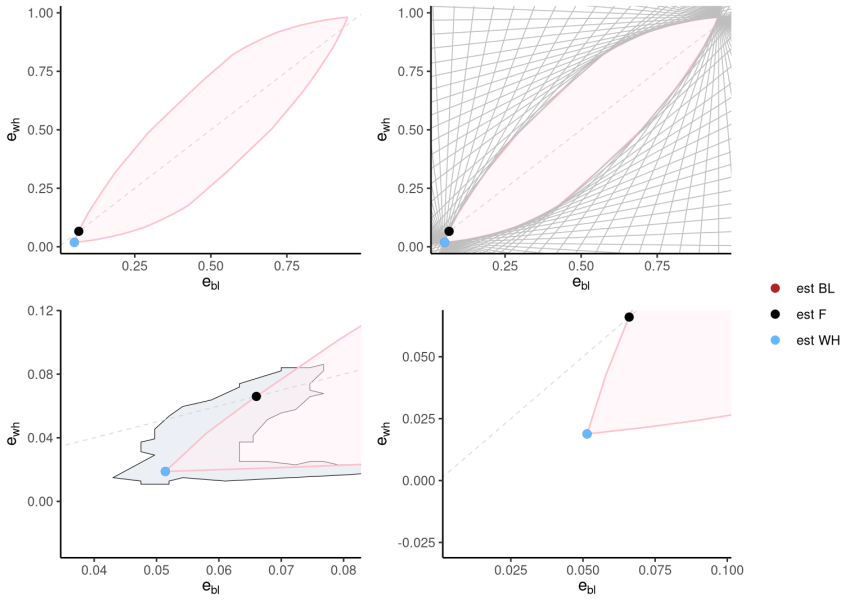
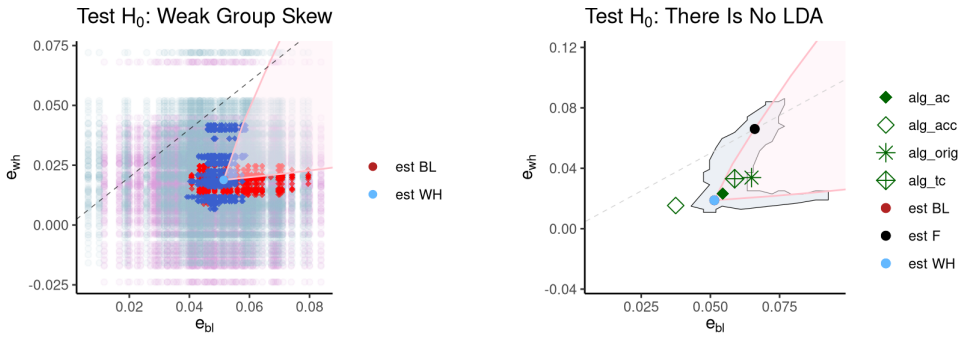


FIGURE C.1.—Top-left panel: $\widehat{\mathcal{E}}$; top-right panel: $\widehat{\mathcal{E}}$ along with one hundred supporting hyperplanes; bottom-left panel: zoom-in to $\widehat{\mathcal{F}}$ and the 95% confidence set around this frontier; bottom-right panel: further zoom-in to the best group-specific points BL and WH, and the fairest point F. $\Delta\boldsymbol{\theta}$ is estimated by logit lasso.



(a) Candidate values for (BL, WH)

(b) 3 experimental algorithms

FIGURE C.2.—Panel (a): Plum-colored (respectively, light-blue colored) circles correspond to candidate values for e_{BL} (e_{WH}) sampled from a normal distribution centered at $\widehat{e}_{\mathsf{BL}}$ ($\widehat{e}_{\mathsf{WH}}$), and red (blue) diamonds correspond to non-rejected values. Panel (b): $\widehat{\mathcal{F}}$ along with its 95% confidence set and the estimated group risks for four algorithms considered by OPVM: the original algorithm used by the hospital (asterisk); one that predicts total cost (hollow diamond with a cross); one that predicts avoidable costs (filled diamond); and one that predicts the number of active chronic conditions (hollow diamond). $\Delta\boldsymbol{\theta}$ is estimated by logit lasso.

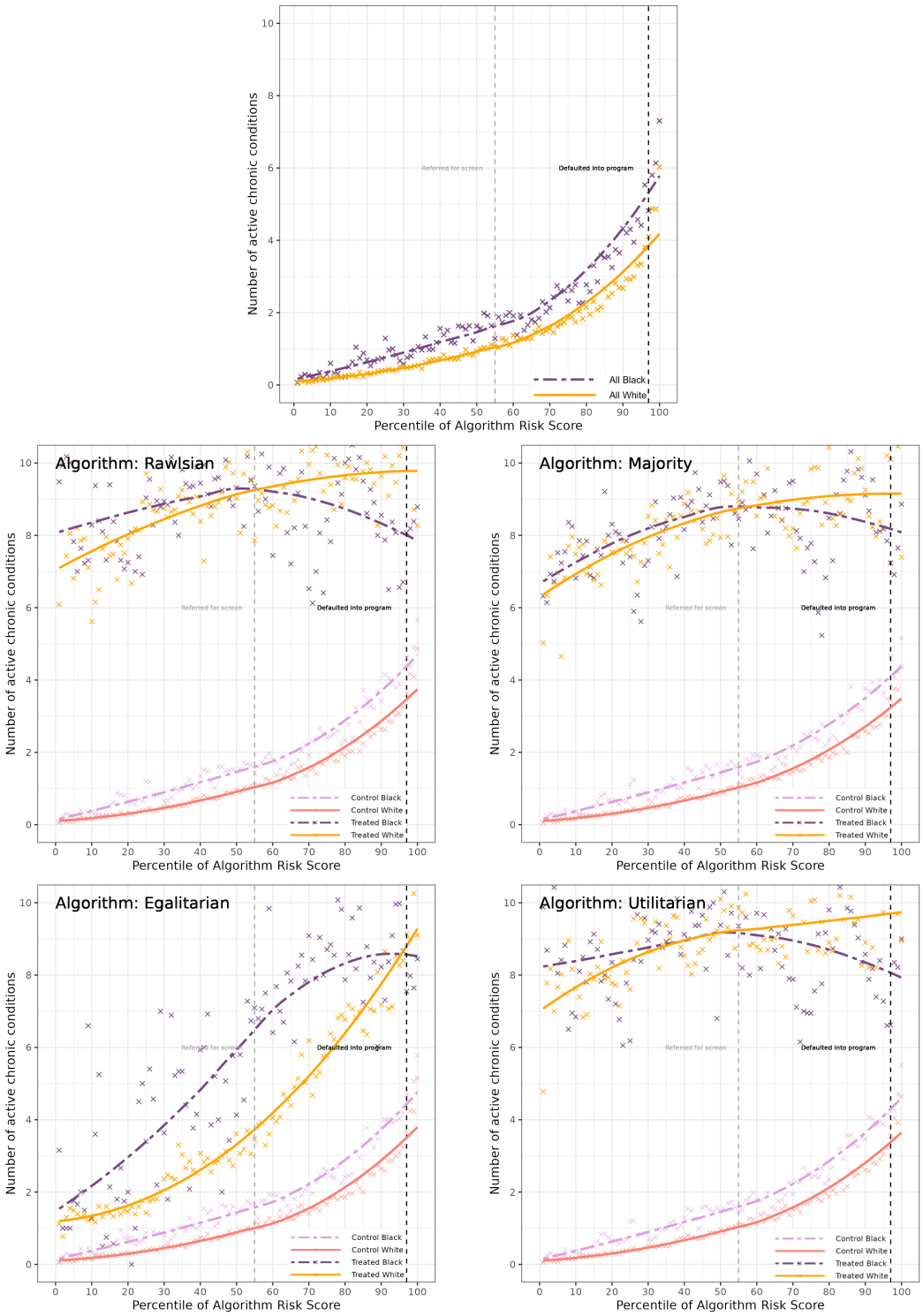


FIGURE C.3.—Average number of active chronic conditions within each risk-score percentile bin by treatment group under the alternative algorithms on the FA frontier subject to 3% capacity constraint, averaged across 20 replications of the 50-50 split. $\Delta\boldsymbol{\theta}$ is estimated by logit lasso.

C.2. *Variability of Empirical Results due to the Randomness in the Nuisance Estimation*

Both random forests and lasso involve randomness in their respective construction: for forests implemented by the `grf` package, randomness comes from subsampling in the construction of individual trees and randomly splitting features at each tree node, whereas for lasso implemented by the `glmnet` package, randomness comes from choosing the optimal penalty parameter via cross-validation. Therefore, even if the same seed is set for reproducibility whenever possible, results will vary across different seeds. For this reason, we provide an assessment of how the empirical results reported in Section 8.2 vary across different seeds by repeating the empirical exercises 20 times, with results for forests and lasso reported respectively in Table C.III and Table C.IV.

TABLE C.I

RESULTS FOR THE LDA TEST AND CONFIDENCE SETS FOR THE DISTANCE TO F (FOR $\alpha = 0.05$)

Test H_0 : *There Is No LDA*				
	Original	Total Costs	Avoid. Costs	Act. Chr. Cond.
Estimated Risks	$(0.065, 0.034)$	$(0.059, 0.033)$	$(0.054, 0.023)$	$(0.037, 0.015)$
Test Statistic	2.043	0.787	0.440	2.774
Critical Value	2.014	1.865	1.937	1.886
Conclusion	Rejected	Not Rejected	Not Rejected	Rejected
Distance to $F = (0.063, 0.063)$				
Estimated Distance	0.0009	0.0009	0.0017	0.0029
Confidence Set	$(0.000, 0.002)$	$(0.000, 0.002)$	$(0.000, 0.003)$	$(0.001, 0.006)$

Top panel: LDA test statistics and 0.05-level critical values associated with the original algorithm and the three experimental algorithms (predicting, respectively, total costs; avoidable costs; number of active chronic conditions) analyzed by OPVM. Bottom panel: estimated squared-Euclidean distance to the F point and corresponding confidence set for this distance. $\Delta\boldsymbol{\theta}$ is estimated by logit lasso

TABLE C.II

FRACTION OF BLACK PATIENTS TREATED AMONG ALL TREATED

	Algorithms from Obermeyer et al.		**Algorithms on the FA-Frontier**			
Capacity Threshold	Original	Counterfactual	Rawlsian	Majority	Egalitarian	Utilitarian
55	0.120	0.184	0.173	0.173	0.174	0.172
69	0.128	0.255	0.217	0.200	0.171	0.202
82	0.138	0.327	0.241	0.224	0.130	0.223
89	0.151	0.407	0.264	0.247	0.118	0.249
94	0.167	0.498	0.324	0.284	0.124	0.294
97	0.184	0.592	0.369	0.318	0.143	0.339

The distribution of the number of active chronic conditions is such that the 55^{th} to the 68^{th} percentiles all correspond to 1 active chronic condition, the 69^{th}-81^{st} correspond to 2, the 82^{nd}-88^{th} correspond to 3, the 89^{th}-92^{nd} correspond to 4, the 94^{th}-95^{th} correspond to 5, and the 96^{th}-97^{th} correspond to 6. $\Delta\boldsymbol{\theta}$ is estimated by logit lasso.

TABLE C.III
VARIABILITY OF EMPIRICAL RESULTS: RANDOM FORESTS

	Mean	SD	Min	25-th	50-th	75-th	Max
Weak Skew Test							
Conclusion	0	0	0	0	0	0	0
LDA Test							
Original Algorithm:							
Test Statistic	3.338	0.318	2.666	3.173	3.301	3.562	3.823
Critical Value	1.894	0.051	1.810	1.866	1.896	1.919	2.004
Conclusion	1	0	1	1	1	1	1
Algorithm that Predicts Total Costs:							
Test Statistic	2.149	0.327	1.475	1.977	2.111	2.361	2.761
Critical Value	1.844	0.066	1.736	1.804	1.829	1.876	1.998
Conclusion	0.8	0.410	0	1	1	1	1
Algorithm that Predicts Avoidable Costs:							
Test Statistic	1.488	0.054	1.398	1.440	1.493	1.521	1.577
Critical Value	1.721	0.060	1.612	1.682	1.724	1.754	1.836
Conclusion	0	0	0	0	0	0	0
Algorithm that Predicts the Number of Active Chronic Conditions:							
Test Statistic	1.194	0.346	0.674	1.007	1.135	1.395	1.812
Critical Value	1.632	0.049	1.516	1.600	1.634	1.654	1.725
Conclusion	0.15	0.366	0	0	0	0	1
Confidence Set for the Distance to F							
Estimated F	0.052	0.003	0.049	0.050	0.052	0.054	0.058
Original Algorithm:							
Estimated Distance	0.0005	0.0000	0.0005	0.0005	0.0005	0.0005	0.0006
Lower 95% CI	0.0001	0.0001	0.0000	0.0000	0.0000	0.0001	0.0002
Upper 95% CI	0.0012	0.0003	0.0008	0.0009	0.0011	0.0014	0.0018
Algorithm that Predicts Total Costs:							
Estimated Distance	0.0004	0.0001	0.0003	0.0004	0.0004	0.0004	0.0006
Lower 95% CI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
Upper 95% CI	0.0013	0.0004	0.0007	0.0010	0.0013	0.0016	0.0022
Algorithm that Predicts Avoidable Costs:							
Estimated Distance	0.0009	0.0002	0.0007	0.0007	0.0008	0.0009	0.0013
Lower 95% CI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
Upper 95% CI	0.0024	0.0005	0.0018	0.0020	0.0023	0.0027	0.0033
Algorithm that Predicts the Number of Active Chronic Conditions:							
Estimated Distance	0.0016	0.0003	0.0012	0.0013	0.0016	0.0017	0.0023
Lower 95% CI	0.0003	0.0002	0.0000	0.0002	0.0003	0.0004	0.0006
Upper 95% CI	0.0041	0.0006	0.0030	0.0035	0.0042	0.0046	0.0051

Table C.III reports the variability of the empirical results in Section 8.2 due to randomness in the estimation of $\Delta\boldsymbol{\theta}$ using random forests, reported as the mean, standard deviation, minimum, the 25-th percentile, 50-th percentile, 75-th percentile, and the maximum across 20 replications. Test conclusions are recorded as 1 if the conclusion is rejection, and 0 otherwise.

TABLE C.IV
VARIABILITY OF EMPIRICAL RESULTS: LOGIT LASSO

	Mean	SD	Min	25-th	50-th	75-th	Max
Weak Skew Test							
Conclusion	0	0	0	0	0	0	0
LDA Test							
Original Algorithm:							
Test Statistic	2.037	0.182	1.784	1.874	2.001	2.166	2.390
Critical Value	1.976	0.074	1.762	1.937	1.989	2.024	2.066
Conclusion	0.550	0.510	0	0	1	1	1
Algorithm that Predicts Total Costs:							
Test Statistic	0.804	0.173	0.562	0.651	0.788	0.908	1.154
Critical Value	1.915	0.055	1.801	1.890	1.914	1.961	1.995
Conclusion	0	0	0	0	0	0	0
Algorithm that Predicts Avoidable Costs:							
Test Statistic	0.478	0.171	0.100	0.375	0.491	0.591	0.821
Critical Value	1.913	0.057	1.783	1.888	1.923	1.941	2.006
Conclusion	0	0	0	0	0	0	0
Algorithm that Predicts the Number of Active Chronic Conditions:							
Test Statistic	2.769	0.174	2.470	2.668	2.754	2.889	3.184
Critical Value	1.873	0.058	1.768	1.841	1.880	1.905	2.020
Conclusion	1	0	1	1	1	1	1
Confidence Set for the Distance to F							
Estimated F	0.063	0.001	0.061	0.062	0.062	0.063	0.065
Original Algorithm:							
Estimated Distance	0.0008	0.0001	0.0007	0.0008	0.0008	0.0009	0.0010
Lower 95% CI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
Upper 95% CI	0.0020	0.0003	0.0014	0.0017	0.0019	0.0021	0.0026
Algorithm that Predicts Total Costs:							
Estimated Distance	0.0009	0.0001	0.0008	0.0008	0.0009	0.0009	0.0010
Lower 95% CI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
Upper 95% CI	0.0022	0.0003	0.0017	0.0021	0.0022	0.0025	0.0025
Algorithm that Predicts Avoidable Costs:							
Estimated Distance	0.0016	0.0001	0.0014	0.0015	0.0016	0.0017	0.0018
Lower 95% CI	0.0003	0.0001	0.0000	0.0002	0.0003	0.0004	0.0005
Upper 95% CI	0.0035	0.0004	0.0030	0.0032	0.0035	0.0038	0.0042
Algorithm that Predicts the Number of Active Chronic Conditions:							
Estimated Distance	0.0028	0.0002	0.0026	0.0027	0.0028	0.0029	0.0032
Lower 95% CI	0.0009	0.0003	0.0005	0.0008	0.0009	0.0011	0.0016
Upper 95% CI	0.0054	0.0004	0.0047	0.0051	0.0054	0.0057	0.0062

Table C.IV reports the variability of the empirical results in Section 8.2 due to randomness in the estimation of $\Delta\boldsymbol{\theta}$ using logit lasso, reported as the mean, standard deviation, minimum, the 25-th percentile, 50-th percentile, 75-th percentile, and the maximum across 20 replications. Test conclusions are recorded as 1 if the conclusion is rejection, and 0 otherwise.