

IDENTIFICATION AND INFERENCE
FOR ALGORITHMIC FRONTIERS WITH SELECTIVE LABELS

YIQI LIU

Department of Economics, Cornell University

FRANCESCA MOLINARI

Department of Economics, Cornell University

AMILCAR VELEZ

Department of Economics, Cornell University

This paper provides identification results to characterize a fairness-accuracy (FA) frontier, and statistical inference tools to test hypotheses and build a confidence set for the FA-frontier, when outcomes are observed only for selected individuals. When the selection process is unrestricted but loss is measured in specific ways, we provide a characterization of the sharp identification region of the FA-frontier. Under an assumption of unconfoundedness conditional on observables (and unrestricted loss functions), we obtain point identification and propose a de-biased machine learning estimator, derive its asymptotic distribution, and show how this can be used to carry out inference for the FA-frontier. In work in progress, we extend the partial identification results to a broader class of loss functions.

KEYWORDS: Algorithmic fairness, selective labels, statistical inference, support function.

1. INTRODUCTION

Over the last decade, algorithms have come to play an increasingly prevalent role in everyday life, often by supporting high-stake decisions through predictions of quantities such

Yiqi Liu: y13467@cornell.edu

Francesca Molinari: fm72@cornell.edu

Amilcar Velez: amilcare@cornell.edu

Preliminary version. This draft: June 12, 2026. We thank José Montiel-Olea for helpful comments.

as re-offense risk, repayment likelihood, and health care needs. These predictions enter institutional decision processes that determine who is detained or released, who is granted credit, and which patients receive outreach and care-management interventions. The increasing reliance on algorithms has prompted the emergence of a set of regulatory and oversight processes, known as *AI governance*, that rule the design, deployment, documentation, and monitoring of AI systems, including, in some settings, restrictions on the use of protected attributes as model inputs. A distinctive feature of these regimes is their evidentiary orientation: for high-risk AI systems, the EU AI Act requires technical documentation, logging capabilities, and appropriate accuracy and robustness over the lifecycle ([European Parliament and Council, 2024](#), Articles 11, 12, 15, 26), while related governance proposals emphasize dataset documentation, model reporting, and internal auditing (see, e.g., [Gebru et al., 2021](#), [Mitchell et al., 2019](#), [Raji et al., 2020](#)).

This paper extends the research agenda set out in [Liu and Molinari \(2026\)](#) to provide a toolkit for nonparametric estimation of and statistical inference on the fairness-accuracy trade-off that characterizes algorithms. Fairness and accuracy evaluations play a central role in governance, as they determine both the validity of model-supported decisions and whether those decisions impose disproportionate harms on legally protected groups.¹ At the same time, governance questions are rarely about a single fixed model. In public oversight and non-discrimination doctrine, the relevant question is often whether a challenged procedure can be replaced by an alternative that achieves comparable accuracy with less adverse impact, shifting attention from a single algorithm to a potentially large class of admissible decision rules, e.g., those obtained under constraints on inputs, model classes, or the use of sensitive attributes. Regulators and oversight bodies are called to adjudicate (i) whether a deployed rule is Pareto efficient, (ii) whether design restrictions enable less discriminatory alternatives (LDAs) with comparable accuracy, and (iii) whether such LDAs

¹Algorithms may exhibit systematic differences in performance across legally protected groups, both in predictive accuracy and in the decisions they induce (e.g., [Angwin et al., 2016](#), [Arnold et al., 2021](#), [Obermeyer et al., 2019](#), [Cowgill and Tucker, 2020](#), [Berk et al., 2021](#)). Yet, improving parity in expected losses often requires sacrificing accuracy for at least one group, and conversely increasing accuracy for one group may worsen disparities.

exist within an admissible class of decision rules.² As in [Liu and Molinari \(2026\)](#), we build on the theoretical framework put forward by [Liang, Lu, Mu, and Okumura \(2026, LLMO henceforth\)](#), which in our view is well-suited to formalize these questions because it characterizes a fairness-accuracy (FA) frontier under broad preferences over group-specific expected losses as well as the implications of banning attributes as inputs to the algorithms.

Our first innovation relative to [Liu and Molinari \(2026\)](#)'s groundwork is to allow for accuracy and fairness of an algorithm to be assessed through two different loss functions, thereby aligning our treatment with the literature in computer science ([Corbett-Davies et al., 2017](#), [Menon and Williamson, 2018](#), [Kim et al., 2020](#), [Gillis et al., 2024](#)). We show that the support function of a convex feasible set $\mathcal{E} \subset \mathbb{R}^3$ of expected accuracy losses for the two groups and the difference in expected fairness losses between the groups plays a crucial role in characterizing the FA-frontier and carrying out inference.³ We extend the results in [LLMO](#)'s Appendix O by providing a simple and testable necessary and sufficient condition under which the FA-frontier coincides with the Pareto frontier. When this condition is not satisfied, a policymaker who values fairness may choose to deploy an algorithm that is Pareto dominated in accuracy loss, to ameliorate its performance in fairness loss.

Our second innovation addresses the reality that in many high-stakes settings, outcomes are observed only for selected individuals: pretrial misconduct (or lack thereof) is observed only if bail is granted; repayment behavior is observed only when credit is extended; certain health outcomes are observed only for patients who receive treatment or remain under follow-up within a system; etc. Hence, instead of observing the outcome vector of interest Y^* , one observes ZY^* , where $Z \in \{0, 1\}$ indicates observability. This *selective labels* problem creates challenges for governance compliance similar to the ones long recognized in

²For example, US employment discrimination law prohibits selection procedures that cause disparate impact unless the employer demonstrates they are “job related for the position in question and consistent with business necessity.” Even then the practice is unlawful if an alternative practice would serve the employer’s legitimate interests with less discriminatory impact ([U.S. Congress, 1991](#)).

³In [Liu and Molinari \(2026\)](#)'s analysis and [LLMO](#)'s main text, the same loss function is used to measure fairness and accuracy, so that the feasible set \mathcal{E} is in \mathbb{R}^2 . [Liu and Molinari \(2026\)](#) propose to use the support function of $\mathcal{E} \subset \mathbb{R}^2$ as the unifying tool for statistical inference.

the vast econometrics literature on selectively observed data (e.g., Heckman, 1979, Manski, 1989). Accuracy metrics computed using only observed outcomes may misrepresent true performance, potentially overstating accuracy where it matters most (e.g., if those denied bail would not have re-offended). Hence, assessments of whether a system achieves “appropriate levels of accuracy” (EU AI Act, European Parliament and Council, 2024, Article 15) or whether LDAs exist (U.S. Congress, 1991) may be incorrect and required documentation may systematically exclude the most policy-relevant counterfactual outcomes.

When classification loss is used to measure accuracy and statistical parity is used to measure fairness (as in Menon and Williamson, 2018), and no restrictions are imposed on the selection process, we characterize the sharp identification region for the FA-frontier in the spirit of Manski (1989) (see Molinari, 2020, Canay et al., 2026, for reviews of the partial identification literature). Doing so is challenging because the FA-frontier is ultimately about *algorithms*: it equals the set of expected losses $\varepsilon(a) \in \mathcal{E}$ associated with algorithms $a \in \mathcal{A}(\mathcal{X})$ (measurable functions mapping covariates $X \in \mathcal{X}$ to $[0, 1]$) such that no other algorithm $\tilde{a} \in \mathcal{A}(\mathcal{X})$ yields expected losses $\varepsilon(\tilde{a})$ that FA-dominate $\varepsilon(a)$ in a sense made precise in Definition 5. When labels are selectively observed, both $\varepsilon(a)$ and \mathcal{E} are partially but not point identified. Hence, to pass judgment on whether algorithm a yields expected losses $\varepsilon(a)$ on the frontier, care needs to be taken to couple the *same* candidate distribution for the missing labels in constructing both $\varepsilon(a)$ and \mathcal{E} , and then assess membership across all admissible distributions for the missing labels. We reduce this daunting task to solving a finite dimensional optimization problem.

When the selection process is restricted by a missing-at-random (MAR) assumption, where conditional on attributes, outcome observability does not convey additional information on Y^* , we obtain point identification of the FA-frontier. For any loss function, we provide a debiased machine learning (DML) estimator for the FA-frontier based on inverse probability weighting. Our DML approach is particularly appropriate for governance contexts, where high-dimensional administrative data enable flexible nonparametric modeling of selection and formal inference tools are needed for legally defensible determinations. While ideas transfer from the program evaluation and high-dimensional inference literature

(e.g., [Chernozhukov et al. \(2024\)](#)), we extend the methods to carry out inference on a new set-valued estimand, the FA-frontier. We derive the asymptotic distribution of our estimator and test statistics building on [Liu and Molinari \(2026\)](#), [Semenova \(2023\)](#), [Chandrasekhar et al. \(2018\)](#), [Chernozhukov et al. \(2018\)](#), [Beresteanu and Molinari \(2008\)](#), and [Fang and Santos \(2019\)](#), provide inverse-probability weighting correction in the DML estimator, and asymptotically valid confidence sets for the FA frontier and tests for LDA existence.

Accompanying software implementing both the MAR-based and partial identification approaches is under development, along with empirical applications of our methods.

Related Literature. A large literature in computer science, statistics, and economics studies algorithmic fairness; see [Chouldechova and Roth \(2018\)](#), [Barocas et al. \(2023\)](#), and [Corbett-Davies et al. \(2023\)](#) for overviews. Much of this work treats fairness as a constraint or regularizer in an objective that prioritizes predictive performance (e.g., [Dwork et al., 2012](#), [Berk et al., 2017](#)), or studies impossibility results and trade-offs among fairness criteria ([Kleinberg et al., 2016](#)). A related literature computes fairness-accuracy Pareto frontiers or ranges of disparities for specific model classes, typically as an optimization or auditing exercise rather than as a problem of statistical inference (e.g., [Wei and Niethammer, 2022](#), [Little et al., 2022](#), [Coston et al., 2021](#)). When labels are perfectly observed, [Auerbach et al. \(2024\)](#) provide a sample-splitting based test for LDA existence, and [Fallah et al. \(2026\)](#) study minimax-optimality of the sample analog of a Pareto-optimal linear rule and propose a uniform high-probability bound on its empirical error.

Within the algorithmic fairness literature, other work has focused on selective labels (e.g., [Lakkaraju et al., 2017](#), [Kleinberg et al., 2018](#), [Rambachan et al., 2025](#), [Khan et al., 2025](#)). [Rambachan et al. \(2025\)](#) develop a general framework for robust evaluation and design of predictive algorithms under selective labels and unobserved confounding, delivering bounds for performance measures under alternative restrictions on the selection process. We view their contribution as addressing a distinct question from ours: how to evaluate predictive performance in the presence of counterfactual outcomes. We focus instead on identification and inference for the fairness-accuracy trade-off of group losses across all feasible algorithms using a given information set, and on hypothesis tests and confidence

statements about frontier properties such as LDA existence that directly map to governance questions and design restrictions.

Outline. Section 2 lays out notation and derives the FA-frontier extending [LLMO](#) and [Liu and Molinari \(2026\)](#). Section 3 formalizes the selective labels problem, derives the sharp identification region for the FA-frontier for specific loss functions and unrestricted selection process, and obtains point identification under a MAR assumption for any loss function. Section 4 puts forward our DML estimator and derives its asymptotic distribution. Section 5 proposes a test for equality between the Pareto and FA-frontier. It then puts forward our test for existence of an LDA in the presence of selectively observed labels and derives its asymptotic distribution, building on [Liu and Molinari \(2026\)](#). Section 6 concludes. Our main proofs are in Appendix A; Appendix B reports auxiliary results.

2. SETUP

Let a population of individuals be described by an outcome vector $Y^* \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$, a binary group identity $G \in \{r, b\}$ (red or blue), and a vector of covariates $X \in \mathcal{X} \subset \mathbb{R}^{d_X}$. For example, one component of Y^* may denote an individual’s number of active chronic illnesses in the subsequent year and the other component the cost of care; G may denote the individual’s race, and X may include age, gender, biomarkers, comorbidity, and medication variables. As in [Liu and Molinari \(2026\)](#), we leave the relation between G and X unspecified, but require G not to be a deterministic function of X . Each individual receives a binary decision $D \in \{0, 1\}$, e.g., whether they are automatically enrolled in a high-risk care management program; or granted bail; or granted or denied a loan (while we assume G and D to be binary, the results can be extended to multiple groups and multiple decisions). An algorithm $a : \mathcal{X} \mapsto [0, 1]$ assigns a probability distribution to D conditional on X ; e.g., the algorithm assigns each patient a health risk score in $[0, 1]$; or a recidivism probability; or a repayment probability. For simplicity, we take $a(X)$ to be the only input to the decision, and hence $D \sim a(X)$. We let $\mathcal{A}(\mathcal{X})$ denote the set of all (measurable) algorithms that map from the input space \mathcal{X} to a probability distribution over D . Extending the framework in [Liu and Molinari \(2026\)](#), we let $\ell^A : \{0, 1\} \times \mathcal{Y} \mapsto \mathbb{R}$ be a loss function used to measure the *accuracy* of an algorithm for an individual with outcome $y \in \mathcal{Y}$, and $\ell^F : \{0, 1\} \times \mathcal{Y} \mapsto \mathbb{R}$ a

loss function used to measure the algorithm’s *fairness*. For example, one could use classification error to measure accuracy ($\ell^A(d, y) = \mathbb{1}(d \neq y_1)$), which in a health care example 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. And one could use statistical parity to measure fairness ($\ell^F(d, y) = d$), hence judging an algorithm more fair if the proportion of either group receiving the two decisions is closer.

This paper is concerned with the case where instead of observing Y^* , we observe $Y = ZY^*$, where $Z \in \{0, 1\}$ is an indicator variable that records when the outcome of interest Y^* is observed in the testing data. In other words, we allow for selectively observed labels. Before analyzing this problem, however, we provide a tractable characterization of the FA-frontier, under the assumption that $(Y^*, X, G, Z) \sim \mathbb{P}^*$ is observed, when the loss function used to measure accuracy differs from that used to measure fairness. In doing so, we extend the analysis in Appendix O.1 of [LLMO](#). Armed with our novel characterization, we then study what can be learned about the FA-frontier and how can valid statistical inference be carried out, when in fact only the selectively observed labels Y are available.

2.1. The Fairness-Accuracy Frontier

We denote $\mu_g \equiv \mathbb{P}^*(G = g)$ the population proportion of group $g \in \{r, b\}$ and, for $d \in \{0, 1\}$, the (unobserved) *ideal labels* associated with loss function $\iota \in \{A, F\}$ by

$$L_d^{g,\iota} \equiv \frac{\ell^\iota(d, Y^*) \mathbb{1}\{G=g\}}{\mu_g}. \quad (1)$$

We denote the conditional expectation of $L_d^{g,\iota}$ taken with respect to $\mathbb{P}^*(Y^*, G|X)$ by

$$\theta_d^{g,\iota}(X) \equiv \mathbb{E}^*[L_d^{g,\iota}|X] = \frac{\mathbb{E}^*[\ell^\iota(d, Y^*) \mathbb{1}\{G=g\}|X]}{\mu_g}.$$

DEFINITION 1—Ideal labels and conditional expectations: For $d \in \{0, 1\}$, define

$$\mathbf{L}_d \equiv \begin{bmatrix} L_d^{r,A} \\ L_d^{b,A} \\ L_d^{r,F} - L_d^{b,F} \end{bmatrix} \in \mathbb{R}^3, \quad (2)$$

and let $\boldsymbol{\theta}_d(X) \equiv \mathbb{E}^*[\mathbf{L}_d|X] \in \mathbb{R}^3$ and $\Delta\boldsymbol{\theta}(X) \equiv \boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X) \in \mathbb{R}^3$.

To make sure that the ideal labels are well-defined, we assume:

ASSUMPTION 1—Moment restrictions for both losses: *For all $d \in \{0, 1\}$, $g \in \{r, b\}$, and $\iota \in \{A, F\}$, for some constants $0 < c_1 < 1$, $0 < c_2 < \infty$: (i) $\mu_g \in (c_1, 1 - c_1)$ and $\text{ess sup}_{X \in \mathcal{X}} \mathbb{E}^* \left[(L_d^{g, \iota})^2 | X \right] < c_2$; (ii) $\text{ess inf}_{X \in \mathcal{X}} \text{eig}_{\min} [\text{Var}^*(\mathbf{L}_d | X)] > c_1$, with eig_{\min} denoting the smallest eigenvalue.*

DEFINITION 2—Algorithm space: An *algorithm* is a Borel-measurable map $a : \mathcal{X} \rightarrow [0, 1]$, where $a(x)$ represents the probability of choosing $D = 1$ given $X = x$. Denote $\mathcal{A}(\mathcal{X})$ the set of all such measurable maps. Write \mathcal{A} when the covariate space is clear from context.

For each $a \in \mathcal{A}$ and $g \in \{r, b\}$, define the group-wise expected loss associated with loss function ℓ^ι , $\iota \in \{A, F\}$, when $D \sim a(X)$ as:

$$e_g^\iota(a) \equiv \mathbb{E}^*[\ell^\iota(D, Y^*) | G = g] = \mathbb{E}^*[a(X)\theta_1^{g, \iota}(X) + (1 - a(X))\theta_0^{g, \iota}(X)]. \quad (3)$$

LLMO (Appendix O.1) derive the FA-frontier and study its properties when the loss function used to measure accuracy differs from that used to measure fairness, which is the framework used in this paper. Here we adapt their definitions to our notation.⁴ They work in \mathbb{R}^2 with a feasible set defined with respect to the *accuracy* loss:

$$\mathcal{E}^A = \{(e_r^A(a), e_b^A(a)) : a \in \mathcal{A}\} \subset \mathbb{R}^2. \quad (4)$$

LLMO measure the *unfairness* of an algorithm a through $|f(a)|$, with $f : \mathcal{A} \rightarrow \mathbb{R}$ a linear functional. Their framework includes as a special case $f(a) = e_r^F(a) - e_b^F(a)$, the disparity in expected fairness loss between groups, and for simplicity we specialize our discussion to that case. For each pair of expected accuracy losses $c = (c_r, c_b) \in \mathcal{E}^A$, define the minimal achievable unfairness as

$$d(c) \equiv \min\{|f(a)| : (e_r^A(a), e_b^A(a)) = c, a \in \mathcal{A}\}. \quad (5)$$

⁴They measure accuracy with $\ell : \{0, 1\} \times \mathcal{Y} \rightarrow \mathbb{R}$ (our ℓ^A) and fairness with $\tilde{\ell} : \{0, 1\} \times \mathcal{Y} \rightarrow \mathbb{R}$ (our ℓ^F).

DEFINITION 3—LLMO’s FA-dominance and FA-frontier : For $c, c' \in \mathcal{E}^A$, we say $c' \succ_{FALLMO} c$ if $c'_r \leq c_r$, $c'_b \leq c_b$, $d(c') \leq d(c)$, with at least one inequality strict. Then the FA-frontier relative to the accuracy loss is:

$$\mathcal{F}^A \equiv \{c \in \mathcal{E}^A : \nexists c' \in \mathcal{E}^A \text{ such that } c' \succ_{FALLMO} c\} \subset \mathbb{R}^2. \quad (6)$$

Given the group-wise expected losses induced by algorithm $a \in \mathcal{A}$ for $\iota \in \{A, F\}$ in (3), we instead propose to work with the *feasible set* for $\{(e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a)) : a \in \mathcal{A}\}$, where we explicitly include a coordinate measuring the difference in expected fairness loss for the two groups. Doing so allows us to characterize the properties of the FA frontier, extending LLMO’s results, and to provide a representation of it that is amenable to estimation and inference, even in the presence of selectively observed data.

DEFINITION 4—Feasible set: Given \mathbb{P}^* and \mathcal{A} , as a ranges in \mathcal{A} , the *feasible set* for $(e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a))$ is

$$\mathcal{E} \equiv \left\{ (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in \mathbb{R}^3 : \exists a \in \mathcal{A} \text{ such that } \varepsilon_1 = e_r^A(a), \varepsilon_2 = e_b^A(a), \varepsilon_3 = e_r^F(a) - e_b^F(a) \right\}.$$

To simplify notation, we omit the dependence of \mathcal{E} on $(\mathbb{P}^*, \mathcal{A})$. In Lemma B.2, we show that under Assumptions 1-2, \mathcal{E} is non-empty, compact, and convex. Definition 4 leads to an alternative formulation of FA-dominance and to an FA-frontier in \mathbb{R}^3 , as follows:⁵

DEFINITION 5—FA-dominance and FA-frontier: Given $\varepsilon, \varepsilon' \in \mathcal{E}$, $\varepsilon' \succ_{FA} \varepsilon$ if $\varepsilon'_1 \leq \varepsilon_1$, $\varepsilon'_2 \leq \varepsilon_2$, $|\varepsilon'_3| \leq |\varepsilon_3|$, with at least one inequality strict. The FA-frontier is:

$$\mathcal{F} \equiv \{\varepsilon \in \mathcal{E} : \nexists \varepsilon' \in \mathcal{E} \text{ such that } \varepsilon' \succ_{FA} \varepsilon\} \subset \mathbb{R}^3. \quad (7)$$

We also define the Pareto frontier, which ignores considerations of fairness:

$$\mathcal{PF} \equiv \left\{ \varepsilon \in \mathcal{E} : \nexists \varepsilon' \in \mathcal{E} \text{ s.t. } \varepsilon'_1 \leq \varepsilon_1, \varepsilon'_2 \leq \varepsilon_2, \text{ with at least one strict inequality} \right\}, \quad (8)$$

⁵Auerbach et al. (2024, p.8) use a similar definition of dominance.

and the points in \mathbb{R}^3 that minimize, respectively, e_r^A , e_b^A , and $|e_r^F - e_b^F|$:

$$R \equiv \arg \min_{\varepsilon \in \mathcal{E}} \varepsilon_1, \quad B \equiv \arg \min_{\varepsilon \in \mathcal{E}}, \quad F \equiv \arg \min_{\varepsilon \in \mathcal{E}} |\varepsilon_3|. \quad (9)$$

In Appendix B.1.1 we show that the shape of our FA-frontier \mathcal{F} depends on the position of \mathcal{E} , \mathcal{PF} , R and B relative to the hyperplane $\{\varepsilon_3 = 0\} = \{e_r^F = e_b^F\}$. Figure 1 illustrates several examples. In the next section, we establish that our FA-frontier and LLMO's FA-frontier coincide in what we argue is the only way that matters, and we provide a characterization of \mathcal{F} that does *not* depend on the position of \mathcal{E} , \mathcal{PF} , R and B .

2.2. Support Function-Based Characterization of the FA-Frontier

To begin, we characterize (similarly to Liu and Molinari, 2026, Propositions 3.1-3.2) the support function of \mathcal{E} , denoted $h_{\mathcal{E}}(q)$ for q a direction vector in $\mathbb{S}^2 \equiv \{q \in \mathbb{R}^3 : \|q\| = 1\}$. This functional turns out to be crucial for all our results. We assume:

ASSUMPTION 2—Margin condition: *There exists a constant $m \in (0, 1]$ such that for every $\delta > 0$, $\sup_{q \in \mathbb{S}^2} \mathbb{P}^*(|q^\top \Delta \boldsymbol{\theta}(X)| \leq \delta) \lesssim \delta^m$.*

THEOREM 2.1—Support function of \mathcal{E} : *Under Assumptions 1(i)-2, the support function of \mathcal{E} , denoted $h_{\mathcal{E}}(q)$, and its gradient with respect to $q \in \mathbb{S}^2$ are, respectively,*

$$h_{\mathcal{E}}(q) \equiv \sup_{\varepsilon \in \mathcal{E}} q^\top \varepsilon = \mathbb{E}^*[q^\top \mathbf{L}_0 + q^\top (\mathbf{L}_1 - \mathbf{L}_0) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}], \quad (10)$$

$$\nabla_q h_{\mathcal{E}}(q) = \mathbb{E}^*[\mathbf{L}_0 + (\mathbf{L}_1 - \mathbf{L}_0) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}], \quad (11)$$

and the set \mathcal{E} is strictly convex.

Eq. (11) implies that $\mathcal{S}_{\mathcal{E}}(q) \equiv \{\varepsilon \in \mathcal{E} : q^\top \varepsilon = h_{\mathcal{E}}(q)\}$, the support set of \mathcal{E} in direction $q \in \mathbb{S}^2$, equals the singleton $\nabla_q h_{\mathcal{E}}(q)$ (Schneider, 1993, Corollary 1.7.3). It yields that

$$R = \mathbb{E}^*[\mathbf{L}_0 + (\mathbf{L}_1 - \mathbf{L}_0) \mathbb{1}\{[-1, 0, 0]^\top \Delta \boldsymbol{\theta}(X) > 0\}],$$

$$B = \mathbb{E}^*[\mathbf{L}_0 + (\mathbf{L}_1 - \mathbf{L}_0) \mathbb{1}\{[0, -1, 0]^\top \Delta \boldsymbol{\theta}(X) > 0\}].$$

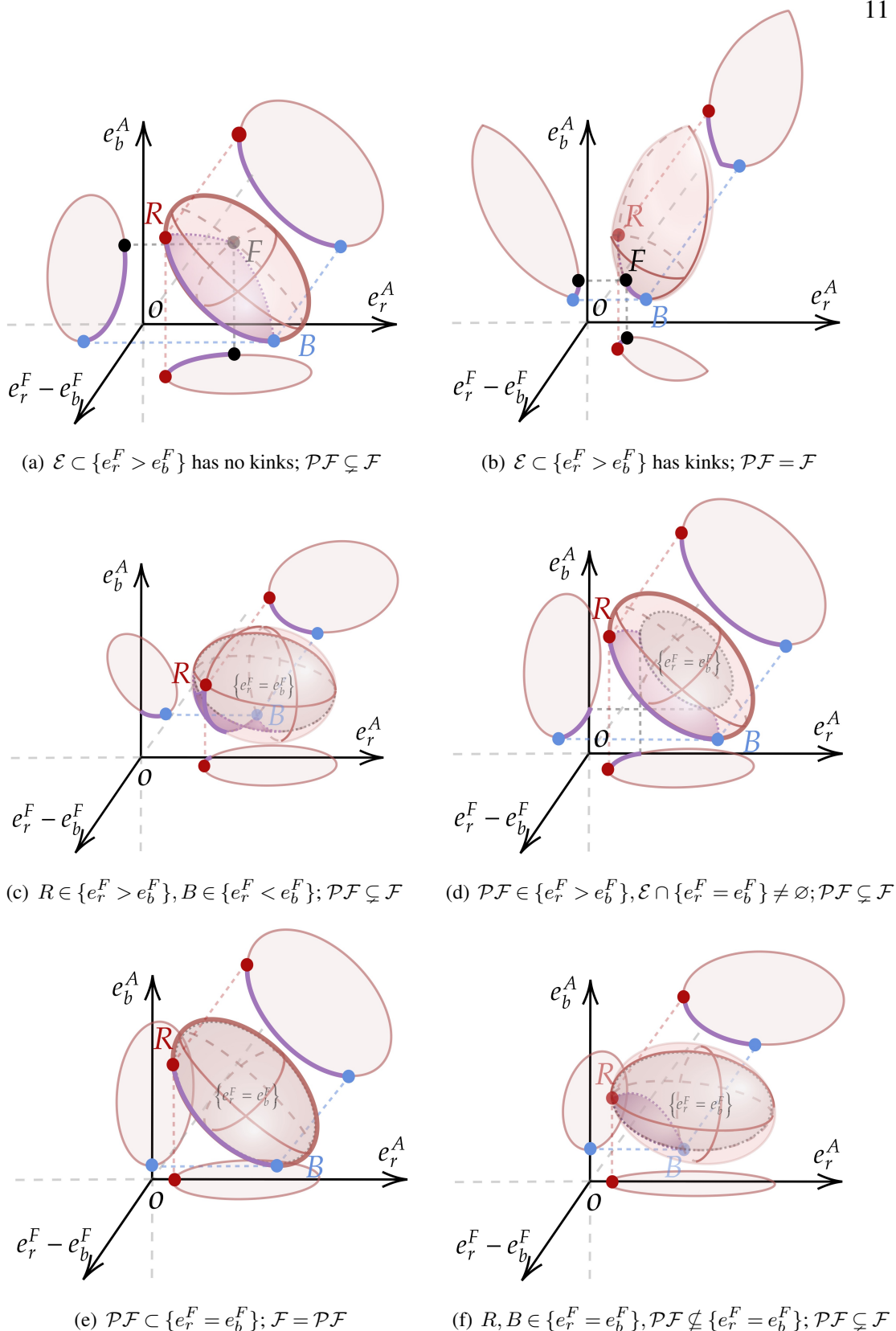


FIGURE 1.— \mathcal{E} is the pink ellipsoid; its projections onto the two-dimensional coordinate planes are plotted as ellipses. \mathcal{F} is the purple shaded area/curve; $\mathcal{P}\mathcal{F}$ is the purple curve connecting R and B . Rear-surface curves on \mathcal{E} are dashed; front-surface curves are solid. Illustrations inspired by analytic expressions.

The next corollary, proven in [Liu and Molinari \(2026, Corollary 3.1\)](#), characterizes precisely which algorithms yield expected losses on the boundary of \mathcal{E} :

COROLLARY 2.1: *Let Assumptions 1(i)-2 hold. Let $\partial\mathcal{E} = \{\mathcal{S}_{\mathcal{E}}(q) : q \in \mathbb{S}^2\}$ denote the boundary of \mathcal{E} . Then for any algorithm $a \in \mathcal{A}(\mathcal{X})$, $\varepsilon(a) \equiv (e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a)) \in \partial\mathcal{E}$ if and only if for some $q \in \mathbb{S}^2$, $a(X) = \mathbb{1}\{q^\top \Delta\theta(X) > 0\}$, \mathbb{P}_X^* -a.s.*

Corollary 2.1 echos the generalized Neyman-Pearson Lemma ([Lehmann and Romano, 2008, Theorem 3.6.1](#)).⁶ One of its important consequences is that *no* algorithm such that $a(X) \in (0, 1)$ for a set of X of positive probability can yield group risks on $\partial\mathcal{E}$.

As the goal of this section is to characterize \mathcal{F} , given $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*) \in \mathcal{E}$, we define

$$\mathcal{C}(\varepsilon^*) \equiv \{\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in \mathbb{R}^3 : \varepsilon_1 \leq \varepsilon_1^*, \varepsilon_2 \leq \varepsilon_2^*, |\varepsilon_3| \leq |\varepsilon_3^*|\}, \quad (12)$$

the set of group-wise expected accuracy losses and fairness disparity $\varepsilon \in \mathbb{R}^3$ —whether or not they are feasible—that are both weakly more accurate and weakly fairer than ε^* . In Lemma B.1, we show that the support function of $\mathcal{C}(\varepsilon^*)$ equals:

$$h_{\mathcal{C}(\varepsilon^*)}(q) \equiv \sup_{\varepsilon \in \mathcal{C}(\varepsilon^*)} q^\top \varepsilon = \begin{cases} q_1 \varepsilon_1^* + q_2 \varepsilon_2^* + |q_3| |\varepsilon_3^*| & \text{if } q_1 \geq 0, q_2 \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (13)$$

Given an algorithm $a^* \in \mathcal{A}$ and $\varepsilon^* \equiv (e_r^A(a^*), e_b^A(a^*), e_r^F(a^*) - e_b^F(a^*))$, we establish in Theorem 2.2 that $\varepsilon^* \in \mathcal{F}$ if and only if the set $\mathcal{C}(\varepsilon^*)$, depicted in the two panels of Figure 2 as the shaded green hyper-rectangle corresponding to two different values of ε^* , can be properly separated from \mathcal{E} (for a definition of proper separation, see [Schneider, 1993, p.12](#)). Panel (a) depicts a case where $\varepsilon^* \in \mathcal{F}$ and proper separation occurs; panel (b) depicts a case where $\varepsilon^* \notin \mathcal{F}$ and proper separation fails. We also show that $\varepsilon^* \in \mathcal{F}$ if and only if $(\varepsilon_1^*, \varepsilon_2^*) \in \mathcal{F}^A$ with $\varepsilon_1^* = e_r^A(a^*)$ and $\varepsilon_2^* = e_b^A(a^*)$, and $|\varepsilon_3^*|$ equals the minimum absolute disparity in (5) associated with $(\varepsilon_1^*, \varepsilon_2^*)$. To do so, we define the projection $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ by $\pi(x_1, x_2, x_3) = (x_1, x_2)$, and let $\pi(\mathcal{E}) = \{\pi(\varepsilon) : \varepsilon \in \mathcal{E}\} = \mathcal{E}^A$.

⁶To apply that theorem, in its notation set $m = 1$, $c_1 = 0$, $\mu = \mathbb{P}$, and let $f_1(x) = 0$ for all $x \in \mathcal{X}$ and $f_2(x) = q^\top \Delta\theta(x)$. The critical functions ϕ are our algorithms $a \in \mathcal{A}(\mathcal{X})$.

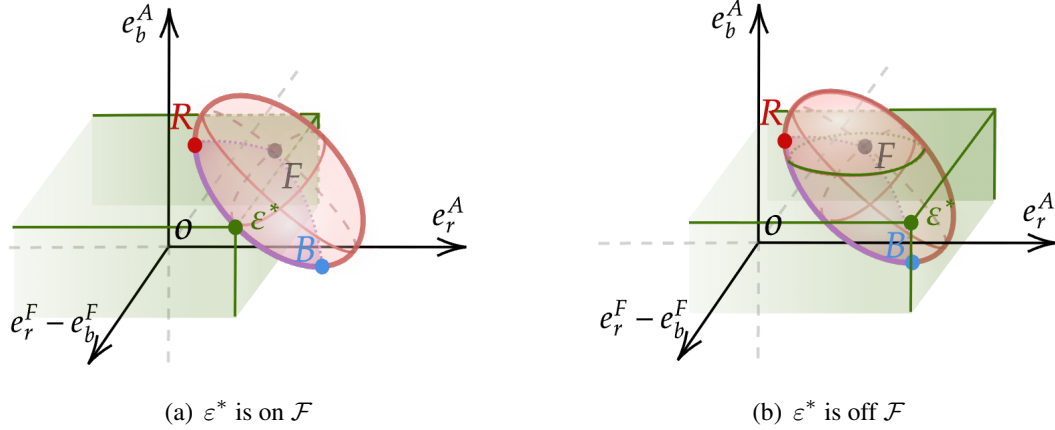


FIGURE 2.—The set $\mathcal{C}(\varepsilon^*)$, which collects all FA improvements relative to ε^* , is the region shaded in green. Panel (a) shows an example where ε^* lies on the frontier and there exists a hyperplane that properly separates $\mathcal{C}(\varepsilon^*)$ and \mathcal{E} , whereas in panel (b) ε^* is not on the frontier and no hyperplane can separate $\mathcal{C}(\varepsilon^*)$ and \mathcal{E} .

THEOREM 2.2: *Let $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*) \in \mathcal{E}$ and define $\tilde{\mathbb{S}}^2 \equiv \{q \in \mathbb{S}^2 : q_1 \geq 0, q_2 \geq 0\}$. Under Assumptions 1(i)-2, the following are equivalent:*

- (i) $\varepsilon^* \in \mathcal{F}$;
- (ii) $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$;
- (iii) $\min_{q \in \tilde{\mathbb{S}}^2} (h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)) = 0$;
- (iv) $\pi(\varepsilon^*) = (\varepsilon_1^*, \varepsilon_2^*) \in \mathcal{F}^A$ and $|\varepsilon_3^*| = d((\varepsilon_1^*, \varepsilon_2^*))$.

An immediate consequence of Theorem 2.2 is that \mathcal{F} captures as many features of the trade-off between accuracy and fairness as \mathcal{F}^A does in LLMO's original contribution, while preserving convexity of the feasible set and yielding a simple characterization of the FA-frontier.⁷ Indeed, using the equivalence (i) \Leftrightarrow (iii) in Theorem 2.2, similarly to Liu and Molinari (2026, Proposition 3.3), we have

$$\mathcal{F} = \left\{ \varepsilon^* \in \mathcal{E} : \min_{q \in \tilde{\mathbb{S}}^2} [h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)] = 0 \right\}. \quad (14)$$

⁷In fact, \mathcal{F} yields a finer characterization than \mathcal{F}^A . Consider algorithms $a, a' \in \mathcal{A}$ such that $\varepsilon := \varepsilon(a) = (0.5, 0.5, 1)$ and $\varepsilon' := \varepsilon(a') = (0.5, 0.5, 0)$. Then $\varepsilon' \succ_{FA} \varepsilon$ but $c' \not\succeq_{FALLMO} c$ for $c = (\varepsilon_1, \varepsilon_2)$ and $c' = (\varepsilon'_1, \varepsilon'_2)$.

When the same loss function is used to measure accuracy and fairness, as in [LLMO](#)'s main text and in [Liu and Molinari \(2026\)](#), it is simple to characterize when the FA-frontier coincides with the Pareto frontier in (8).⁸ When different loss functions are used, such characterization is harder to obtain, and [LLMO](#) provide only a sufficient condition for it. Working with the feasible set $\mathcal{E} \subset \mathbb{R}^3$ and its support function, we are able to provide a simple, necessary and sufficient, testable condition under which $\mathcal{PF} = \mathcal{F}$. The characterization is valuable because when \mathcal{F} does not coincide with \mathcal{PF} , the trade-off between fairness and accuracy becomes substantial: a policymaker who values fairness may choose to deploy an algorithm that is Pareto dominated in accuracy loss, to ameliorate its performance relative to fairness loss. To derive the result, it is useful to distinguish between cases in which \mathcal{E} has kinks, and cases where it does not, through the following condition.

ASSUMPTION 3—No Kinks: *For any $q, v \in \mathbb{S}^2, q \neq v$,*

$$\mathbb{P}(q^\top \Delta \boldsymbol{\theta}(X) > 0, v^\top \Delta \boldsymbol{\theta}(X) < 0) > 0. \quad (15)$$

Assumption 3 is a necessary and sufficient condition under which \mathcal{E} has no kinks (see Supplemental Appendix B.2.3 in [Bontemps et al., 2012](#) and Remark 3.3 in [Liu and Molinari, 2026](#)). It yields injectivity of the support map of \mathcal{E} , $\mathcal{S}_{\mathcal{E}}(q) = \arg \max_{\varepsilon \in \mathcal{E}} q^\top \varepsilon$ (see Theorem 2.1 and the discussion following it).

The next result characterizes when \mathcal{PF} and \mathcal{F} are exactly the same.

THEOREM 2.3: *Under Assumptions 1(i)-2, $\mathcal{PF} = \mathcal{F}$ if and only if for all $\varepsilon^* \in \mathcal{F}$ there exists $\alpha \in [0, 1]$ such that for $q(\alpha) = \frac{[\alpha, 1-\alpha, 0]^\top}{\|[\alpha, 1-\alpha, 0]^\top\|}$, $\varepsilon^* = \arg \min_{\eta \in \mathcal{E}} q(\alpha)^\top \eta$. If Assumption 3 also holds, then $\mathcal{PF} = \mathcal{F}$ if and only if $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$.*

Intuitively, $\mathcal{PF} = \mathcal{F}$ when every point on the FA-frontier is Pareto-optimal. When each point on the boundary of \mathcal{E} has a unique supporting direction (which is guaranteed by Assumption 3), this happens if and only if $\mathcal{PF} \subset \{\varepsilon_3 = 0\} = \{\varepsilon : e_r^F = e_b^F\}$; in this case,

⁸In that case, $\ell^A = \ell^F$ and $e_g^A = e_g^F, g \in \{r, b\}$, so \mathcal{E} can be defined in \mathbb{R}^2 and [LLMO](#) show that the Pareto and the FA frontiers coincide if and only if R and B , defined in (9), lie on opposite sides of the 45° line.

all points along the Pareto frontier are fair and there is no conflict between improving fairness and improving accuracy. On the other hand, when \mathcal{E} has kinks, one may have $\mathcal{PF} = \mathcal{F}$ even though $\mathcal{PF} \not\subseteq \{\varepsilon_3 = 0\}$. The complexity of the geometry of \mathcal{F} in relation to the geometry of \mathcal{E} and the usefulness of our Theorem 2.3 are illustrated in Figure 1. If \mathcal{E} has no kinks and lies completely to one side of the hyperplane $\{\varepsilon_3 = 0\}$ (Panel (a), where $\mathcal{E} \subset \{\varepsilon_3 > 0\} = \{e_r^F > e_b^F\}$) or R and B lie on opposite sides of that hyperplane (Panel (c)), then $\mathcal{PF} \subsetneq \mathcal{F}$ since $\mathcal{PF} \not\subset \{\varepsilon_3 = 0\}$. On the other hand, if \mathcal{E} has kinks properly facing the origin, then it can be that $\mathcal{PF} = \mathcal{F}$ even when \mathcal{E} lies completely to one side of the hyperplane $\{\varepsilon_3 = 0\}$ (Panel (b)). If \mathcal{E} crosses the hyperplane $\{\varepsilon_3 = 0\} = \{e_r^F = e_b^F\}$ with \mathcal{PF} fully contained in one of $\{\varepsilon_3 > 0\}$ or $\{\varepsilon_3 < 0\}$, then $\mathcal{PF} \subsetneq \mathcal{F}$ regardless of the presence or absence of kinks (Panel (d)), see Corollary B.2. If \mathcal{E} has no kinks and $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$, then $\mathcal{PF} = \mathcal{F}$ (Panel (e)), yet having only $\{R, B\} \in \{\varepsilon_3 = 0\}$ but not the entire Pareto frontier may yield $\mathcal{PF} \subsetneq \mathcal{F}$ (Panel (f)). Hence, the characterization in Theorem 2.3 is particularly useful to avoid case-by-case distinctions and pre-tests.

3. SELECTIVELY OBSERVED LABELS

Having fully characterized the FA-frontier in the idealized case when $(Y^*, X, G, Z) \sim \mathbb{P}^*$ is observed, we turn to our goal of providing identification results and inference procedures for \mathcal{F} when only $(Y, X, G, Z) \sim \mathbb{P}$ is observed, with $Y = ZY^*$ the selected labels and Z a selection indicator. This problem has been studied, e.g., by Lakkaraju et al. (2017), Kleinberg et al. (2018), Rambachan et al. (2025), Khan et al. (2025). When evaluating whether algorithmic policies can improve upon human decision-making, selectively observed labels are compounded by the possibility that Z is determined by a decision maker (e.g., a judge) who observes information beyond the recorded covariates made available to the algorithm. In this case, for a specific choice of loss functions, we report partial identification results. We then provide point identification results and statistical inference procedures for the case where label observability is induced by deployment of a decision rule in $\mathcal{A}(\mathcal{X})$, so that Z is determined by an algorithm that uses only the inputs in \mathcal{X} .

3.1. Partial Identification of the FA-Frontier

In this section, we allow Z to depend on variables unavailable to the algorithms in $\mathcal{A}(\mathcal{X})$. This creates substantial challenges for identification and inference, because not only the feasible set \mathcal{E} is partially identified, but even for a fixed algorithm $a^* \in \mathcal{A}(\mathcal{X})$, the vector of expected losses $\varepsilon^* \equiv \varepsilon(a^*) := (e_r^A(a^*), e_b^A(a^*), e_r^F(a^*) - e_b^F(a^*))$ is also only partially identified. Consequently, extra care needs to be taken, when characterizing what can be learned about \mathcal{F} in (14), to correctly couple the possible values of \mathcal{E} and ε^* associated with each candidate distribution for the unobserved labels.

To make progress on this task, we specialize our analysis to the case where Y^* is a binary scalar variable, the accuracy loss measures classification error, $\ell^A(d, y) = \mathbb{1}\{d \neq y\}$, and the fairness loss measures statistical parity, $\ell^F(d, y) = \mathbb{1}\{d = 1\}$. Write $Y = ZY^*$ for the observed binary outcome. In this case the fairness coordinate depends only on the decision rule and the observable distribution of (X, G) , so lack of point identification comes entirely from the selectively observed outcomes entering the two accuracy coordinates.

Denote the unobserved conditional expectation of Y^* when $Z = 0$ by

$$\lambda_g^*(x) \equiv \mathbb{E}^*[Y^* | X = x, G = g, Z = 0], \quad g \in \{r, b\}$$

and $\lambda^*(x) \equiv [\lambda_r^*(x), \lambda_b^*(x)]^\top$. Without restrictions on the selection process, λ^* is unknown. As $Y^* \in \{0, 1\}$, every measurable map $\lambda \equiv (\lambda_r, \lambda_b) : \mathcal{X} \rightarrow [0, 1]^2$ is compatible with \mathbb{P} , the observed law of (Y, X, G, Z) .⁹ Let $\Lambda \equiv \{\lambda : \mathcal{X} \rightarrow [0, 1]^2\}$ denote this class of functions.

We next derive the sharp identification region for $h_{\mathcal{E}}(q)$. To do so, we need to introduce some notation. Define the observable functions

$$A_0(x) \equiv \mathbb{E} \left[\left(\frac{ZY\mathbb{1}\{G=r\}}{\mu_r}, \frac{ZY\mathbb{1}\{G=b\}}{\mu_b}, 0 \right)^\top \middle| X = x \right], \quad (16)$$

$$A_1(x) \equiv \mathbb{E} \left[\left(\frac{(1-ZY)\mathbb{1}\{G=r\}}{\mu_r}, \frac{(1-ZY)\mathbb{1}\{G=b\}}{\mu_b}, \frac{\mathbb{1}\{G=r\}}{\mu_r} - \frac{\mathbb{1}\{G=b\}}{\mu_b} \right)^\top \middle| X = x \right], \quad (17)$$

⁹By “compatible with the observed data” we mean that it yields a law of (Y^*, X, G, Z) that reproduces the observed distribution of (Y, X, G, Z) and satisfies $Y = ZY^*$ almost surely.

$$B(x) \equiv \mathbb{E} \left[\left(\begin{array}{cc} \frac{(1-Z)\mathbb{1}\{G=r\}}{\mu_r} & 0 \\ 0 & \frac{(1-Z)\mathbb{1}\{G=b\}}{\mu_b} \\ 0 & 0 \end{array} \right) \middle| X = x \right]. \quad (18)$$

Lemma A.1 shows that the functions $\theta_d(\cdot)$ in Definition 1 associated with distribution $\lambda \in \Lambda$ satisfy $\theta_0(x; \lambda) = A_0(x) + B(x)\lambda(x)$ and $\theta_1(x; \lambda) = A_1(x) - B(x)\lambda(x)$. Hence, the feasible set associated with distribution λ and its support function are, respectively,

$$\mathcal{E}(\lambda) = \{\mathbb{E}[(1 - a(X))\theta_0(X; \lambda) + a(X)\theta_1(X; \lambda)], a \in \mathcal{A}(\mathcal{X})\}, \quad (19)$$

$$h_{\mathcal{E}}(q; \lambda) = \mathbb{E}[\max\{q^\top \theta_0(X; \lambda), q^\top \theta_1(X; \lambda)\}]. \quad (20)$$

Accordingly, the sharp identified set for $h_{\mathcal{E}}(q)$ in direction q is $\{h_{\mathcal{E}}(q; \lambda) : \lambda \in \Lambda\}$.

For a fixed algorithm a^* , let $A_{d,j}(x)$ denote the j th coordinate of $A_d(x)$, $d \in \{0, 1\}$, in (16)-(17) and define

$$C(q; a^*) \equiv \sum_{j=1}^2 q_j \mathbb{E}[a^*(X)A_{1,j}(X) + (1 - a^*(X))A_{0,j}(X)] + |q_3| |\mathbb{E}[a^*(X)A_{1,3}(X)]|. \quad (21)$$

Then we can express $h_{\mathcal{C}(\varepsilon(a^*; \lambda))}$ in (13) associated with $\lambda \in \Lambda$ as

$$h_{\mathcal{C}(\varepsilon(a^*; \lambda))}(q) = C(q; a^*) + \mathbb{E}^*[(1 - 2a^*(X)) q^\top B(X)\lambda(X)]. \quad (22)$$

We now present an assumption on the observable distribution \mathbb{P} under which the margin condition (Assumption 2) holds for all distributions consistent with the data.

ASSUMPTION 4: *There exists a constant $m \in (0, 1]$ such that, for every $\delta > 0$,*

$$\sup_{q \in \mathbb{S}^2} \mathbb{P}_X \left(\inf_{s \in [0, 1]^2} |q^\top \{A_1(X) - A_0(X) - 2B(X)s\}| \leq \delta \right) \lesssim \delta^m.$$

Under Assumption 4, Theorem 2.2 characterizes frontier membership for a given λ , where crucially we use the same candidate unobserved distribution λ both in obtaining $h_{\mathcal{E}}(\cdot)$ and $h_{\mathcal{C}(\varepsilon^*)}(\cdot)$. Recall $\varepsilon^*(a^*; \lambda)$ denotes the vector of expected losses induced by a^*

under unobserved distribution λ ; then a^* lies on the FA-frontier under λ if and only if

$$\min_{q \in \tilde{\mathbb{S}}^2} [h_{\mathcal{C}(\varepsilon(a^*; \lambda))}(q) + h_{\mathcal{E}}(-q; \lambda)] = 0.$$

The next result shows how to check whether a^* can generate expected losses on the frontier for *some* distribution of the unobserved labels, through a finite dimensional optimization problem that involves functionals of the observed distribution \mathbb{P} .

THEOREM 3.1: *Let Assumptions 1(i) and 4 hold. Suppose $Y^* \in \{0, 1\}$, $\ell^A(d, y) = \mathbb{1}\{d \neq y\}$, and $\ell^F(d, y) = \mathbb{1}\{d = 1\}$. Then a given algorithm a^* generates expected losses on the FA-frontier for some distribution consistent with the observed data if and only if*

$$\min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E} [\max\{J_0(\lambda_1; q, a^*), J_1(\lambda_0; q, a^*)\}] = 0. \quad (23)$$

where $\lambda_0 = (0, 0)^\top$, $\lambda_1 = (1, 1)^\top$, and

$$\begin{aligned} J_0(\lambda; q, a^*) &\equiv C(q; a^*) - q^\top A_0(X) - 2a^*(X)q^\top B(X)\lambda, \\ J_1(\lambda; q, a^*) &\equiv C(q; a^*) - q^\top A_1(X) + 2(1 - a^*(X))q^\top B(X)\lambda. \end{aligned} \quad (24)$$

Theorem 3.1 shows that once the source of partial identification is reduced to the functional λ , the search over the infinite-dimensional class Λ collapses to a finite-dimensional criterion involving only the corner distributions λ_0 and λ_1 .

Our next goal is to characterize the sharp identification region of \mathcal{F} using a finite dimensional optimization problem that does not involve $a \in \mathcal{A}(\mathcal{X})$ and $\lambda \in \Lambda$. Recall $\mathcal{E}(\lambda)$ in (19) and define the FA frontier associated with distribution λ and the *FA-frontier envelope* as

$$\mathcal{F}(\lambda) \equiv \{\varepsilon \in \mathcal{E}(\lambda) : \nexists \varepsilon' \in \mathcal{E}(\lambda) \text{ such that } \varepsilon' \succ_{FA} \varepsilon\}, \quad (25)$$

$$\mathcal{F}^\exists \equiv \bigcup_{\lambda \in \Lambda} \mathcal{F}(\lambda). \quad (26)$$

For a given $q \in \tilde{\mathbb{S}}^2$, collect all expected loss vectors achievable by an algorithm that takes the threshold form specified in Corollary 2.1 for some distribution λ and thus are the boundary

points of $\mathcal{E}(\lambda)$, in the following set¹⁰

$$\mathcal{M}_q \equiv \left\{ \mathbb{E} [a(X)\boldsymbol{\theta}_1(X; \lambda) + (1 - a(X))\boldsymbol{\theta}_0(X; \lambda)] : \right. \\ \left. a(x) \in \arg \min_{d \in \{0,1\}} q^\top \boldsymbol{\theta}_d(x; \lambda), \mathbb{P}_X\text{-a.s.}, \lambda \in \Lambda \right\}. \quad (27)$$

In Lemma B.5 we show that for every $q \in \tilde{\mathbb{S}}^2$, the set \mathcal{M}_q is nonempty, compact, and convex. In Lemma B.7 we show that its support function, $h_{\mathcal{M}_q}(v), v \in \mathbb{S}^2$, is given by

$$h_{\mathcal{M}_q}(v) = \mathbb{E} [\max\{\kappa_0(v, q; X), \kappa_1(v, q; X)\}],$$

where $\kappa_0(v, q; x)$ and $\kappa_1(v, q; x)$, formally defined in (72)-(73), are known functions of $A_0(x), A_1(x), B(x)$. We then have the following characterization of \mathcal{F}^\exists :

THEOREM 3.2: *Let Assumptions 1 and 4 hold. Then*

$$\mathcal{F}^\exists = \bigcup_{q \in \tilde{\mathbb{S}}^2} (\mathcal{M}_q \cap \{\varepsilon \in \mathbb{R}^3 : q_3 \varepsilon_3 \geq 0\}). \quad (28)$$

For $\varepsilon \in \mathbb{R}^3$, define

$$T_{\text{env}}(\varepsilon) \equiv \inf_{q \in \tilde{\mathbb{S}}^2 : q_3 \varepsilon_3 \geq 0} [\sup_{v \in \mathbb{S}^2} \{v^\top \varepsilon - h_{\mathcal{M}_q}(v)\}]_+. \quad (29)$$

Then, for every $\varepsilon \in \mathbb{R}^3$, $\varepsilon \in \mathcal{F}^\exists \iff T_{\text{env}}(\varepsilon) = 0$.

REMARK 3.1: In work in progress, we obtain a debiased machine learning estimator for $J_d(\lambda; q, a^*)$, $d \in \{0, 1\}$, and we develop an inference procedure to test the hypothesis that a given algorithm generates expected losses on the FA-frontier. We also explore extending Theorem 3.1 to other loss functions.

¹⁰The minimization in the definition of \mathcal{M}_q is a sign convention that we use because the first two coordinates are losses (we could alternatively take $\arg \max_d -q^\top \boldsymbol{\theta}_d(x; \lambda)$ for $q \in \tilde{\mathbb{S}}^2$).

3.2. Point Identification Under Missing at Random Assumptions

While the distribution \mathbb{P} of (Y, G, X, Z) is point identified by the observed data, in the absence of additional assumptions the distribution \mathbb{P}^* of (Y^*, G, X, Z) is not. The previous section shows that, as a consequence, the support function $h_{\mathcal{E}}(q)$ and $\varepsilon^*(a)$, $a \in \mathcal{A}(\mathcal{X})$, are not point identified either, as the ideal labels and their conditional expectations (\mathbf{L}_d and $\theta_d(X)$, respectively) are a function of Y^* and \mathbb{P}^* .

We leverage the fact that all algorithms in $\mathcal{A}(\mathcal{X})$ can only use the covariates X for training (and no other unobserved variables) to argue plausibility of a selection on observables assumption that we use to point identify the distribution \mathbb{P}^* .

ASSUMPTION 5—Ideal labels missing at random (MAR): $(Y^*, G) \perp Z|X$ and $\pi(X) \equiv \mathbb{E}[Z|X] \in (0, 1)$ *a.s.*

Assumption 5 may look stronger than the typical missing at random (unconfoundedness or selection on observables) assumption, which might instead require $Y^* \perp Z|X, G$. Our approach remains valid under this weaker condition, but here we work with Assumption 5 because it yields lighter notation and simpler expressions. We argue that, in many settings of interest, the additional requirement $G \perp Z|X$ is a natural consequence of how Z is generated. Specifically, suppose that $Z = D$, where $D \in \{0, 1\}$ is the binary decision an individual receives (e.g., granted or denied a loan; granted or denied bail), and that this decision is produced by a deployed decision rule in $\mathcal{A}(\mathcal{X})$ that takes X as input and, following regulations, does not use G or apply any group-dependent post-processing. Then, for all $g \in \{r, b\}$, $\mathbb{P}(D = 1|X, G = g) = \mathbb{P}(D = 1|X) = a(X)$, so that $G \perp Z|X$ holds. Combined with the standard selection on observables condition, $Y^* \perp Z|X, G$, this yields Assumption 5. The more restrictive part of Assumption 5 is therefore the requirement $Y^* \perp Z|X, G$. This condition is plausible when the assignment mechanism that determines Z depends on recorded variables capturing the main drivers of both selection and outcomes, and when discretion at the decision-making stage is limited or can be accounted for. In such settings, conditioning on (X, G) may reasonably approximate conditioning on the information that drives selection, so that Z does not convey additional information

about Y^* beyond (X, G) . When instead the assignment process exploits unrecorded “soft information” (e.g., interviews, free text, documents, or unlogged history) or when discretion is systematically related to expected outcomes, $Y^* \perp Z | X, G$ may fail, motivating our partial identification analysis in Section 3.1.

The next lemma shows that, under Assumption 5 the support function of \mathcal{E} is point identified through a standard inverse propensity weighting identity:

LEMMA 3.1: *Let Assumptions 1(i) and 5 hold. Then for $\Delta \mathbf{L} \equiv \mathbf{L}_1 - \mathbf{L}_0$,*

$$\boldsymbol{\theta}_d(X) = \mathbb{E} \left[\frac{Z \mathbf{L}_d}{\pi(X)} \middle| X \right], \quad (30)$$

$$h_{\mathcal{E}}(q) = \mathbb{E} \left[\frac{q^\top \mathbf{L}_0 Z}{\pi(X)} + \left(\frac{q^\top \Delta \mathbf{L} Z}{\pi(X)} \right) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\} \right], \quad \text{for all } q \in \mathbb{S}^2, \quad (31)$$

$$e_g^\iota(a) = \mathbb{E} \left[\frac{a(X) L_1^{g,\iota} Z}{\pi(X)} + \frac{(1-a(X)) L_0^{g,\iota} Z}{\pi(X)} \right], \quad g = r, b \text{ and } \iota = A, F. \quad (32)$$

4. A DEBIASED MACHINE LEARNING ESTIMATOR AND ITS ASYMPTOTIC DISTRIBUTION

4.1. The DML Estimator

Our goal is to estimate the support function $h_{\mathcal{E}}(q)$ using a random sample $\{(Y_i, G_i, X_i, Z_i) : 1 \leq i \leq n\}$ drawn from the distribution \mathbb{P} . By Lemma 3.1, $h_{\mathcal{E}}(q)$ is identified by (31), which involves nuisance functions $\pi(X)$ and $\Delta \boldsymbol{\theta}(X)$ that need to be estimated in the first step. As the ideal labels \mathbf{L} depend on $\boldsymbol{\mu} \equiv [\mu_r, \mu_b]^\top$ (see (1)-(2)), we also need to estimate $\boldsymbol{\mu}$. In order to build an estimator for $h_{\mathcal{E}}(q)$ that is locally robust to the first-step estimation error, we propose using the debiased machine learning (DML) approach. Consider the estimand

$$h_{\mathcal{E}}(q) = \mathbb{E} [\zeta_i(q; \mathbf{L}, \boldsymbol{\eta})], \quad (33)$$

where $\boldsymbol{\eta} \equiv (\Delta \boldsymbol{\theta}, \pi, \boldsymbol{\theta}_0)$ is a vector of nuisance functions and

$$\zeta_i(q; \mathbf{L}, \boldsymbol{\eta}) \equiv \frac{q^\top \mathbf{L}_{0,i} Z_i}{\pi(X_i)} + \left(\frac{q^\top \Delta \mathbf{L}_i Z_i}{\pi(X_i)} \right) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} + \alpha^h(q; X_i) \left(1 - \frac{Z_i}{\pi(X_i)} \right), \quad (34)$$

$$\alpha^h(q; X) \equiv q^\top \boldsymbol{\theta}_0(X) + q^\top \Delta \boldsymbol{\theta}(X) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}. \quad (35)$$

The local effect of the first-step function π on the original estimand $h_{\mathcal{E}}(q)$ in (33) is accounted for by $\alpha^h(q; X_i) \left(1 - \frac{Z_i}{\pi(X_i)} \right)$. As we show in the proof of Theorem 4.1, under

Assumptions 6 and 7 presented below, it is not necessary to account for the local effect of the first-step function $\Delta\theta$ on $h_{\mathcal{E}}(q)$. Finally, we note that the estimand in (33) is based on \mathbf{L} , which by (1) relies on the population means $\mu_g, g \in \{r, b\}$.

Following the DML approach (e.g., Chernozhukov et al., 2018, Velez, 2024), we first randomly split the indices $[n] \equiv \{1, \dots, n\}$ into K equal-sized folds \mathcal{I}_k , i.e., $\cup_{k=1}^K \mathcal{I}_k = [n]$. We denote by n_k the size of \mathcal{I}_k .¹¹ We then estimate $\hat{\boldsymbol{\eta}}(X_i)$ as $\hat{\boldsymbol{\eta}}_k(X_i)$ for $i \in \mathcal{I}_k$ and $k = 1, \dots, K$, where $\hat{\boldsymbol{\eta}}_k(\cdot)$ is estimated using all data except the portion with indices in fold \mathcal{I}_k . Finally, the first-stage estimators for the nuisance functions are used to form a second-stage estimator for $h_{\mathcal{E}}(q)$ based on the expression of the estimand defined in (33):

$$\hat{h}_{\mathcal{E}}(q; \hat{\mathbf{L}}, \hat{\boldsymbol{\eta}}) \equiv \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \zeta_i(q; \hat{\mathbf{L}}, \hat{\boldsymbol{\eta}}_k) \right) = \frac{1}{n} \sum_{i=1}^n \zeta_i(q; \hat{\mathbf{L}}, \hat{\boldsymbol{\eta}}), \quad (36)$$

where $\hat{\mathbf{L}}$ is defined as in (1)-(2) with μ_g replaced by $\hat{\mu}_g \equiv n^{-1} \sum_{i=1}^n \mathbf{1}\{G_i = g\}$.

We also propose a DML estimator for $\varepsilon^* = \varepsilon(a^*) \in \mathcal{E}$. By (32) in Lemma 3.1 and Definition 3, we can write $\varepsilon^* = \mathbb{E}[\xi_i(a^*; \mathbf{L}, \boldsymbol{\eta})]$, where for a given algorithm $a \in \mathcal{A}$,

$$\xi_i(a; \mathbf{L}, \boldsymbol{\eta}) \equiv (\mathbf{L}_{0,i} + a(X_i)\Delta\mathbf{L}_i) \frac{Z_i}{\pi(X_i)} + \boldsymbol{\alpha}^e(a; X_i) \left(1 - \frac{Z_i}{\pi(X_i)}\right), \quad (37)$$

$$\boldsymbol{\alpha}^e(a; X) \equiv \boldsymbol{\theta}_0(X) + a(X)\Delta\boldsymbol{\theta}(X), \quad (38)$$

and again the local effect of the first-step function π on the original estimand ε^* is accounted for by $\boldsymbol{\alpha}^e(X_i) \left(1 - \frac{Z_i}{\pi(X_i)}\right)$. Using this notation, we propose to estimate ε^* by

$$\hat{\varepsilon}^* \equiv \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \xi_i(a^*; \hat{\mathbf{L}}, \hat{\boldsymbol{\eta}}_k) \right) = \frac{1}{n} \sum_{i=1}^n \xi_i(a^*; \hat{\mathbf{L}}, \hat{\boldsymbol{\eta}}), \quad (39)$$

where $\hat{\mathbf{L}}$ stacks the feasible labels and $\{\hat{\boldsymbol{\eta}}_k\}_{k=1}^K$ are estimators for the nuisance functions using cross-fitting, as we described above.

¹¹When n is not divisible by K , the number of observations in some folds will be $\lfloor n/K \rfloor$ while in others $\lfloor n/K \rfloor + 1$, where $\lfloor n/K \rfloor$ is the greatest integer less than or equal to n/K .

4.2. Asymptotic Distribution of the DML Estimator

For the nuisance functions' estimation error to be asymptotically negligible for the DML estimator, the nuisance estimators must satisfy suitable regularity conditions set out next.

ASSUMPTION 6: *There is a known partition of X , $X = (X_1, X_2)$, with $X_1 \in \mathcal{X}_1 \subset \mathbb{R}^p$ and $X_2 \in \mathcal{X}_2 \subset \mathbb{R}^{d_X - p}$ such that:*

- (a) *For every $\delta > 0$, $\sup_{q \in \mathbb{S}^2} \mathbb{P}(|q^\top \Delta \boldsymbol{\theta}(X)| \leq \delta | X_2) \lesssim \delta$.*
- (b) *There is a constant $c > 0$ such that $\inf_{q \in \mathbb{S}^2} \text{Var}[|q^\top \Delta \boldsymbol{\theta}(X)| | X_2] \geq c$.*
- (c) *There are unknown functions $\nu : \mathcal{X}_1 \rightarrow \mathbb{R}^{d_\nu}$ and $\gamma : \mathcal{X}_2 \rightarrow \mathbb{R}^{d_\gamma}$, and a known function $F : \mathbb{R}^{d_\nu} \times \mathbb{R}^{d_\gamma} \times \mathbb{R}^2 \rightarrow [\epsilon, 1 - \epsilon] \times \mathbb{R}^3$, such that:*

$$(\pi(X), \Delta \boldsymbol{\theta}(X)) = F(\nu(X_1), \gamma(X_2), \mu)$$

and the partial derivatives of F are uniformly bounded for any $\nu \in \mathbb{R}^{d_\nu}$, $\gamma \in \mathbb{R}^{d_\gamma}$, and $\mu \in \mathbb{B}_\epsilon(\mu_r, \mu_b) \equiv \{\mu \in \mathbb{R}^2 : \|\mu - (\mu_r, \mu_b)\| < \epsilon\}$, that is

$$\sup_{\nu \in \mathbb{R}^{d_\nu}} \sup_{\gamma \in \mathbb{R}^{d_\gamma}} \sup_{\mu \in \mathbb{B}_\epsilon(\mu_r, \mu_b)} \|DF(\nu, \gamma, \mu)\| < \infty.$$

A sufficient condition for Assumption 6 is given in [Liu and Molinari \(2026, Assumption 3\)](#), where a partially linear structure is assumed for $\Delta \boldsymbol{\theta}$.

Let $\mathfrak{F}_k \equiv \sigma((Y_i, G_i, X_i, Z_i) : i \notin \mathcal{I}_k)$ be the σ -algebra generated by the data used for constructing the estimator $\widehat{\boldsymbol{\eta}}_k$ for the nuisance $\boldsymbol{\eta} \equiv (\Delta \boldsymbol{\theta}, \pi, \boldsymbol{\theta}_0)$. Let $\widehat{\boldsymbol{\mu}}_k$ be an estimator for $\boldsymbol{\mu} \equiv [\mu_r, \mu_b]^\top$ that uses only the data outside fold \mathcal{I}_k (similar to how we define $\widehat{\boldsymbol{\eta}}_k$):

$$\widehat{\boldsymbol{\mu}}_k = \left(\frac{1}{n - n_k} \sum_{i \notin \mathcal{I}_k} \mathbb{1}\{G_i = r\}, \frac{1}{n - n_k} \sum_{i \notin \mathcal{I}_k} \mathbb{1}\{G_i = b\} \right)^\top.$$

We assume that the estimators for $\Delta \boldsymbol{\theta}(X)$ and $\pi(X)$ share the same structure as in Assumption 6. Furthermore, we treat $\nu(X_1)$ as a low-dimensional nonparametric component that can be estimated uniformly well over \mathcal{X}_1 , whereas $\gamma(X_2)$ captures a high-dimensional component that can be estimated sufficiently well in the sense that its mean squared error converges to zero sufficiently fast. We assume that $\boldsymbol{\theta}_0$ can be estimated sufficiently well.

ASSUMPTION 7: *There exist estimators $\widehat{\nu}_k$, $\widehat{\gamma}_k$, and $\widehat{\boldsymbol{\theta}}_{0,k}$ for $k = 1, \dots, K$ such that*

- (a) $(\widehat{\pi}_k(X), \Delta \widehat{\boldsymbol{\theta}}_k(X)) = F(\widehat{\nu}_k(X_1), \widehat{\gamma}_k(X_2), \widehat{\boldsymbol{\mu}}_k)$.
- (b) $\sup_{x_1 \in \mathcal{X}_1} \|\widehat{\nu}_k(x_1) - \nu(x_1)\| = o_p(n^{-1/4})$.
- (c) $\mathbb{E} [\|\widehat{\gamma}_k(X_2) - \gamma(X_2)\|^2 | \mathfrak{F}_k] = o_p(n^{-1/2})$.
- (d) $\mathbb{E} [\|\widehat{\boldsymbol{\theta}}_{0,k}(X) - \boldsymbol{\theta}_0(X)\|^2 | \mathfrak{F}_k] = o_p(n^{-1/2})$.

Example machine learning methods that, under suitable conditions, are $n^{1/4}$ -consistent in the L^2 -norm include ℓ_1 -penalized methods, boosting, and neural nets (see, e.g., [Chernozhukov et al., 2018](#), and references therein, for a discussion of machine learners compatible with the rate requirement).

Similarly to [Liu and Molinari \(2026, Thm. 4.1\)](#), our main asymptotic result shows that $\widehat{h}_{\mathcal{E}}(q; \widehat{\mathbf{L}}, \widehat{\boldsymbol{\eta}})$ converges to a Gaussian process uniformly in $q \in \mathbb{S}^2$, where the score $\zeta_i(q; \mathbf{L}, \boldsymbol{\eta})$ in (34) is the influence function that governs the part of the limit distribution of $\widehat{h}_{\mathcal{E}}(q; \widehat{\mathbf{L}}, \widehat{\boldsymbol{\eta}})$ due to the uncertainty in $\widehat{\boldsymbol{\eta}}$, and the remaining part is attributed to estimating $\boldsymbol{\mu}$:

THEOREM 4.1: *Let Assumptions 1-3 and 5-7 hold. Then,*

$$\sqrt{n} \left(\widehat{h}_{\mathcal{E}}(q; \widehat{\mathbf{L}}, \widehat{\boldsymbol{\eta}}) - h_{\mathcal{E}}(q) \right) \Rightarrow \mathbb{G}[\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta})] \quad \text{in } \ell^\infty(\mathbb{S}^2),$$

where

$$\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) \equiv \zeta_i(q; \mathbf{L}, \boldsymbol{\eta}) + \sum_{g \in \{r, b\}} \Gamma_g^h(q) \left(1 - \frac{\mathbb{1}\{G_i = g\}}{\mu_g} \right), \quad (40)$$

$$\text{and } \Gamma_g^h(q) \equiv \mathbb{E} \left[\mathbb{1}\{G_i = g\} (q^\top \mathbf{L}_{0,i} + q^\top \Delta \mathbf{L}_i \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}) \frac{Z_i}{\pi(X_i)} \right]$$

REMARK 4.1: One could alternatively construct a DML estimator of $h_{\mathcal{E}}(q)$ using $\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta})$ in (40) instead of $\zeta_i(q; \mathbf{L}, \boldsymbol{\eta})$ in (34). Doing so would be valid because $\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta})$ satisfies the Neyman orthogonality condition with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$. In practice, when $\widehat{\boldsymbol{\eta}}$ is obtained via cross-fitting and $\boldsymbol{\mu}$ is estimated by sample means, $\widehat{h}_{\mathcal{E}}(q; \widehat{\mathbf{L}}, \widehat{\boldsymbol{\eta}})$ is numerically equivalent to $\frac{1}{n} \sum_{i=1}^n \zeta_i^*(q; \widehat{\mathbf{L}}, \widehat{\boldsymbol{\eta}})$. Hence, our proposed estimator implicitly accounts for first-step estimation error in $\widehat{\boldsymbol{\mu}}$.

5. TESTING PROCEDURES

We provide test statistics and their asymptotic distribution to test whether (i) the FA-frontier coincides with the Pareto frontier; and (ii) a less discriminatory alternative to a given algorithm exists.

5.1. Testing $\mathcal{PF} = \mathcal{F}$

Theorem 2.3 provides a necessary and sufficient condition for $\mathcal{PF} = \mathcal{F}$. Here we put forward an equivalent characterization based on the support function process, which is amenable to testing the null hypothesis $\mathcal{PF} = \mathcal{F}$ against the alternative $\mathcal{PF} \neq \mathcal{F}$.

PROPOSITION 5.1: *Let Assumptions 1-2 hold. For C a constant pinned down by Assumption 1, $\mathcal{E} \subset \mathbb{B}_C$, with \mathbb{B}_C the ball in \mathbb{R}^3 of radius C . Define*

$$T^{LDA}(\varepsilon) \equiv \left[\sup_{q \in \mathbb{S}^2} \{q^\top \varepsilon - h_{\mathcal{E}}(q)\} \right]_+ + \left[\inf_{q \in \tilde{\mathbb{S}}^2} \{h_{\mathcal{C}(\varepsilon)}(q) + h_{\mathcal{E}}(-q)\} \right]_+,$$

$$\psi_{\mathcal{PF}}(\varepsilon) \equiv \left[\inf_{\alpha \in [0,1]} \{q(\alpha)^\top \varepsilon + h_{\mathcal{E}}(-q(\alpha))\} \right]_+, \quad q(\alpha) \equiv \frac{[\alpha, 1-\alpha, 0]^\top}{\|[\alpha, 1-\alpha, 0]^\top\|}.$$

Then $\mathcal{PF} = \mathcal{F} \iff \sup_{\varepsilon \in \mathbb{B}_C: T^{LDA}(\varepsilon)=0} \psi_{\mathcal{PF}}(\varepsilon) = 0$.

If Assumption 3 also holds, fix any $\bar{s} > 0$ and define

$$\Delta_{\mathcal{PF}}^h(\bar{s}) \equiv \sup_{\alpha \in [0,1]} \left[h_{\mathcal{E}}(-q(\alpha)) - \inf_{s \in [-\bar{s}, \bar{s}]} \sqrt{1+s^2} h_{\mathcal{E}}\left(\frac{-q(\alpha) + su_3}{\sqrt{1+s^2}}\right) \right], \quad (41)$$

where $u_3 = [0, 0, 1]^\top$. Then $\mathcal{PF} = \mathcal{F} \iff \Delta_{\mathcal{PF}}^h(\bar{s}) = 0$.

Eq. (41) translates the geometric condition $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$ in Theorem 2.3 into a criterion involving only the support function. For each Pareto-supporting direction $-q(\alpha)$, the expression inside the supremum checks whether $-q(\alpha)$ already minimizes the support value over fairness-direction tilts $\{-q(\alpha) + su_3 : s \in [-\bar{s}, \bar{s}]\}$. In the proof of Proposition 5.1, we show that the derivative of this tilted support function at $s = 0$ equals the fairness coordinate of the Pareto point; by convexity, $s = 0$ minimizes the support value over $[-\bar{s}, \bar{s}]$ if and only if this derivative—and thus the fairness coordinate—is zero. Hence, the

criterion in (41) is zero exactly when *all* Pareto points have zero fairness disparity, i.e. $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$, which is equivalent to $\mathcal{PF} = \mathcal{F}$ by Theorem 2.3.

REMARK 5.1: In work in progress, we derive the limit distribution of sample analogs of $\psi_{\mathcal{PF}}(\varepsilon)$ and $\Delta_{\mathcal{PF}}^h$, and valid bootstrap critical values.

5.2. Assessing the Existence of a Less Discriminatory Alternative

When evaluating whether there exists a less discriminatory alternative (LDA) to a given algorithm, regulators and policymakers need to confront the fact that they can only rely on finite data to do so. Concretely, given an algorithm $a^* \in \mathcal{A}(\mathcal{X})$, we call another algorithm $\tilde{a} \in \mathcal{A}$ an LDA if it yields expected losses that are at least as accurate as those associated with a^* for both groups, and at least as fair, with one of these inequalities strict. In terms of the FA-dominance notion and FA-frontier reported in Definition 5, the absence of an LDA is equivalent to the fact that a^* yields expected losses with

$$\varepsilon^* \equiv \varepsilon(a^*) = (e_r^A(a^*), e_b^A(a^*), e_r^F(a^*) - e_b^F(a^*)) \in \mathcal{F}. \quad (42)$$

By Theorem 2.2, (42) and (43) below are equivalent, hence we can use the latter to test the hypothesis that no LDA exists against the alternative that it does, as follows:

$$H_0 : \min_{q \in \mathbb{S}^2} [h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)] = 0, \quad (43)$$

$$H_A : \min_{q \in \mathbb{S}^2} [h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)] > 0, \quad (44)$$

where $h_{\mathcal{C}(\varepsilon^*)}(q)$ is defined in (13). We refer to (43) as the LDA hypothesis.

We propose to reject the LDA hypothesis for large values of the following test statistic:

$$T_n^{\text{LDA}} \equiv \sqrt{n} \left(\left[\max_{q \in \mathbb{S}^2} (q^\top \widehat{\varepsilon}^\omega - \widehat{h}_{\mathcal{E}}^\omega(q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}})) \right]_+ + \left[\min_{q \in \mathbb{S}^2} (h_{\mathcal{C}(\widehat{\varepsilon}^\omega)}(q) + \widehat{h}_{\mathcal{E}}^\omega(-q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}})) \right]_+ \right), \quad (45)$$

where $\widehat{h}_{\mathcal{E}}^\omega$ is a weighted estimator for $h_{\mathcal{E}}$ defined in (55) in Appendix A.3 and $\widehat{\varepsilon}^\omega$ is a weighted estimator for ε^* defined in (83) in Appendix B.3.2. The former uses weights $\omega_i^h =$

$1 - \Xi_i/2$ and the latter uses weights $\omega_i^\varepsilon = 1 + \Xi_i/2$, $i = 1, \dots, n$, with Ξ_i a Rademacher random variable taking values in $\{-1, 1\}$ with uniform probability and independent of the sample data and of the first stage training split. The test statistic in (45) was proposed in [Liu and Molinari \(2026, Section 6.2\)](#). The reason why it resorts to using reweighted estimators for $h_{\mathcal{E}}(q)$ and ε^* is that under the null hypothesis that $\varepsilon(a^*) \in \mathcal{F}$, [Corollary 2.1](#) yields that there exists a direction vector $q^* \in \tilde{\mathbb{S}}$ such that $a^*(X) = \mathbb{1}\{q^{*\top} \Delta \boldsymbol{\theta}(X) > 0\}$, \mathbb{P}_X^* -a.s. As a consequence, one can show that under [Assumption 3](#), for ε^* a point at which $h_{\mathcal{C}(\varepsilon^*)}(q^*)$ is locally differentiable, the limit distribution of a the test statistic like the one in (45) but with $\omega_i^h = \omega_i^\varepsilon = 1$ for all $i = 1, \dots, n$, has a degenerate limit distribution. [Liu and Molinari \(2026\)](#) show that this test statistic with degenerate limit distribution asymptotically rejects a true null with probability zero, and a false null under *local* $(1/\sqrt{n})$ alternatives motivated by practical significance criteria with probability one (for a discussion of “practical significance” see [Blattner and Spiess, 2022, Sec. 6](#)). The use of Rademacher weights regularizes the limit distribution of the test statistic, trading asymptotic size α for the null and local asymptotic unbiasedness at local alternatives closer to the null than $1/\sqrt{n}$ times the practical significance threshold, for power above nominal size instead of one just above that threshold.

Our contribution is deriving the test’s asymptotic distribution as well as a valid bootstrap procedure to estimate the associated critical values, in the presence of selectively observed labels. We consider a bootstrap-based critical value $\hat{c}_{1-\alpha}$ defined in [Procedure 1](#) below. To simplify notation, let $\boldsymbol{\beta} \equiv (h_{\mathcal{E}}(q), \varepsilon^*) \in \ell^\infty(\mathbb{S}^2) \times \mathbb{R}^3$ denote the vector of population values and let $\hat{\boldsymbol{\beta}}^\omega \equiv (\hat{h}_{\mathcal{E}}^\omega(q), \hat{\varepsilon}^\omega) \in \ell^\infty(\mathbb{S}^2) \times \mathbb{R}^3$ denote the vector of their estimators. We can write $T_n^{\text{LDA}} = \sqrt{n} \left(\phi(\hat{\boldsymbol{\beta}}^\omega) - \phi(\boldsymbol{\beta}) \right)$, where ϕ is defined in [\(59\)](#) in [Appendix A](#).

PROCEDURE 1—Bayesian Bootstrap for the Quantiles of T_n^{LDA} :

1. Draw $\{W_i\}_{i=1}^n$ *i.i.d.* from the exponential distribution with mean 1 independent of the sample $\{(Y_i, G_i, X_i, Z_i)\}_{i=1}^n$ and the weights $\{\Xi_i\}_{i=1}^n$. Let the bootstrap analogue of $\hat{\boldsymbol{\beta}}^\omega$ be $\tilde{\boldsymbol{\beta}}^\omega \equiv (\tilde{h}_{\mathcal{E}}^\omega, \tilde{\varepsilon}^\omega)$, where $\tilde{h}_{\mathcal{E}}^\omega(q; \tilde{\mathbf{L}}^h, \tilde{\boldsymbol{\eta}}) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{W_i \omega_i^h}{W^h} \right) \zeta_i(q; \tilde{\mathbf{L}}^h, \tilde{\boldsymbol{\eta}})$, $\tilde{\varepsilon}^\omega \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{W_i \omega_i^\varepsilon}{W^\varepsilon} \right) \xi_i(a; \tilde{\mathbf{L}}^\varepsilon, \tilde{\boldsymbol{\eta}})$, $\tilde{\mathbf{L}}^h$ and $\tilde{\mathbf{L}}^\varepsilon$ are defined as in [\(1\)-\(2\)](#) with μ_g replaced

by $\tilde{\mu}_g^h = n^{-1} \sum_{i=1}^n \left(\frac{W_i \omega_i^h}{\bar{W}^h} \right) \mathbb{1}\{G_i = g\}$ and $\tilde{\mu}_g^\varepsilon = n^{-1} \sum_{i=1}^n \left(\frac{W_i \omega_i^\varepsilon}{\bar{W}^\varepsilon} \right) \mathbb{1}\{G_i = g\}$, respectively, and $\bar{W}^h = n^{-1} \sum_{i=1}^n W_i \omega_i^h$ and $\bar{W}^\varepsilon = n^{-1} \sum_{i=1}^n W_i \omega_i^\varepsilon$.

2. Numerically approximate $\phi'_\beta(\cdot)$, the directional derivative of $\phi(\cdot)$ at β , by

$$\hat{\phi}'_\beta(\ddot{\beta}) = \frac{1}{s_n} \left(\phi \left(\hat{\beta}^\omega + s_n(\ddot{\beta}) \right) - \phi \left(\hat{\beta}^\omega \right) \right),$$

where $\ddot{\beta} \in \ell^\infty(\mathbb{S}^2) \times \mathbb{R}^3$ is a candidate direction at which we evaluate $\phi'_\beta(\cdot)$ and s_n is a vanishing sequence of step sizes such that $\sqrt{n}s_n \rightarrow \infty$.

3. Obtain $\hat{\phi}'_\beta(\sqrt{n}\{\tilde{\beta}^\omega - \hat{\beta}^\omega\})$ and calculate

$$\hat{c}_{1-\alpha} \equiv \inf \left\{ c : \mathbb{P} \left(\hat{\phi}'_\beta(\sqrt{n}\{\tilde{\beta}^\omega - \hat{\beta}^\omega\}) \leq c \mid \{(Y_i, G_i, X_i, Z_i, \Xi_i)\}_{i=1}^n \right) \geq 1 - \alpha \right\}.$$

Let us define the LDA test by $\varphi_n^{\text{LDA}} = \mathbb{1}\{T_n^{\text{LDA}} > \hat{c}_{1-\alpha+\kappa} + \kappa\}$ for a given significance level $\alpha \in (0, 1)$, where $\kappa > 0$ is an arbitrarily small positive constant. The constant κ can be taken equal to 10^{-6} as in [Andrews and Shi \(2013\)](#). We can take $\kappa = 0$ when the asymptotic distribution of T_n^{LDA} is continuous at its $(1 - \alpha)$ -quantile; $\kappa > 0$ appears in the critical value used in the LDA test to handle the technicalities arising from possible discontinuity points of the asymptotic distribution of T_n^{LDA} .

The next result establishes the consistency of the Bayesian bootstrap outlined in [Procedure 1](#) and guarantees the asymptotic correct size of the proposed test φ_n^{LDA} .

THEOREM 5.1: *Let Assumptions 1–7 hold. Then, the Bayesian bootstrap outlined in [Procedure 1](#) is consistent. Furthermore, if the LDA hypothesis holds, the test φ_n^{LDA} has asymptotically correct size, $\limsup_{n \rightarrow \infty} \mathbb{E}[\varphi_n^{\text{LDA}}] \leq \alpha$.*

Given [Theorem 5.1](#), we can build an asymptotically valid confidence set for \mathcal{F} by test inversion, similarly to [Liu and Molinari \(2026, Proposition 5.1\)](#).

6. CONCLUSIONS

In this paper, we consider the problem of identification and inference for the FA-frontier put forward in [LLMO](#), when labels (outcomes) are selectively observed. We build on the methodology developed by [Liu and Molinari \(2026\)](#), who assume away the selective labels

problem to focus on deriving moment-inequality representations for the frontier. Selective labels challenge one’s ability to (point) identify features of the conditional distribution of the true labels, including the FA-frontier. Naively computed accuracy metrics using only observed outcomes may misrepresent true performance, potentially overstating accuracy where it matters most and distorting conclusions on whether LDAs exist.

Under the assumption that labels are missing at random conditional on a rich set of observed covariates, we obtain point identification of the FA-frontier using inverse propensity score weighting and derive the asymptotic distribution of the test statistic that we use to characterize whether an algorithm is on the frontier. The inference method that we propose is based on a debiased machine learning estimator, which is particularly appropriate in this context, where high-dimensional administrative data enable flexible nonparametric modeling of the selection process and hence make the missing at random assumption more credible, but formal inference with valid confidence statements remains necessary for legally defensible determinations, making the use of methods that control the bias of the first-step nonparametric estimation crucial.

When the selection process is left completely unrestricted (hence, we dispense with the missing at random assumption) we provide a tractable characterization of the sharp identification region for the FA-frontier, for a specific class of loss functions. In work in progress, we derive a DML estimator and its asymptotic theory (along with valid bootstrap methods) to build a confidence set for the partially identified FA-frontier and to test the hypothesis that no LDA exists to a given algorithm.

REFERENCES

- ADLER, ROBERT J. AND JONATHAN E. TAYLOR (2007): *Random Fields and Geometry*, Springer Monographs in Mathematics, New York: Springer. [47]
- ANDREWS, DONALD W. K. (1994): “Chapter 37 Empirical process methods in econometrics,” Elsevier, vol. 4 of *Handbook of Econometrics*, 2247–2294. [45]
- ANDREWS, DONALD W. K. AND XIAOXIA SHI (2013): “Inference based on conditional moment inequalities,” *Econometrica*, 81, 609–666. [28]
- 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*. [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-Accuracy Improvability of Algorithms,” *arXiv preprint arXiv:2405.04816*. [5, 9]
- BAROCAS, SOLON, MORITZ HARDT, AND ARVIND NARAYANAN (2023): *Fairness and Machine Learning: Limitations and Opportunities*, MIT Press, <http://www.fairmlbook.org>. [5]
- BERESTEANU, ARIE AND FRANCESCA MOLINARI (2008): “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76, 763–814. [5]
- BERK, RICHARD, HODA HEIDARI, SHAHIN JABBARI, MATTHEW JOSEPH, MICHAEL KEARNS, JAMIE MORGENTERN, 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]
- BLATTNER, LAURA AND JANN SPIESS (2022): “Machine Learning Explainability and Fairness: Insights from Consumer Lending,” *FinRegLab Empirical White Paper*. [27]
- BONTEMPS, CHRISTIAN, THIERRY MAGNAC, AND ERIC MAURIN (2012): “Set identified linear models,” *Econometrica*, 80, 1129–1155. [14, 37]
- CANAY, IVAN A., GASTÓN ILLANES, AND AMILCAR VELEZ (2026): “A User’s Guide for Inference in Models Defined by Moment Inequalities,” *Journal of Econometrics*, forthcoming. [4]
- 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,” CeMMAP working paper CWP09/19, available at <https://www.cemmap.ac.uk/publication/id/13913>. [5]
- 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,” *The Econometrics Journal*, 21, C1–C68. [5, 22, 24, 64, 67]
- CHERNOZHUKOV, VICTOR, CHRISTIAN HANSEN, NATHAN KALLUS, MARTIN SPINDLER, AND VASILIS SYRGKANIS (2024): *Applied Causal Inference Powered by ML and AI*, Online; Version 0.1.1. [5]
- CHOULDECHOVA, ALEXANDRA AND AARON ROTH (2018): “The Frontiers of Fairness in Machine Learning,” *arXiv preprint arXiv:1810.08810*. [5]
- CORBETT-DAVIES, SAM, JOHANN D. GAEBLER, HAMED NILFOROSHAN, RAVI SHROFF, AND SHARAD GOEL (2023): “The measure and mismeasure of fairness,” *Journal of Machine Learning Research*, 24, 1–117. [5]
- CORBETT-DAVIES, SAM, EMMA PIERSON, AVI FELLER, SHARAD GOEL, AND AZIZ HUQ (2017): “Algorithmic Decision Making and the Cost of Fairness,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 797–806. [3]

- COSTON, AMANDA, ASHESH RAMBACHAN, AND ALEXANDRA CHOULDECHOVA (2021): “Characterizing Fairness Over the Set of Good Models Under Selective Labels,” in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, vol. 139, 2144–2155. [5]
- 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]
- EUROPEAN PARLIAMENT AND COUNCIL (2024): “Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence,” OJ L, 2024/1689, 12.7.2024. [2, 4]
- FALLAH, ALIREZA, MICHAEL I. JORDAN, AND ANNIE ULICHNEY (2026): “The Statistical Fairness-Accuracy Frontier,” *arXiv preprint arXiv:2508.17622*. [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. [5, 47]
- GEBRU, TIMNIT, JAMIE MORGENSTERN, BRIANA VECCHIONE, JENNIFER WORTMAN VAUGHAN, HANNA WALLACH, HAL DAUMÉ III, AND KATE CRAWFORD (2021): “Datasheets for Datasets,” *Communications of the ACM*, 64, 86–92. [2]
- GILLIS, TALIA B., VITALY MEURSAULT, AND BERK USTUN (2024): “Operationalizing the Search for Less Discriminatory Alternatives in Fair Lending,” in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, ACM, 377–387. [3]
- HECKMAN, JAMES J. (1979): “Sample Selection Bias as a Specification Error,” *Econometrica*, 47, 153–161. [4]
- HONG, HAN AND JESSIE LI (2018): “The numerical delta method,” *Journal of Econometrics*, 206, 379–394, special issue on Advances in Econometric Theory: Essays in honor of Takeshi Amemiya. [47]
- KAIDO, HIROAKI (2016): “A dual approach to inference for partially identified econometric models,” *Journal of Econometrics*, 192, 269 – 290. [47]
- KHAN, SHAKEEB, ELIE TAMER, AND QINGSONG YAO (2025): “Inference on High Dimensional Selective Labeling Models,” *arXiv preprint arXiv:2410.18381v4*. [5, 15]
- KIM, JOON SIK, JIAHAO CHEN, AND AMEET TALWALKAR (2020): “FACT: A Diagnostic for Group Fairness Trade-offs,” in *Proceedings of the 37th International Conference on Machine Learning*, PMLR, vol. 119 of *Proceedings of Machine Learning Research*, 5264–5274. [3]
- KLEINBERG, JON, HIMABINDU LAKKARAJU, JURE LESKOVEC, JENS LUDWIG, AND SENDHIL MULLAINATHAN (2018): “Human Decisions and Machine Predictions,” *The Quarterly Journal of Economics*, 133, 237–293. [5, 15]
- KLEINBERG, JON, SENDHIL MULLAINATHAN, AND MANISH RAGHAVAN (2016): “Inherent Trade-Offs in the Fair Determination of Risk Scores,” *arXiv preprint arXiv:1609.05807*. [5]

- LAKKARAJU, HIMABINDU, JON KLEINBERG, JURE LESKOVEC, JENS LUDWIG, AND SENDHIL MULLAINATHAN (2017): “The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables,” *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [5, 15]
- LEHMANN, ERICH L. AND JOSEPH P. ROMANO (2008): *Testing Statistical Hypotheses*, Springer New York. [12]
- LIANG, ANNIE, JAY LU, XIAOSHENG MU, AND KYOHEI OKUMURA (2026): “Algorithm Design: A Fairness-Accuracy Frontier,” *Journal of Political Economy*, 134, 1401–1467. [3, 6, 7, 8, 9, 10, 13, 14, 28]
- LITTLE, CAMILLE OLIVIA, MICHAEL WEYLANDT, AND GENEVERA I. ALLEN (2022): “To the Fairness Frontier and Beyond: Identifying, Quantifying, and Optimizing the Fairness-Accuracy Pareto Frontier,” *arXiv preprint arXiv:2206.00074*. [5]
- LIU, YIQI AND FRANCESCA MOLINARI (2026): “Inference for an Algorithmic Fairness-Accuracy Frontier,” *revision in progress. Previous version available as arXiv preprint arXiv:2402.08879v2*. [2, 3, 5, 6, 10, 12, 13, 14, 23, 24, 27, 28, 63]
- MANSKI, CHARLES F. (1989): “Anatomy of the Selection Problem,” *Journal of Human Resources*, 24, 343–360. [4]
- MENON, ADITYA KRISHNA AND ROBERT C. WILLIAMSON (2018): “The Cost of Fairness in Binary Classification,” in *Proceedings of Machine Learning Research*, PMLR, vol. 81, 107–118. [3, 4]
- MITCHELL, MARGARET, SIMONE WU, ANDREW ZALDIVAR, PARKER BARNES, LUCY VASSERMAN, BEN HUTCHINSON, ELENA SPITZER, INIOLUWA DEBORAH RAJI, AND TIMNIT GEBRU (2019): “Model Cards for Model Reporting,” in *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*)*, 220–229. [2]
- MOLCHANOV, I. (2017): *Theory of Random Sets*, London: Springer, 2 ed. [58, 59]
- MOLCHANOV, ILYA AND FRANCESCA MOLINARI (2018): *Random Sets in Econometrics*, Econometric Society Monograph Series, Cambridge University Press, Cambridge UK. [58]
- 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]
- 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]
- RAJI, INIOLUWA DEBORAH, ANDREW SMART, REBECCA N. WHITE, MARGARET MITCHELL, TIMNIT GEBRU, BEN HUTCHINSON, JAMILA SMITH-LOUD, DANIEL THERON, AND PARKER BARNES (2020): “Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT*)*, 33–44. [2]
- RAMBACHAN, ASHESH, AMANDA COSTON, AND EDWARD H. KENNEDY (2025): “Robust Design and Evaluation of Predictive Algorithms under Unobserved Confounding,” *arXiv preprint arXiv:2212.09844v6*. [5, 15]

- RAO, M. M. AND R. J. SWIFT (2006): *Probability Theory With Applications*, vol. 582 of *Mathematics and Its Applications*, New York: Springer-Verlag. [49]
- ROCKAFELLAR, R. T. (1997): *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Princeton University Press. [42]
- ROMANO, JOSEPH P. AND AZEEM M. SHAIKH (2012): “On the uniform asymptotic validity of subsampling and the bootstrap,” *The Annals of Statistics*, 40, 2798 – 2822. [47, 48]
- SCHNEIDER, ROLF (1993): *Convex Bodies: The Brunn-Minkowski Theory*, Encyclopedia of Mathematics and its Applications, Cambridge University Press. [10, 12, 35, 51]
- SEMENOVA, VIRA (2023): “Debiased machine learning of set-identified linear models,” *Journal of Econometrics*, 235, 1725–1746. [5]
- U.S. CONGRESS (1991): “Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071,” Codified at 42 U.S.C. § 2000e et seq. Section 2000e-2(k) establishes disparate impact framework and alternative employment practice requirement. [3, 4]
- 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. [44, 45, 48]
- VELEZ, AMILCAR (2024): “On the asymptotic properties of debiased machine learning estimators,” *arXiv preprint arXiv:2411.01864*. [22]
- WEI, SUSAN AND MARC NIETHAMMER (2022): “The fairness-accuracy Pareto front,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15, 287–302. [5]

APPENDIX A: PROOFS OF THE MAIN RESULTS

Assume throughout that $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a standard Borel space.

A.1. Proofs for Section 2

PROOF OF THEOREM 2.1: Recall $\mathcal{E} = \{(e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a)) : a \in \mathcal{A}\}$, so

$$h_{\mathcal{E}}(q) = \sup_{a \in \mathcal{A}} q^{\top} \left(e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a) \right).$$

Using the representation $(e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a)) = \mathbb{E}^*[\boldsymbol{\theta}_0(X)] + \mathbb{E}^*[a(X)\Delta\boldsymbol{\theta}(X)]$, we obtain $q^{\top} (e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a)) = \mathbb{E}^*[q^{\top}\boldsymbol{\theta}_0(X)] + \mathbb{E}^*[a(X)q^{\top}\Delta\boldsymbol{\theta}(X)]$. Hence,

$$h_{\mathcal{E}}(q) = \mathbb{E}^*[q^{\top}\boldsymbol{\theta}_0(X)] + \sup_{a \in \mathcal{A}} \mathbb{E}^*[a(X)q^{\top}\Delta\boldsymbol{\theta}(X)].$$

Because $a(X) \in [0, 1]$ and $q^\top \Delta \boldsymbol{\theta}(X)$ is $\sigma(X)$ -measurable, the pointwise maximizer is

$$a^q(X) \in \arg \max_{t \in [0, 1]} t q^\top \Delta \boldsymbol{\theta}(X) = \begin{cases} 1, & q^\top \Delta \boldsymbol{\theta}(X) > 0, \\ 0, & q^\top \Delta \boldsymbol{\theta}(X) < 0, \end{cases}$$

with arbitrary tie-breaking on $\{q^\top \Delta \boldsymbol{\theta}(X) = 0\}$. Therefore,

$$\sup_{a \in \mathcal{A}} \mathbb{E}^* [a(X) q^\top \Delta \boldsymbol{\theta}(X)] = \mathbb{E}^* [(q^\top \Delta \boldsymbol{\theta}(X))_+].$$

Next, note that $\mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}$ is $\sigma(X)$ -measurable. Therefore, by iterated expectations,

$$\mathbb{E}^* [q^\top \mathbf{L}_0 + q^\top (\mathbf{L}_1 - \mathbf{L}_0) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}] = \mathbb{E}^* [q^\top \boldsymbol{\theta}_0(X) + q^\top \Delta \boldsymbol{\theta}(X) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}].$$

This proves (10). To prove (11), fix $q \in \mathbb{S}^2$ and let $v \in \mathbb{R}^3$ be arbitrary; denote $u_+ \equiv \max\{u, 0\}$. For $t \in \mathbb{R}$ such that $q + tv \neq 0$, using the support-function formula, we have

$$\frac{h_{\mathcal{E}}(q+tv) - h_{\mathcal{E}}(q)}{t} = \mathbb{E}^* \left[v^\top \boldsymbol{\theta}_0(X) + \frac{((q+tv)^\top \Delta \boldsymbol{\theta}(X))_+ - (q^\top \Delta \boldsymbol{\theta}(X))_+}{t} \right].$$

Hence it is enough to study the difference quotient of the map $u \mapsto u_+$.

For each realization $X = x$ such that $q^\top \Delta \boldsymbol{\theta}(x) \neq 0$, the function $t \mapsto ((q + tv)^\top \Delta \boldsymbol{\theta}(x))_+$ is differentiable at $t = 0$, with derivative $v^\top \Delta \boldsymbol{\theta}(x) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(x) > 0\}$. Therefore, pointwise on the event $\{q^\top \Delta \boldsymbol{\theta}(X) \neq 0\}$, as $t \rightarrow 0$,

$$\frac{((q+tv)^\top \Delta \boldsymbol{\theta}(X))_+ - (q^\top \Delta \boldsymbol{\theta}(X))_+}{t} \rightarrow v^\top \Delta \boldsymbol{\theta}(X) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}. \quad (46)$$

Assumption 2 implies $\mathbb{P}^*(q^\top \Delta \boldsymbol{\theta}(X) = 0) = 0$ for every $q \in \mathbb{S}^2$, because for every $\delta > 0$,

$$\mathbb{P}^*(q^\top \Delta \boldsymbol{\theta}(X) = 0) \leq \sup_{\tilde{q} \in \mathbb{S}^2} \mathbb{P}^*(|\tilde{q}^\top \Delta \boldsymbol{\theta}(X)| \leq \delta) \lesssim \delta^m$$

and letting $\delta \downarrow 0$ yields the claim. Thus, the convergence in (46) holds \mathbb{P}^* -almost surely.

Next we obtain an integrable dominating function. Since $u \mapsto u_+$ is 1-Lipschitz,

$$\left| \frac{((q+tv)^\top \Delta \boldsymbol{\theta}(X))_+ - (q^\top \Delta \boldsymbol{\theta}(X))_+}{t} \right| \leq |v^\top \Delta \boldsymbol{\theta}(X)|$$

for all $t \neq 0$. Therefore,

$$\left| v^\top \boldsymbol{\theta}_0(X) + \frac{((q+tv)^\top \Delta \boldsymbol{\theta}(X))_+ - (q^\top \Delta \boldsymbol{\theta}(X))_+}{t} \right| \leq |v^\top \boldsymbol{\theta}_0(X)| + |v^\top \Delta \boldsymbol{\theta}(X)|$$

By Assumption 1, the random vectors $\boldsymbol{\theta}_0(X)$ and $\Delta \boldsymbol{\theta}(X)$ are integrable, so the right-hand side is integrable. Hence, by dominated convergence,

$$\lim_{t \rightarrow 0} \frac{h_{\mathcal{E}}(q+tv) - h_{\mathcal{E}}(q)}{t} = v^\top \mathbb{E}^* [\boldsymbol{\theta}_0(X) + \Delta \boldsymbol{\theta}(X) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}].$$

Thus the directional derivative of $h_{\mathcal{E}}$ at q in direction v exists and is linear in v . It follows that $h_{\mathcal{E}}$ is differentiable at q , with gradient

$$\nabla_q h_{\mathcal{E}}(q) = \mathbb{E}^* [\boldsymbol{\theta}_0(X) + \Delta \boldsymbol{\theta}(X) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X) > 0\}].$$

Using again the law of iterated expectations yields (11).

Finally, by Schneider (1993, Corollary 1.7.3), \mathcal{E} is strictly convex because by (11) its support set is a singleton in each direction $q \in \mathbb{S}^2$. *Q.E.D.*

PROOF OF THEOREM 2.2: Recall $\tilde{\mathbb{S}}^2 \equiv \{q \in \mathbb{S}^2 : q_1 \geq 0, q_2 \geq 0\}$ and that under Assumptions 1-2, \mathcal{E} has a nonempty interior, is compact, and is strictly convex.

Step 1: (i) \iff (ii). By definition,

$$\mathcal{C}(\varepsilon^*) = \{\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in \mathbb{R}^3 : \varepsilon_1 \leq \varepsilon_1^*, \varepsilon_2 \leq \varepsilon_2^*, |\varepsilon_3| \leq |\varepsilon_3^*|\}.$$

Hence $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon \in \mathcal{E} : \varepsilon_1 \leq \varepsilon_1^*, \varepsilon_2 \leq \varepsilon_2^*, |\varepsilon_3| \leq |\varepsilon_3^*|\}$. Next, recall that $\varepsilon^* \in \mathcal{F}$ if and only if there does not exist $\varepsilon \in \mathcal{E}$ with $\varepsilon_1 \leq \varepsilon_1^*$, $\varepsilon_2 \leq \varepsilon_2^*$, and $|\varepsilon_3| \leq |\varepsilon_3^*|$, with at least one strict inequality. Since $\varepsilon^* \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$, this is equivalent to $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$.

Step 2: (ii) \iff (iv). We begin showing (ii) \implies (iv). Assume $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$. We prove first that $|\varepsilon_3^*| = d((\varepsilon_1^*, \varepsilon_2^*))$. If not, then $d((\varepsilon_1^*, \varepsilon_2^*)) < |\varepsilon_3^*|$, and by definition of d , there exists $\tilde{\varepsilon}_3 \in \mathbb{R}$ such that $(\varepsilon_1^*, \varepsilon_2^*, \tilde{\varepsilon}_3) \in \mathcal{E}$ and $|\tilde{\varepsilon}_3| < |\varepsilon_3^*|$. But then $(\varepsilon_1^*, \varepsilon_2^*, \tilde{\varepsilon}_3) \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$ and this point is distinct from ε^* , contradicting $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$. Hence, $|\varepsilon_3^*| = d((\varepsilon_1^*, \varepsilon_2^*))$. We next show that $\pi(\varepsilon^*) = (\varepsilon_1^*, \varepsilon_2^*) \in \mathcal{F}^A$. Suppose not. Then there exists $(c_r, c_b) \in \mathcal{E}^A$ such that $c_r \leq \varepsilon_1^*$, $c_b \leq \varepsilon_2^*$, and $d((c_r, c_b)) \leq d((\varepsilon_1^*, \varepsilon_2^*))$, with at least one strict inequality. As

$(c_r, c_b) \in \mathcal{E}^A$, by definition of d there exists $\tilde{\varepsilon}_3 \in \mathbb{R}$ such that $(c_r, c_b, \tilde{\varepsilon}_3) \in \mathcal{E}$ and $|\tilde{\varepsilon}_3| = d((c_r, c_b))$. Using the equality already proved, $d((\varepsilon_1^*, \varepsilon_2^*)) = |\varepsilon_3^*|$, we obtain $c_r \leq \varepsilon_1^*$, $c_b \leq \varepsilon_2^*$, and $|\tilde{\varepsilon}_3| \leq |\varepsilon_3^*|$, with at least one strict inequality. Thus $(c_r, c_b, \tilde{\varepsilon}_3) \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$ with $(c_r, c_b, \tilde{\varepsilon}_3) \neq \varepsilon^*$, contradicting $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$. Hence, $\pi(\varepsilon^*) \in \mathcal{F}^A$, proving (ii) \Rightarrow (iv).

We now prove (iv) \Rightarrow (ii). Assume $\pi(\varepsilon^*) = (\varepsilon_1^*, \varepsilon_2^*) \in \mathcal{F}^A$ and $|\varepsilon_3^*| = d((\varepsilon_1^*, \varepsilon_2^*))$. Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$. Then $\varepsilon_1 \leq \varepsilon_1^*$, $\varepsilon_2 \leq \varepsilon_2^*$, and $|\varepsilon_3| \leq |\varepsilon_3^*|$. Since $\varepsilon \in \mathcal{E}$, we have

$$d((\varepsilon_1, \varepsilon_2)) \leq |\varepsilon_3| \leq |\varepsilon_3^*| = d((\varepsilon_1^*, \varepsilon_2^*)). \quad (47)$$

Thus, $(\varepsilon_1, \varepsilon_2) \in \mathcal{E}^A$ weakly improves on $(\varepsilon_1^*, \varepsilon_2^*)$ in both group-accuracy losses and in the fairness index d . As $(\varepsilon_1^*, \varepsilon_2^*) \in \mathcal{F}^A$, it follows that $(\varepsilon_1, \varepsilon_2) = (\varepsilon_1^*, \varepsilon_2^*)$ (else we would contradict $(\varepsilon_1^*, \varepsilon_2^*) \in \mathcal{F}^A$). Substituting back $(\varepsilon_1^*, \varepsilon_2^*)$ for $(\varepsilon_1, \varepsilon_2)$ in (47) yields $|\varepsilon_3| = |\varepsilon_3^*|$. If $\varepsilon_3 \neq \varepsilon_3^*$, then necessarily $\varepsilon_3 = -\varepsilon_3^*$. If $\varepsilon_3^* = 0$, this is impossible, so $\varepsilon_3 = \varepsilon_3^*$. If $\varepsilon_3^* \neq 0$, by convexity of \mathcal{E} the midpoint $\bar{\varepsilon} \equiv \frac{\varepsilon + \varepsilon^*}{2} = (\varepsilon_1^*, \varepsilon_2^*, 0)$ belongs to \mathcal{E} . Therefore, $d((\varepsilon_1^*, \varepsilon_2^*)) \leq 0$, which implies $d((\varepsilon_1^*, \varepsilon_2^*)) = 0$. But then $|\varepsilon_3^*| = d((\varepsilon_1^*, \varepsilon_2^*)) = 0$, contradicting $\varepsilon_3^* \neq 0$. Hence $\varepsilon_3 = \varepsilon_3^*$, so $\varepsilon = \varepsilon^*$. We conclude that $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$, proving (iv) \Rightarrow (ii).

Step 3: (ii) \Rightarrow (iii). Assume $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$. As \mathcal{E} is compact and convex with nonempty interior, and $\mathcal{C}(\varepsilon^*)$ is closed and convex, the two sets admit a common supporting hyperplane at their unique common point. Hence, there exists $q \neq 0$ and $\alpha \in \mathbb{R}$ such that

$$q^\top \varepsilon \geq \alpha \geq q^\top c \quad \forall \varepsilon \in \mathcal{E}, \forall c \in \mathcal{C}(\varepsilon^*).$$

As $\varepsilon^* \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$, necessarily $\alpha = q^\top \varepsilon^*$, hence $\inf_{\varepsilon \in \mathcal{E}} q^\top \varepsilon = \sup_{c \in \mathcal{C}(\varepsilon^*)} q^\top c$. Equivalently, $-h_{\mathcal{E}}(-q) = h_{\mathcal{C}(\varepsilon^*)}(q)$. Since $h_{\mathcal{C}(\varepsilon^*)}(q) < +\infty$, we must have $q_1 \geq 0$ and $q_2 \geq 0$, because $\mathcal{C}(\varepsilon^*)$ is unbounded in the negative first- and second-coordinate directions. Since $q \neq 0$, positive homogeneity of support functions implies that, after replacing q by $q/\|q\|$, we may assume $q \in \tilde{\mathbb{S}}^2$ while preserving the equality $h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q) = 0$. Since $\varepsilon^* \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$, for every $q \in \tilde{\mathbb{S}}^2$, $h_{\mathcal{C}(\varepsilon^*)}(q) \geq q^\top \varepsilon^*$ and $h_{\mathcal{E}}(-q) \geq -q^\top \varepsilon^*$, and hence $h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q) \geq 0$. Therefore, $\min_{q \in \tilde{\mathbb{S}}^2} (h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)) = 0$, proving (ii) \Rightarrow (iii).

Step 4: (iii) \Rightarrow (ii). Assume $\min_{q \in \tilde{\mathbb{S}}^2} (h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)) = 0$. As $\tilde{\mathbb{S}}^2$ is compact, \mathcal{E} is compact, and $h_{\mathcal{C}(\varepsilon^*)}$ is finite and continuous on $\tilde{\mathbb{S}}^2$, the minimum is attained at some $\bar{q} \in \tilde{\mathbb{S}}^2$.

Thus, $\sup_{c \in \mathcal{C}(\varepsilon^*)} \bar{q}^\top c = \inf_{\varepsilon \in \mathcal{E}} \bar{q}^\top \varepsilon$. Let $\tilde{\varepsilon} \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$. Then

$$\bar{q}^\top \tilde{\varepsilon} \leq \sup_{c \in \mathcal{C}(\varepsilon^*)} \bar{q}^\top c = \inf_{\varepsilon \in \mathcal{E}} \bar{q}^\top \varepsilon \leq \bar{q}^\top \tilde{\varepsilon}.$$

Hence, every $\tilde{\varepsilon} \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$ satisfies $\bar{q}^\top \tilde{\varepsilon} = \inf_{\varepsilon \in \mathcal{E}} \bar{q}^\top \varepsilon$. Therefore,

$$\mathcal{E} \cap \mathcal{C}(\varepsilon^*) \subset \left\{ \varepsilon \in \mathcal{E} : \bar{q}^\top \varepsilon = \inf_{\tilde{\varepsilon} \in \mathcal{E}} \bar{q}^\top \tilde{\varepsilon} \right\},$$

that is, the intersection is contained in the support set of \mathcal{E} in direction $-\bar{q}$. Since \mathcal{E} is strictly convex, every support set of \mathcal{E} is a singleton. Because $\varepsilon^* \in \mathcal{E} \cap \mathcal{C}(\varepsilon^*)$, the above inclusion implies $\mathcal{E} \cap \mathcal{C}(\varepsilon^*) = \{\varepsilon^*\}$, proving (iii) \Rightarrow (ii).

Combining Steps 1-4, we obtain (i) \iff (ii) \iff (iii) \iff (iv). *Q.E.D.*

PROOF OF THEOREM 2.3: We divide the proof of this theorem into four steps. The first two steps prove the first part of the theorem using Lemma B.3, while the last two steps additionally use the injectivity of the support map $\mathcal{S}_{\mathcal{E}}$ implied by the non-kink condition.

Step 1: Suppose that for any $\varepsilon^* \in \mathcal{F}$ there exists $q^* = q(\alpha) \in \tilde{\mathbb{S}}^2$ (defined in the statement of the theorem) such that $\varepsilon^* = \mathcal{S}_{\mathcal{E}}(-q^*)$ and $q_3^* = 0$. By part 1 of Lemma B.3, this implies that $\varepsilon^* \in \mathcal{PF}$; therefore, $\mathcal{F} \subset \mathcal{PF}$. Since $\mathcal{PF} \subset \mathcal{F}$ by Lemma B.3, we conclude $\mathcal{PF} = \mathcal{F}$.

Step 2: Suppose that $\mathcal{PF} = \mathcal{F}$. Since $\varepsilon^* \in \mathcal{F}$ implies $\varepsilon^* \in \mathcal{PF}$, we conclude that there exist $q^* \in \tilde{\mathbb{S}}^2$ such that $\varepsilon^* = \mathcal{S}_{\mathcal{E}}(-q^*)$ and $q_3^* = 0$ by part 1 of Lemma B.3. We conclude the proof by setting $q(\alpha) = q^*$.

Recall that Assumption 3 (no kinks) and the argument in Bontemps et al. (2012, Supplemental Appendix B.2.3) imply the injectivity of the support map $\mathcal{S}_{\mathcal{E}}$.

Step 3: Suppose that $\mathcal{PF} = \mathcal{F}$. Note that the injectivity of $\mathcal{S}_{\mathcal{E}}$ implies that any $\varepsilon^* \in \mathcal{F}$ has a unique support direction q whose third coordinate is 0 due to $\mathcal{PF} = \mathcal{F}$ and part 1 of Lemma B.3. Now, suppose that $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$ is false, that is, there exist $\varepsilon^* \in \mathcal{PF}$ such that $\varepsilon_3^* \neq 0$. In what follows, we will find $\varepsilon^t \in \mathcal{F}$ such that $\varepsilon^t = \mathcal{S}_{\mathcal{E}}(-q^t)$ and $q_3^t \neq 0$, which will be a contradiction to $\mathcal{PF} = \mathcal{F}$ and the injectivity of $\mathcal{S}_{\mathcal{E}}$.

By part 1 of Lemma B.3, there exist $q = (q_1, q_2, 0) \in \tilde{\mathbb{S}}^2$ such that $\varepsilon^* = \mathcal{S}_{\mathcal{E}}(-q)$. Let $s \equiv \text{sgn}(\varepsilon_3^*) \in \{-1, 1\}$ and for $t > 0$ define

$$q^t = \frac{(q_1, q_2, st)}{\sqrt{q_1^2 + q_2^2 + t^2}}.$$

Then $q^t \rightarrow q$ as $t \downarrow 0$. By continuity of $\mathcal{S}_{\mathcal{E}}$,

$$\varepsilon^t = \mathcal{S}_{\mathcal{E}}(-q^t) \rightarrow \mathcal{S}_{\mathcal{E}}(-q) = \varepsilon^*.$$

Hence, for sufficiently small $t > 0$, $\text{sgn}(\varepsilon_3^t) = s = \text{sgn}(q_3^t)$, so $q_3^t \varepsilon_3^t > 0$. By part 2 of Lemma B.3, this implies $\varepsilon^t \in \mathcal{F}$ such that $\varepsilon^t = \mathcal{S}_{\mathcal{E}}(-q^t)$ and $q_3^t \neq 0$, which is a contradiction.

Step 4: Suppose that $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$. Note that $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$ implies that the third coordinate of both $R = \mathcal{S}_{\mathcal{E}}(-u_1)$ and $B = \mathcal{S}_{\mathcal{E}}(-u_2)$ is zero, where $u_1 = (1, 0, 0)^\top$ and $u_2 = (0, 1, 0)^\top$; therefore, $u_3^\top R = 0$ and $u_3^\top B = 0$ where $u_3 = (0, 0, 1)$. Since $R \neq B$, it follows that $\min_{\varepsilon \in \mathcal{E}} u_3^\top \varepsilon < 0$ due to strict convexity of \mathcal{E} . This implies that $\mathcal{S}_{\mathcal{E}}(-u_3) \notin \mathcal{F}$ by part 2 of Lemma B.3. Similarly, we can conclude $\mathcal{S}_{\mathcal{E}}(u_3) \notin \mathcal{F}$.

Since $q \notin \{u_3, -u_3\}$, it follows that $q_1 + q_2 > 0$. Let $\varepsilon^* \in \mathcal{F}$. By part 2 of Lemma B.3, there exist $q \in \tilde{\mathbb{S}}^2$ such that $\varepsilon^* = \mathcal{S}_{\mathcal{E}}(-q)$ and $\varepsilon_3^* q_3 \geq 0$. Let $\tilde{q} = (q_1, q_2, 0) / \sqrt{q_1^2 + q_2^2}$ and $\rho = \mathcal{S}_{\mathcal{E}}(-\tilde{q})$. Consider the following derivations:

$$q^\top \rho \stackrel{(1)}{=} q_1 \rho_1 + q_2 \rho_2 \stackrel{(2)}{\leq} q_1 \varepsilon_1^* + q_2 \varepsilon_2^* \stackrel{(3)}{\leq} q_1 \varepsilon_1^* + q_2 \varepsilon_2^* + q_3 \varepsilon_3^* \stackrel{(4)}{\leq} q^\top \rho,$$

where (1) holds since $\rho \in \mathcal{PF}$ by part 1 of Lemma B.3 and $\rho_3 = 0$ due to $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$, (2) holds since ρ minimizes $q_1 \varepsilon_1 + q_2 \varepsilon_2$ over \mathcal{E} , (3) holds since $\varepsilon^* \in \mathcal{F}$ and $\varepsilon_3^* q_3 \geq 0$ by part 2 of Lemma B.3, and (4) holds since ε^* minimizes $q_1 \varepsilon_1 + q_2 \varepsilon_2 + q_3 \varepsilon_3$ over \mathcal{E} . Therefore, ρ also minimize $q_1 \varepsilon_1 + q_2 \varepsilon_2 + q_3 \varepsilon_3$ over \mathcal{E} . As a result, due to strict convexity of \mathcal{E} , we have that $\rho = \varepsilon^*$, which implies that $\varepsilon^* \in \mathcal{PF}$. Therefore, $\mathcal{F} \subset \mathcal{PF}$ which implies $\mathcal{F} = \mathcal{PF}$ since $\mathcal{PF} \subset \mathcal{F}$ due to Lemma B.3. *Q.E.D.*

A.2. Proofs for Section 3

A.2.1. Proofs for Section 3.1

For $\lambda \in \Lambda$, let $\theta_d(x; \lambda)$ denote the completion-dependent analogue of $\theta_d(x)$. Under the special losses considered here,

$$\begin{aligned} L_0^{g,A} &= \frac{Y_i^* \mathbb{1}\{G=g\}}{\mu_g}, & L_1^{g,A} &= \frac{(1-Y_i^*) \mathbb{1}\{G=g\}}{\mu_g} \\ L_0^{r,F} - L_0^{b,F} &= 0, & L_1^{r,F} - L_1^{b,F} &= \frac{\mathbb{1}\{G=r\}}{\mu_r} - \frac{\mathbb{1}\{G=b\}}{\mu_b} \end{aligned}$$

The next lemma gives the λ -dependent representations used in Section 3.1.

LEMMA A.1: *For every $\lambda \in \Lambda$ and every $x \in \mathcal{X}$,*

$$\theta_0(x; \lambda) = A_0(x) + B(x)\lambda(x), \quad \theta_1(x; \lambda) = A_1(x) - B(x)\lambda(x). \quad (48)$$

Consequently,

$$h_{\mathcal{E}}(q; \lambda) = \mathbb{E}[\max \{q^\top (A_0(X) + B(X)\lambda(X)), q^\top (A_1(X) - B(X)\lambda(X))\}], \quad (49)$$

$$h_{\mathcal{C}(\varepsilon^*(a; \lambda))}(q) = C(q; a) + \mathbb{E}[(1 - 2a(X))q^\top B(X)\lambda(X)], \quad q \in \tilde{\mathcal{S}}^2, \quad (50)$$

where

$$C(q; a) \equiv \sum_{j=1}^2 q_j \mathbb{E}[a(X)A_{1,j}(X) + (1 - a(X))A_{0,j}(X)] + |q_3| |\mathbb{E}[a(X)A_{1,3}(X)]|. \quad (51)$$

PROOF: Fix $g \in \{r, b\}$. Using $Y = ZY^*$ and the definition of $\lambda_g(x)$,

$$\begin{aligned} \theta_0^{g,A}(x; \lambda) &= \mathbb{E}^* \left[\frac{Y^* \mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right] \\ &= \mathbb{E} \left[\frac{ZY \mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right] + \mathbb{E} \left[\frac{(1-Z)Y^* \mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right] \\ &= \mathbb{E} \left[\frac{ZY \mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right] + \lambda_g(x) \mathbb{E} \left[\frac{(1-Z) \mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right], \end{aligned}$$

which is the g th accuracy coordinate of $A_0(x) + B(x)\lambda(x)$. Likewise,

$$\theta_1^{g,A}(x; \lambda) = \mathbb{E}^* \left[\frac{(1-Y^*) \mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right]$$

$$= \mathbb{E} \left[\frac{(Z(1-Y) + (1-Z))\mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right] - \lambda_g(x) \mathbb{E} \left[\frac{(1-Z)\mathbb{1}\{G=g\}}{\mu_g} \middle| X = x \right],$$

which is the g th accuracy coordinate of $A_1(x) - B(x)\lambda(x)$. For the fairness coordinate,

$$\theta_0^F(x; \lambda) = 0, \quad \theta_1^F(x; \lambda) = \mathbb{E} \left[\frac{\mathbb{1}\{G=r\}}{\mu_r} - \frac{\mathbb{1}\{G=b\}}{\mu_b} \middle| X = x \right],$$

which does not depend on λ . This proves (48). Equation (49) now follows directly from Theorem 2.1 applied to the λ -dependent pair $(\theta_0(\cdot; \lambda), \theta_1(\cdot; \lambda))$.

Next fix an algorithm a . By construction,

$$\varepsilon^*(a; \lambda) = \mathbb{E}[a(X)\theta_1(X; \lambda) + (1 - a(X))\theta_0(X; \lambda)].$$

For $j \in \{1, 2\}$, using (48),

$$\varepsilon_j^*(a; \lambda) = \mathbb{E}[a(X)A_{1,j}(X) + (1 - a(X))A_{0,j}(X)] + \mathbb{E}[(1 - 2a(X))B_{jj}(X)\lambda_j(X)].$$

For the third coordinate, $\varepsilon_3^*(a) = \mathbb{E}[a(X)A_{1,3}(X)]$, so it is point identified. Equation (13) then gives, for $q \in \tilde{\mathbb{S}}^2$,

$$\begin{aligned} h_{\mathcal{C}(\varepsilon^*(a; \lambda))}(q) &= q_1 \varepsilon_1^*(a; \lambda) + q_2 \varepsilon_2^*(a; \lambda) + |q_3| |\varepsilon_3^*(a)| \\ &= C(q; a) + \mathbb{E}[(1 - 2a(X))q^\top B(X)\lambda(X)], \end{aligned}$$

which is (50). *Q.E.D.*

PROOF OF THEOREM 3.1: By Assumption 4, all distributions consistent with the data also satisfy the conditions of Theorem 2.2; hence, algorithm a^* yields $\varepsilon^* \in \mathcal{F}$ if and only if

$$\min_{q \in \tilde{\mathbb{S}}^2} [h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)] = 0.$$

Given the definition of $J_d(\lambda; q, a)$, $d = 0, 1$, in Theorem 3.1, this is equivalent to

$$\min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E}^* [\max \{J_1(\lambda^*(X); q, a^*), J_0(\lambda^*(X); q, a^*)\}] = 0 \quad (52)$$

Since $\lambda^*(X)$ is not identified by the data, we say that a^* is on the fairness-accuracy frontier for some distribution consistent with the data if we can find a $\lambda^{(a^*)}(\cdot) \in \Lambda$ such that (52)

holds with $\lambda^*(X) = \lambda^{(a^*)}(X)$. We can write the existence equivalently as follows:

$$\min_{\lambda(\cdot) \in \Lambda} \min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E} [\max \{J_1(\lambda(X); q, a^*), J_0(\lambda(X); q, a^*)\}] = 0 \quad (53)$$

To simplify notation we use $J_d(\lambda)$ instead of $J_d(\lambda(X); q, a^*)$. Since $J_d(\lambda)$ is linear in λ , we write $J_d(\lambda) = \tilde{A}_d + \tilde{B}_d^\top \lambda$, where $\tilde{B}_d = (\tilde{B}_d^r, \tilde{B}_d^b)^\top$. We use the previous notation to write the left-hand side of (53) as follows

$$\begin{aligned} & \stackrel{(1)}{=} \min_{q \in \tilde{\mathbb{S}}^2} \min_{\lambda(\cdot) \in \Lambda} \mathbb{E}^* \left[\max_{t \in [0,1]} \{t J_1(\lambda) + (1-t) J_0(\lambda)\} \right] \\ & \stackrel{(2)}{=} \min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E} \left[\min_{\lambda(\cdot) \in \Lambda} \max_{t \in [0,1]} \{t J_1(\lambda) + (1-t) J_0(\lambda)\} \right] \\ & \stackrel{(3)}{=} \min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E} \left[\max_{t \in [0,1]} \min_{\lambda(\cdot) \in \Lambda} \left\{ t \left(\tilde{A}_1 + \tilde{B}_1^\top \lambda \right) + (1-t) \left(\tilde{A}_0 + \tilde{B}_0^\top \lambda \right) \right\} \right] \\ & \stackrel{(4)}{=} \min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E} \left[\max_{t \in [0,1]} \left\{ t \tilde{A}_1 + (1-t) \tilde{A}_0 + \min\{0, t \tilde{B}_1^r + (1-t) \tilde{B}_0^r\} + \min\{0, t \tilde{B}_1^b + (1-t) \tilde{B}_0^b\} \right\} \right] \\ & \stackrel{(5)}{=} \min_{q \in \tilde{\mathbb{S}}^2} \mathbb{E} \left[\max \left\{ \tilde{A}_0 + \min\{0, \tilde{B}_0^r\} + \min\{0, \tilde{B}_0^b\}, a^*(X) \tilde{A}_1 + (1-a^*(X)) \tilde{A}_0, \tilde{A}_1 \right\} \right] \end{aligned} \quad (54)$$

where (1) holds because $\max\{a, b\} = \max_{t \in [0,1]} \{ta + (1-t)b\}$, (2) holds by the measurability selection theorem, (3) holds by the minimax theorem since the objective function is linear in t and λ and both domains are convex, (4) by solving in $\lambda(X) \in [0, 1]^2$ conditional on X , and (5) holds by claims 1–3 below. Finally, (54) is sufficient to conclude the proof of the proposition since $\tilde{A}_0 + \min\{0, \tilde{B}_0^r\} + \min\{0, \tilde{B}_0^b\} = J_0(\lambda_1; q, a^*)$, $\tilde{A}_0 = J_0(\lambda_0; q, a^*)$, $\tilde{A}_1 = J_1(\lambda_0; q, a^*)$, and $a^*(X) \in \{0, 1\}$ due to Corollary 2.1.

Claim 1: $t \tilde{B}_1^g + (1-t) \tilde{B}_0^g \leq 0$ if $t \leq t^* = a^*(X)$. Recall that $\tilde{B}_d = (\tilde{B}_d^r, \tilde{B}_d^b)^\top$ and

$$\begin{aligned} \tilde{B}_0 &= -2a^*(X)q^\top B(X) \\ \tilde{B}_1 &= 2(1-a^*(X))q^\top B(X). \end{aligned}$$

Therefore, $t \tilde{B}_1 + (1-t) \tilde{B}_0 = 2(t - a^*(X))q^\top B(X)$, which implies the claim since $q_1 \geq 0$, $q_2 \geq 0$, and $B(X)$ defined in (18) is nonnegative.

Claim 2: $t\tilde{B}_1^g + (1-t)\tilde{B}_0^g \geq 0$ if $t \geq t^* = a^*(X)$. Similar to claim 1.

Claim 3: $J(t)$ is maximized at $t = 0$, $a^*(X)$, and 1, where

$$J(t) \equiv t\tilde{A}_1 + (1-t)\tilde{A}_0 + \min\{0, t\tilde{B}_1^r + (1-t)\tilde{B}_0^r\} + \min\{0, t\tilde{B}_1^b + (1-t)\tilde{B}_0^b\}.$$

Furthermore, $J(a^*(X)) = a^*(X)\tilde{A}_1 + (1-a^*(X))\tilde{A}_0$ and $J(1) = \tilde{A}_1$. The proof of this claim follows by linearity on t and claims 1 and 2. *Q.E.D.*

PROOF OF THEOREM 3.2: By definition, $\mathcal{F}^\exists = \bigcup_{\lambda \in \Lambda} \mathcal{F}(\lambda)$. First let $\varepsilon \in \mathcal{F}^\exists$. Then $\varepsilon \in \mathcal{F}(\lambda)$ for some $\lambda \in \Lambda$. By Lemma B.4, there exists $q \in \tilde{\mathbb{S}}^2$ such that $\varepsilon = \mathcal{S}_{\mathcal{E}(\lambda)}(-q)$ and $q_3\varepsilon_3 \geq 0$. Since $\mathcal{S}_{\mathcal{E}(\lambda)}(-q)$ is generated by a rule minimizing $q^\top \theta_d(x; \lambda)$ pointwise, the corresponding expected loss vector belongs to \mathcal{M}_q . Hence

$$\varepsilon \in \mathcal{M}_q \cap \{\eta \in \mathbb{R}^3 : q_3\eta_3 \geq 0\}.$$

Conversely, suppose $\varepsilon \in \mathcal{M}_q \cap \{\eta \in \mathbb{R}^3 : q_3\eta_3 \geq 0\}$ for some $q \in \tilde{\mathbb{S}}^2$. By definition of \mathcal{M}_q , there exist $\lambda \in \Lambda$ and a q -optimal rule a such that

$$\varepsilon = \mathbb{E}[(1-a(X))\theta_0(X; \lambda) + a(X)\theta_1(X; \lambda)].$$

This vector minimizes $q^\top \eta$ over $\eta \in \mathcal{E}(\lambda)$, so $\varepsilon = \mathcal{S}_{\mathcal{E}(\lambda)}(-q)$. Since $q_3\varepsilon_3 \geq 0$, the representation of $\mathcal{F}(\lambda)$ due to Lemma B.4 implies $\varepsilon \in \mathcal{F}(\lambda)$. Therefore, $\varepsilon \in \mathcal{F}^\exists$, establishing (28).

Using the result in (28), suppose $\varepsilon \in \mathcal{F}^\exists$. Then there exists $q \in \tilde{\mathbb{S}}^2$ such that $q_3\varepsilon_3 \geq 0$ and $\varepsilon \in \mathcal{M}_q$. By Lemma B.5, \mathcal{M}_q is compact and convex, hence representing set membership through support function dominance (Rockafellar, 1997, Chapter 13), we have

$$\rho_q(\varepsilon) \equiv \sup_{v \in \mathbb{S}^2} \{v^\top \varepsilon - h_{\mathcal{M}_q}(v)\} \leq 0, \quad \text{hence} \quad [\rho_q(\varepsilon)]_+ = 0.$$

Since q is feasible in the infimum defining $T_{\text{env}}(\varepsilon)$ in (29), $0 \leq T_{\text{env}}(\varepsilon) \leq [\rho_q(\varepsilon)]_+ = 0$. Therefore $T_{\text{env}}(\varepsilon) = 0$.

Conversely, suppose $T_{\text{env}}(\varepsilon) = 0$. By Lemma B.6, the infimum is attained. Hence there exists $q^* \in \{q \in \tilde{\mathbb{S}}^2 : q_3\varepsilon_3 \geq 0\}$ such that $[\rho_{q^*}(\varepsilon)]_+ = 0$. Thus, $\rho_{q^*}(\varepsilon) \leq 0$. Using again the support function representation of set membership for the compact convex set \mathcal{M}_{q^*} ,

$\varepsilon \in \mathcal{M}_{q^*}$. Since $q^* \in \{q \in \tilde{\mathbb{S}}^2 : q_3 \varepsilon_3 \geq 0\}$, we also have $q_3^* \varepsilon_3 \geq 0$. Therefore,

$$\varepsilon \in \mathcal{M}_{q^*} \cap \{\eta \in \mathbb{R}^3 : q_3^* \eta_3 \geq 0\} \subseteq \mathcal{F}^\exists,$$

where the last inclusion follows from (28). This proves $\varepsilon \in \mathcal{F}^\exists \iff T_{\text{env}}(\varepsilon) = 0$. *Q.E.D.*

A.2.2. Proofs for Section 3.2

PROOF OF LEMMA 3.1: Since $\pi(X) \in (0, 1)$ a.s., the ratio $Z\mathbf{L}_d/\pi(X)$ is well-defined. Applying the law of iterated expectations,

$$\mathbb{E} \left[\frac{Z\mathbf{L}_d}{\pi(X)} \middle| X \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{Z\mathbf{L}_d}{\pi(X)} \middle| X, Y^*, G \right] \middle| X \right] = \mathbb{E} \left[\frac{\mathbf{L}_d}{\pi(X)} \mathbb{E}[Z|X, Y^*, G] \middle| X \right].$$

By Assumption 5, $\mathbb{E}[Z|X, Y^*, G] = \pi(X)$ a.s., and hence $\frac{\mathbf{L}_d}{\pi(X)} \mathbb{E}[Z|X, Y^*, G] = \mathbf{L}_d$, a.s., and the characterization of $\theta_d(X)$ in (30) follows. Repeating the same argument and using (10) yields the characterizations of $h_{\mathcal{E}}(q)$ and $e_g^t(a)$ in (31)-(32). *Q.E.D.*

A.3. Proofs for Section 4

Let $\{\omega_i\}_{i=1}^n$ be random positive weights independent of the data. Define

$$\widehat{h}_{\mathcal{E}}^\omega(q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}} \right) \zeta_i(q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}), \quad (55)$$

where $\bar{\omega} = n^{-1} \sum_{i=1}^n \omega_i$ and $\widehat{\mathbf{L}}^\omega$ is defined as in (1)–(2) with μ_g replaced by $\widehat{\mu}_g^\omega = n^{-1} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}} \right) \mathbb{1}\{G_i = g\}$. Note that $\widehat{h}_{\mathcal{E}}(q; \widehat{\mathbf{L}}, \widehat{\boldsymbol{\eta}}) = \widehat{h}_{\mathcal{E}}^\omega(q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}})$ by taking $\omega_i = 1, \forall i$.

The next result presents a general version of Theorem 4.1. It presents the limiting distribution for the class of weighted DML estimator $\widehat{h}_{\mathcal{E}}^\omega$ of the support function $h_{\mathcal{E}}$.

THEOREM A.1: *Let Assumptions 1–7 hold and let $\{\omega_i\}_{i=1}^n$ be random positive weights independent of the data such that $\mathbb{E}[\omega_i] = 1$ and $\mathbb{E}[\omega_i^2] < C_\omega$. Then,*

$$\sqrt{n} \left(\widehat{h}_{\mathcal{E}}^\omega(q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) - h_{\mathcal{E}}(q) \right) \Rightarrow \mathbb{G}[\omega_i \{ \zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q) \}] \quad \text{in } \ell^\infty(\mathbb{S}^2),$$

where $\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta})$ and $\widehat{h}_{\mathcal{E}}^\omega(q; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}})$ are defined in (40) and (55), respectively.

PROOF: The proof has two parts. We first show in part 1 that

$$\sup_{q \in \mathbb{S}^2} \left| \sqrt{n}(\widehat{h}_{\mathcal{E}}^{\omega}(q, \widehat{\mathbf{L}}^{\omega}; \widehat{\boldsymbol{\eta}}) - h_{\mathcal{E}}(q)) - \mathbb{G}_n[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] \right| = o_p(1). \quad (56)$$

We then use [van der Vaart and Wellner \(1996, Thm. 2.5.2\)](#) to conclude in part 2 that

$$\mathbb{G}_n[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] \Rightarrow \mathbb{G}[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] \quad \text{in } \ell^{\infty}(\mathbb{S}^2). \quad (57)$$

Part 1: Our goal is to establish (56). Define

$$\widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}} \right) \zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}).$$

By [Lemma B.10](#), we have

$$\sup_{q \in \mathbb{S}^2} \left| \sqrt{n}(\widehat{h}_{\mathcal{E}}^{\omega}(q, \widehat{\mathbf{L}}^{\omega}; \widehat{\boldsymbol{\eta}}) - \widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta})) \right| = o_p(1).$$

This implies that (56) holds if the equation below holds,

$$\sup_{q \in \mathbb{S}^2} \left| \sqrt{n}(\widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)) - \mathbb{G}_n[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] \right| = o_p(1). \quad (58)$$

Now, we rewrite the left-hand side of (58) using the identity $\mathbb{G}_n[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] = \bar{\omega} \sqrt{n}(\widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q))$. That is

$$\left| 1 - \frac{1}{\bar{\omega}} \right| \sup_{q \in \mathbb{S}^2} \left| \mathbb{G}_n[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] \right|.$$

Since $|1 - \frac{1}{\bar{\omega}}| = o_p(1)$ by the definition of the weights ω_i and the law of large numbers. Therefore, (58) holds if $\sup_{q \in \mathbb{S}^2} \left| \mathbb{G}_n[\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}] \right| = O_p(1)$, which holds by (57) and the Continuous Mapping Theorem. Note that we can use (57) since the proof of part 2 is independent of the proof of part 1. This completes the proof of part 1.

Part 2: Our goal is to establish (57). Consider the random vector

$$W_i = (\mathbf{L}_{0,i}, \Delta \mathbf{L}_i, Z_i, \pi(X_i), \Delta \boldsymbol{\theta}(X_i), \boldsymbol{\theta}_0(X_i), \omega_i) \in \mathcal{W}$$

and the class of functions $\mathcal{Q} \equiv \{f : \mathcal{W} \rightarrow \mathbb{R} : f(W_i) = \omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}, q \in \mathbb{S}^2\}$. Here, we are using the true value of the nuisance function $\boldsymbol{\eta}$ —which is fixed—and treat-

ing $\boldsymbol{\eta}(X_i)$ as a random vector. In order to use [van der Vaart and Wellner \(1996, Thm. 2.5.2\)](#) to complete the proof of part 2, we need to verify that (i) an envelope function of the class \mathcal{Q} has finite second-moment, and (ii) the uniform entropy bound in [van der Vaart and Wellner \(1996, \(2.5.1\)\)](#) holds. To verify (i), we note that the minimum envelope $\sup_{q \in \mathcal{S}^2} |\omega_i \{\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}) - h_{\mathcal{E}}(q)\}|$ has finite second-moment due to Assumptions 1 and 6. To verify (ii), we rely on [Andrews \(1994, Thms. 2 and 3\)](#) that states that the class of functions of type I and their mixing (by addition and product operation) verify the uniform entropy bound. We conclude by noting that \mathcal{Q} is of type I, since any function in \mathcal{Q} can be written as the sum of (i) linear functions in q , (ii) indicators of linear functions in q , and (iii) product of (i) and (ii). *Q.E.D.*

PROOF OF THEOREM 4.1: It follows from Theorem A.1 by taking $\omega_i = 1, \forall i$. *Q.E.D.*

A.4. Proofs for Section 5.1

PROOF OF PROPOSITION 5.1: Theorem 2.2 yields $T^{\text{LDA}}(\varepsilon) = 0 \iff \varepsilon \in \mathcal{F}$. Next, we show that for $\varepsilon \in \mathcal{E}$, $\psi_{\mathcal{P}\mathcal{F}}(\varepsilon) = 0 \iff \varepsilon \in \mathcal{P}\mathcal{F}$. Indeed, for every $\alpha \in [0, 1]$,

$$q(\alpha)^\top \varepsilon + h_{\mathcal{E}}(-q(\alpha)) = q(\alpha)^\top \varepsilon - \inf_{\eta \in \mathcal{E}} q(\alpha)^\top \eta \geq 0.$$

So, $\psi_{\mathcal{P}\mathcal{F}}(\varepsilon) = 0$ if and only if there exists $\alpha \in [0, 1]$ with $q(\alpha)^\top \varepsilon = \inf_{\eta \in \mathcal{E}} q(\alpha)^\top \eta$, i.e., $\varepsilon = \arg \min_{\eta \in \mathcal{E}} q(\alpha)^\top \eta$. By definition, this is equivalent to $\varepsilon \in \mathcal{P}\mathcal{F}$. Since $\{\varepsilon \in \mathbb{B}_C : T^{\text{LDA}}(\varepsilon) = 0\} = \mathcal{F}$, we have $\sup_{\varepsilon \in \mathbb{B}_C : T^{\text{LDA}}(\varepsilon) = 0} \psi_{\mathcal{P}\mathcal{F}}(\varepsilon) = \sup_{\varepsilon \in \mathcal{F}} \psi_{\mathcal{P}\mathcal{F}}(\varepsilon)$. Because $\psi_{\mathcal{P}\mathcal{F}}(\varepsilon) \geq 0$ for every $\varepsilon \in \mathcal{E}$, the right hand side equals 0 if and only if $\psi_{\mathcal{P}\mathcal{F}}(\varepsilon) = 0$ for all $\varepsilon \in \mathcal{F}$.

For $v \neq 0$, by positive homogeneity of $h_{\mathcal{E}}(v)$, $h_{\mathcal{E}}(v) = \|v\| h_{\mathcal{E}}\left(\frac{v}{\|v\|}\right)$, so that, for every $\alpha \in [0, 1]$ and $s \in [-\bar{s}, \bar{s}]$, $\sqrt{1+s^2} h_{\mathcal{E}}\left(\frac{-q(\alpha)+su_3}{\sqrt{1+s^2}}\right) = h_{\mathcal{E}}(-q(\alpha) + su_3)$ for $u_3 = [0, 0, 1]^\top$. Indeed, $\| -q(\alpha) + su_3 \| = \sqrt{1+s^2}$, since $q(\alpha)_3 = 0$ and $\|q(\alpha)\| = 1$.

Fix $\alpha \in [0, 1]$, set $q = q(\alpha)$, and consider the scalar function

$$g_\alpha(s) \equiv h_{\mathcal{E}}(-q + su_3) = \sup_{\eta \in \mathcal{E}} (-q + su_3)^\top \eta.$$

The function g_α is convex because it is the supremum of affine functions of s . By Assumptions 1-2, \mathcal{E} is compact and strictly convex, so each support set is a singleton. Hence the support point $\mathcal{S}_\mathcal{E}(-q) = \arg \max_{\eta \in \mathcal{E}} (-q)^\top \eta$ is well-defined.

We next show that g_α is differentiable at zero and that $g'_\alpha(0) = u_3^\top \mathcal{S}_\mathcal{E}(-q)$. Let $\varepsilon^* = \mathcal{S}_\mathcal{E}(-q)$. For $s \neq 0$, define the (unique) support point $\varepsilon^s \equiv \arg \max_{\eta \in \mathcal{E}} (-q + su_3)^\top \eta$. For $s > 0$, optimality of ε^* at $s = 0$ and optimality of ε^s at s give

$$(\varepsilon^*)_3 \leq \frac{g_\alpha(s) - g_\alpha(0)}{s} \leq (\varepsilon^s)_3.$$

For $s < 0$, the same inequalities hold with the order reversed:

$$(\varepsilon^s)_3 \leq \frac{g_\alpha(s) - g_\alpha(0)}{s} \leq (\varepsilon^*)_3.$$

The support map is continuous by Berge's maximum theorem, so $\varepsilon^s \rightarrow \varepsilon^*$ as $s \rightarrow 0$. Therefore, $g'_\alpha(0) = (\varepsilon^*)_3 = u_3^\top \mathcal{S}_\mathcal{E}(-q)$. Since g_α is convex and differentiable at zero, and since zero is an interior point of $[-\bar{s}, \bar{s}]$,

$$g_\alpha(0) = \inf_{s \in [-\bar{s}, \bar{s}]} g_\alpha(s) \iff g'_\alpha(0) = 0.$$

Combining this with the derivative calculation yields

$$h_\mathcal{E}(-q(\alpha)) = \inf_{s \in [-\bar{s}, \bar{s}]} \sqrt{1 + s^2} h_\mathcal{E} \left(\frac{-q(\alpha) + su_3}{\sqrt{1 + s^2}} \right) \iff u_3^\top \mathcal{S}_\mathcal{E}(-q(\alpha)) = 0.$$

The expression inside the supremum in (41) is nonnegative, as $s = 0$ is in the interval $[-\bar{s}, \bar{s}]$. Hence, $\Delta_{\mathcal{P}\mathcal{F}}^h(\bar{s}) = 0$ if and only if $u_3^\top \mathcal{S}_\mathcal{E}(-q(\alpha)) = 0$ for every $\alpha \in [0, 1]$. By Lemma B.3, $\mathcal{P}\mathcal{F} = \{\mathcal{S}_\mathcal{E}(-q(\alpha)) : \alpha \in [0, 1]\}$, so $\Delta_{\mathcal{P}\mathcal{F}}^h(\bar{s}) = 0$ if and only if $\mathcal{P}\mathcal{F} \subset \{\varepsilon \in \mathcal{E} : \varepsilon_3 = 0\}$. Under Assumption 3, Theorem 2.3 gives $\mathcal{P}\mathcal{F} = \mathcal{F}$ if and only if $\mathcal{P}\mathcal{F} \subset \{\varepsilon \in \mathcal{E} : \varepsilon_3 = 0\}$. Combining the last two equivalences proves $\mathcal{P}\mathcal{F} = \mathcal{F} \iff \Delta_{\mathcal{P}\mathcal{F}}^h(\bar{s}) = 0$. *Q.E.D.*

A.5. Proofs for Section 5.2

Let $[\cdot]_+ = \max\{\cdot, 0\}$. Define $\phi : \ell^\infty(\mathbb{S}^2) \times \mathbb{R}^3 \rightarrow \mathbb{R}$ as

$$\phi(f, \varepsilon) \equiv \left[\max_{q \in \mathbb{S}^2} (q^\top \varepsilon - f(q)) \right]_+ + \left[\min_{q \in \tilde{\mathbb{S}}^2} (q_1 \varepsilon_1 + q_2 \varepsilon_2 + |q_3 \varepsilon_3| + f(-q)) \right]_+, \quad (59)$$

where $f \in \ell^\infty(\mathbb{S}^2)$ and $\varepsilon \in \mathbb{R}^3$. Note that $T_n^{\text{LDA}} = \sqrt{n} \left(\phi(\widehat{\beta}^\omega) - \phi(\beta) \right)$.

PROOF OF THEOREM 5.1: The proof has two parts. In the first part, we use [Fang and Santos \(2019, Thms. 2.1 and 3.2\)](#) to show that the Bayesian bootstrap is consistent. In the second part, we use [Romano and Shaikh \(2012, Lemma A.1\)](#) to prove that the test φ_n^{LDA} has asymptotically correct size.

Part 1: By [Kaido \(2016, Lemma D.3\)](#) and standard arguments, it follows that ϕ defined in (59) is Hadamard directionally differentiable at β tangentially to $\mathbb{D}_0 = C(\mathbb{S}^2) \times \mathbb{R}^3 \subset \mathbb{D} = \ell^\infty(\mathbb{S}^2) \times \mathbb{R}^3$, where $C(\mathbb{S}^2)$ is the class of continuous functions on \mathbb{S}^2 . This guarantees Assumption 1 in [Fang and Santos \(2019\)](#). By Cramer-Wold and Theorems [A.1](#) and [B.1](#), we have $\sqrt{n}(\widehat{\beta}^\omega - \beta) \Rightarrow \mathbb{G}_\beta$. Let $K(q, q')$ be the covariance matrix of \mathbb{G}_β for $q, q' \in \mathbb{S}^2$. It can be show that $K(q, q')$ is continuous in both $q, q' \in \mathbb{S}^2$, which implies that $\mathbb{G}_\beta \in \mathbb{D}_0$ with probability 1 by [Adler and Taylor \(2007, Lemma 1.3.1\)](#). This guarantees Assumption 2 in [Fang and Santos \(2019\)](#). By [Fang and Santos \(2019, Thm. 2.1\)](#), we conclude $\sqrt{n} \left(\phi(\widehat{\beta}^\omega) - \phi(\beta) \right) \Rightarrow \phi'_\beta(\mathbb{G}_\beta)$.

To prove the consistency of the bootstrap, it is sufficient to show that

$$\sup_{f \in \mathcal{BL}_1} \left| \mathbb{E} \left[f(\widehat{\phi}'_\beta(\sqrt{n}\{\widetilde{\beta}^\omega - \widehat{\beta}^\omega\})) \mid \{(Y_i, G_i, X_i, Z_i, \Xi_i)\}_{i=1}^n \right] - \mathbb{E} \left[f(\phi'_\beta(\mathbb{G}_\beta)) \right] \right| = o_p(1), \quad (60)$$

where \mathcal{BL}_1 is the set of 1-Lipschitz functions $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $|f|_\infty \leq 1$.

It is sufficient to verify Assumptions 3 and 4 in [Fang and Santos \(2019\)](#) to conclude (60) by [Fang and Santos \(2019, Thm. 3.2\)](#). Note that Assumption 3 part (i), (iii), and (iv) hold by construction. By [Hong and Li \(2018, Thm. 3.1\)](#), we conclude $\widehat{\phi}'_\beta$ defined in Procedure 1 verifies Assumption 4 in [Fang and Santos \(2019\)](#). Therefore, it is sufficient to verify Assumption 3 part (ii), i.e., the Bayesian bootstrap works for the estimator $\widehat{\beta}^\omega$:

$$\sup_{f \in \mathcal{BL}_1} \left| \mathbb{E} \left[f \left(\sqrt{n} \left(\widetilde{\beta}^\omega - \widehat{\beta}^\omega \right) \right) \mid \{(Y_i, G_i, X_i, Z_i, \Xi_i)\}_{i=1}^n \right] - \mathbb{E} \left[f(\mathbb{G}_\beta) \right] \right| = o_p(1). \quad (61)$$

Define

$$\widehat{\beta}^{\omega,*} \equiv \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i^h}{\omega^h} \right) \zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}), \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i^\varepsilon}{\omega^\varepsilon} \right) \xi_i^*(a^*; \mathbf{L}, \boldsymbol{\eta}) \right) \quad (62)$$

$$\widetilde{\beta}^{\omega,*} \equiv \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{W_i \omega_i^h}{W^h} \right) \zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}), \frac{1}{n} \sum_{i=1}^n \left(\frac{W_i \omega_i^\varepsilon}{W^\varepsilon} \right) \xi_i^*(a^*; \mathbf{L}, \boldsymbol{\eta}) \right) \quad (63)$$

By Lemma B.14, we have $\left\| \sqrt{n} \left(\widetilde{\beta}^{\omega,*} - \widehat{\beta}^{\omega,*} \right) - \sqrt{n} \left(\beta^{\omega,*} - \widehat{\beta}^{\omega,*} \right) \right\| = o_p(1)$; therefore, we conclude that (61) holds if the following equation holds

$$\sup_{f \in \mathcal{B}\mathcal{L}_1} \left| \mathbb{E} \left[f \left(\sqrt{n} \left(\widetilde{\beta}^{\omega,*} - \widehat{\beta}^{\omega,*} \right) \right) \mid \{(Y_i, G_i, X_i, Z_i, \Xi_i)\}_{i=1}^n \right] - \mathbb{E}[f(\mathbb{G}_\beta)] \right| = o(1).$$

Finally, the previous equation holds by definition of $\widetilde{\beta}^{\omega,*}$, $\widehat{\beta}^{\omega,*}$, by [van der Vaart and Wellner \(1996, Thm. 3.6.13\)](#), and because $\sqrt{n} \left(\widehat{\beta}^{\omega,*} - \beta \right) \Rightarrow \mathbb{G}_\beta$ since $E[\omega^h] = E[\omega^\varepsilon] = 1$.

This completes the proof of the bootstrap consistency.

Part 2: For an arbitrarily small $\kappa > 0$, $\mathbb{P} \left(\sup_{x \in \mathbb{R}} \widetilde{F}_n(x) - F_\infty(x) > \kappa \right) = o(1)$, where

$$\widetilde{F}_n(x) = \mathbb{P} \left(\widehat{\phi}'_\beta(\sqrt{n} \{ \widetilde{\beta}^{\omega,*} - \widehat{\beta}^{\omega,*} \}) \leq x \mid \{(Y_i, G_i, X_i, Z_i, \Xi_i)\}_{i=1}^n \right), \quad F_\infty(x) = \mathbb{P} \left(\phi'_\beta(\mathbb{G}_\beta) \leq x \right).$$

Let $c_{1-\alpha}$ be $(1-\alpha)$ -quantile of $\phi'_\beta(\mathbb{G}_\beta)$, i.e., $F_\infty(c_{1-\alpha}) \geq 1-\alpha$. Recall that $\widehat{c}_{1-\alpha+\kappa}$ is the $(1-\alpha+\kappa)$ -quantile of the distribution of $\widehat{\phi}'_\beta(\sqrt{n} \{ \widetilde{\beta}^{\omega,*} - \widehat{\beta}^{\omega,*} \})$. The following derivation completes the proof of part 2:

$$\begin{aligned} \mathbb{P}(T_n^{\text{LDA}} > \widehat{c}_{1-\alpha+\kappa} + \kappa) &\stackrel{(1)}{\leq} \mathbb{P}(T_n^{\text{LDA}} > c_{1-\alpha} + \kappa) + o(1) \stackrel{(2)}{\leq} \mathbb{P}(T_n^{\text{LDA}} \geq c_{1-\alpha} + \kappa/2) + o(1) \\ &\stackrel{(3)}{\leq} \mathbb{P}(\phi'_\beta(\mathbb{G}_\beta) \geq c_{1-\alpha} + \kappa/2) + o(1) \\ &\stackrel{(4)}{\leq} \mathbb{P}(\phi'_\beta(\mathbb{G}_\beta) > c_{1-\alpha}) + o(1) \stackrel{(5)}{\leq} \alpha + o(1), \end{aligned}$$

where (1) holds since $\mathbb{P}(\widehat{c}_{1-\alpha+\kappa} \geq c_{1-\alpha}) = 1 - o(1)$ by [Romano and Shaikh \(2012, Lemma A.1\)](#), (2) and (4) hold since $\kappa > 0$, (3) holds since $T_n^{\text{LDA}} \xrightarrow{d} \phi'_\beta(\mathbb{G}_\beta)$, and (5) holds by the definition of $c_{1-\alpha}$. *Q.E.D.*

APPENDIX B: AUXILIARY RESULTS

B.1. Lemmas and Auxiliary Results for Section 2

LEMMA B.1—Support function of $\mathcal{C}(\varepsilon^*)$: For $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*) \in \mathcal{E}$ and $q \in \mathbb{R}^3$, the support function of the closed convex set $\mathcal{C}(\varepsilon^*)$ in (12) is as given in (13).

PROOF: We have $h_{\mathcal{C}(\varepsilon^*)}(q) = \sup_{\varepsilon \in \mathcal{C}(\varepsilon^*)} q^\top \varepsilon = \sup_{\substack{\varepsilon_1 \leq \varepsilon_1^*, \varepsilon_2 \leq \varepsilon_2^* \\ |\varepsilon_3| \leq |\varepsilon_3^*|}} (q_1 \varepsilon_1 + q_2 \varepsilon_2 + q_3 \varepsilon_3)$. If $q_1 < 0$, then $\sup_{\varepsilon_1 \leq \varepsilon_1^*} q_1 \varepsilon_1 = +\infty$ (taking $\varepsilon_1 \rightarrow -\infty$). Similarly if $q_2 < 0$. For $q_1, q_2 \geq 0$, the supremum over $\varepsilon_j \leq \varepsilon_j^*$ gives $q_j \varepsilon_j^*$, $j = 1, 2$. For the third coordinate:

$$\sup_{|\varepsilon_3| \leq |\varepsilon_3^*|} q_3 \varepsilon_3 = \begin{cases} q_3 |\varepsilon_3^*| & \text{if } q_3 \geq 0, \\ -q_3 |\varepsilon_3^*| = |q_3| |\varepsilon_3^*| & \text{if } q_3 < 0. \end{cases}$$

Thus $h_{\mathcal{C}(\varepsilon^*)}(q) = q_1 \varepsilon_1^* + q_2 \varepsilon_2^* + |q_3| |\varepsilon_3^*|$ for $q_1, q_2 \geq 0$.

Q.E.D.

LEMMA B.2: Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be the probability space on which (Y, G, X) are defined. Then: (i) \mathcal{E} is convex. (ii) Under Assumption 1, \mathcal{E} is compact. (iii) Under Assumptions 1-2, \mathcal{E} has a non-empty interior in \mathbb{R}^3 , and hence $\dim(\mathcal{E}) = 3$.

PROOF: Recall that $\mathcal{A} \equiv \{a : \mathcal{X} \rightarrow [0, 1] \text{ Borel-measurable}\}$.

Part (i). Each coordinate map $a \mapsto e_g^l(a)$ is affine in a by (3). Since \mathcal{A} is convex and $(e_r^A(a), e_b^A(a), e_r^F(a) - e_b^F(a))$ is an affine image of a , its range \mathcal{E} is convex.

Part (ii). Define

$$K \equiv \{u \in L^\infty(\Omega, \mathfrak{F}, \mathbb{P}) : u \text{ is } \sigma(X)\text{-measurable and } 0 \leq u \leq 1 \text{ a.s.}\}.$$

As $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is standard Borel, the Doob–Dynkin lemma (Rao and Swift, 2006, Prop. 3, Ch. 1) yields $K = \{a(X) : a \in \mathcal{A}\}$ up to a.s. equality. Hence, $\mathcal{E} = \mathbb{E}[\theta_0(X)] + T(K)$, for

$$T(u) \equiv \mathbb{E}[u \Delta \theta(X)] \in \mathbb{R}^3. \tag{64}$$

We first show that K is weak-* compact in L^∞ , where L^∞ is equipped with the topology $\sigma(L^\infty, L^1)$. Since $0 \leq u \leq 1$ a.s. for every $u \in K$, we have $\|u\|_\infty \leq 1$, so K is contained in

the closed unit ball of L^∞ . Since $L^\infty = (L^1)^*$, the Banach-Alaoglu theorem yields that the closed unit ball of L^∞ is compact for the weak-* topology $\sigma(L^\infty, L^1)$.

To prove weak-* compactness of K , it suffices to show that K is weak-* closed, since K is contained in the weak-* compact unit ball of L^∞ . Accordingly, let $u_\alpha \in K$ and suppose $u_\alpha \rightarrow u$ weak-* in L^∞ . We show that $u \in K$. First, u is $\sigma(X)$ -measurable. Indeed, for every $g \in L^1(\Omega, \mathfrak{F}, \mathbb{P})$, $\mathbb{E}[u_\alpha(g - \mathbb{E}[g|\sigma(X)])] = 0$, because u_α is $\sigma(X)$ -measurable. Passing to the limit using weak-* convergence gives $\mathbb{E}[u(g - \mathbb{E}[g|\sigma(X)])] = 0$, for all $g \in L^1$. Taking $g = \mathbb{1}_A$ for $A \in \mathfrak{F}$, we obtain

$$\int_A u d\mathbb{P} = \int_A \mathbb{E}[u|\sigma(X)] d\mathbb{P}, \quad \forall A \in \mathfrak{F},$$

hence $u = \mathbb{E}[u|\sigma(X)]$ a.s. Therefore u is $\sigma(X)$ -measurable.

Next, the bounds $0 \leq u \leq 1$ a.s. are preserved. To show $u \geq 0$ a.s., suppose otherwise. Then $\mathbb{P}(u < 0) > 0$, so for some $\varepsilon > 0$, $A_\varepsilon \equiv \{u \leq -\varepsilon\}$ has positive probability. Taking $g = \mathbb{1}_{A_\varepsilon} \in L^1$, weak-* convergence yields $\mathbb{E}[u_\alpha g] \rightarrow \mathbb{E}[u g]$. Since $u_\alpha \geq 0$ a.s., we have $\mathbb{E}[u_\alpha g] \geq 0$ for all α , whereas $\mathbb{E}[u g] = \mathbb{E}[u \mathbb{1}_{A_\varepsilon}] \leq -\varepsilon \mathbb{P}(A_\varepsilon) < 0$, a contradiction. Thus $u \geq 0$ a.s. Applying the same argument to $1 - u_\alpha$ shows $u \leq 1$ a.s. Therefore $u \in K$, so K is weak-* closed. Hence K is weak-* compact.

Now consider the map $T : K \rightarrow \mathbb{R}^3$ in (64). This map is well-defined because $u \in L^\infty$ with $|u| \leq 1$ a.s. and each component of $\Delta\theta(X)$ is in L^1 by Assumption 1, so each component of $u\Delta\theta(X)$ is in L^1 . Moreover, for each component $j = 1, 2, 3$, the map from u to the component of $\mathbb{E}[u\Delta\theta(X)]$ is weak-* continuous on L^∞ , since each component of $\Delta\theta(X)$ is in L^1 . Therefore $T : K \rightarrow \mathbb{R}^3$ is continuous when K is endowed with the weak-* topology and \mathbb{R}^3 with its usual Euclidean topology. Since K is weak-* compact and T is continuous, $T(K)$ is compact in \mathbb{R}^3 . Therefore, $\mathcal{E} = \mathbb{E}[\theta_0(X)] + T(K)$ is compact.

Part (iii). Let $\mathcal{A}_{\text{disp}} \equiv \{\varepsilon : \mathcal{X} \rightarrow \mathbb{R} \text{ measurable} : |\varepsilon(X)| \leq 1/2 \text{ } P_X\text{-a.s.}\} = \{\varepsilon : \mathcal{X} \rightarrow \mathbb{R} \text{ measurable} : \|\varepsilon\|_\infty \leq 1/2\}$ where $\|\varepsilon\|_\infty \equiv \text{ess sup}_{x \sim P_X} |\varepsilon(x)|$. For any algorithm $a \in \mathcal{A}$, let $\varepsilon(X) \equiv a(X) - \frac{1}{2}$ be the corresponding displacement in $\mathcal{A}_{\text{disp}}$. Let $e^0 = \mathbb{E}[\theta_0(X)] + \frac{1}{2}\mathbb{E}[\Delta\theta(X)]$. Then for any $a \in \mathcal{A}$ we have $e(a) - e^0 = \mathbb{E}[\varepsilon(X)\Delta\theta(X)]$. Hence, we can

write $\mathcal{E} = e^0 + \mathcal{E}^0$, where \mathcal{E}^0 is the set of feasible displacements:

$$\mathcal{E}^0 \equiv \{ \mathbb{E}[\varepsilon(X)\Delta\boldsymbol{\theta}(X)], \varepsilon : \mathcal{X} \rightarrow [-\frac{1}{2}, \frac{1}{2}] \text{ measurable} \},$$

and the equality holds because every measurable $a : \mathcal{X} \rightarrow [0, 1]$ corresponds to a unique $\varepsilon \in [-\frac{1}{2}, \frac{1}{2}]$ and vice versa. We next show that \mathcal{E}^0 contains a Euclidean ball around 0, which implies $\mathcal{E} = e^0 + \mathcal{E}^0$ contains a ball around e^0 . Fix $q \in \mathbb{R}^3$ with $\|q\| = 1$. Consider the support function of \mathcal{E}^0 in direction q :

$$h_{\mathcal{E}^0}(q) \equiv \sup_{e \in \mathcal{E}^0} q^\top e = \sup_{\|\varepsilon\|_\infty \leq 1/2} \mathbb{E}[\varepsilon(X)q^\top \Delta\boldsymbol{\theta}(X)].$$

Pointwise maximization yields the maximizer $\varepsilon_q(X) = \frac{1}{2} \text{sign}(q^\top \Delta\boldsymbol{\theta}(X))$ (with any value in $[-\frac{1}{2}, \frac{1}{2}]$ on the null set where $q^\top \Delta\boldsymbol{\theta}(X) = 0$), so

$$h_{\mathcal{E}^0}(q) = \frac{1}{2} \mathbb{E}[|q^\top \Delta\boldsymbol{\theta}(X)|]. \quad (65)$$

By Assumption 2, there exists a constant $C > 0$ such that for all $\delta > 0$, $\mathbb{P}(|q^\top \Delta\boldsymbol{\theta}(X)| \leq \delta) \leq C\delta^m$, uniformly in $q \in \mathbb{S}^2 \equiv \{q \in \mathbb{R}^3 : \|q\| = 1\}$. Choose $\delta_0 \equiv (1/(2C))^{1/m} > 0$. Then for every $q \in \mathbb{S}^2$, $\mathbb{P}(|q^\top \Delta\boldsymbol{\theta}(X)| > \delta_0) \geq 1 - C\delta_0^m = \frac{1}{2}$, and hence $\mathbb{E}[|q^\top \Delta\boldsymbol{\theta}(X)|] \geq \delta_0 \mathbb{P}(|q^\top \Delta\boldsymbol{\theta}(X)| > \delta_0) \geq \frac{\delta_0}{2}$. Combining with (65) yields the uniform bound

$$h_{\mathcal{E}^0}(q) \geq \frac{\delta_0}{4} \quad \forall q \in \mathbb{S}^2. \quad (66)$$

Hence, (66) implies that \mathcal{E}^0 contains the Euclidean ball $\mathbb{B}(0, r)$ with radius $r = \delta_0/4$ (see, e.g., Schneider, 1993, Section 1.7). Therefore, $e^0 + \mathbb{B}(0, r) \subseteq \mathcal{E}$ so \mathcal{E} has a nonempty interior in \mathbb{R}^3 , and hence $\dim(\mathcal{E}) = 3$. *Q.E.D.*

B.1.1. Characterization of the Geometry of \mathcal{F} using the Support-Set of \mathcal{E}

Let Assumptions 1-2 hold and let $\mathcal{S}_{\mathcal{E}}(q)$ be the support-set of $\mathcal{E} \subset \mathbb{R}^3$,

$$\mathcal{S}_{\mathcal{E}}(q) = \arg \max_{\varepsilon \in \mathcal{E}} q^\top \varepsilon, \quad q \in \mathbb{S}^2.$$

Since \mathcal{E} is compact and strictly convex, $\mathcal{S}_{\mathcal{E}}(q)$ is a well-defined continuous function.

LEMMA B.3: Let \mathcal{PF} and \mathcal{F} be the Pareto and FA frontiers defined in (14) and (8), respectively, and let $\tilde{\mathbb{S}}^2 \equiv \{q \in \mathbb{S}^2 : q_1 \geq 0, q_2 \geq 0\}$. Under Assumptions 1-2, we have

1. $\mathcal{PF} = \{\varepsilon = \mathcal{S}_{\mathcal{E}}(-q) : q \in \tilde{\mathbb{S}}^2, q_3 = 0\}$
2. $\mathcal{F} = \{\varepsilon = \mathcal{S}_{\mathcal{E}}(-q) : q \in \tilde{\mathbb{S}}^2, q_3 \varepsilon_3 \geq 0\}$

In particular, $\mathcal{PF} \subseteq \mathcal{F}$.

PROOF: The proof has two parts. We first characterize the Pareto and FA frontiers using support functions. We then use the characterization to conclude the lemma.

Part 1: For $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*) \in \mathcal{E}$, we define

$$\mathcal{C}^P(\varepsilon^*) \equiv \{\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \in \mathbb{R}^3 : \varepsilon_1 \leq \varepsilon_1^*, \varepsilon_2 \leq \varepsilon_2^*\}, \quad (67)$$

the set of group-wise expected accuracy losses that are weakly more accurate than ε^* . It can be verified that the support function of $\mathcal{C}^P(\varepsilon^*)$ equals:

$$h_{\mathcal{C}^P(\varepsilon^*)}(q) \equiv \sup_{\varepsilon \in \mathcal{C}^P(\varepsilon^*)} q^\top \varepsilon = \begin{cases} q_1 \varepsilon_1^* + q_2 \varepsilon_2^* & \text{if } q_1 \geq 0, q_2 \geq 0, q_3 = 0 \\ +\infty & \text{otherwise.} \end{cases} \quad (68)$$

Let $\varepsilon^* = (\varepsilon_1^*, \varepsilon_2^*, \varepsilon_3^*) \in \mathcal{E}$, then $\varepsilon^* \in \mathcal{PF}$ if and only if

$$\min_{q \in \tilde{\mathbb{S}}^2, q_3=0} \left(h_{\mathcal{C}^P(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q) \right) = 0, \quad (69)$$

and $\varepsilon^* \in \mathcal{F}$ if and only if

$$\min_{q \in \tilde{\mathbb{S}}^2} \left(h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q) \right) = 0, \quad (70)$$

where (69) follows by similar arguments to those in the proof of Theorem 2.2 (i)-(iii), and (70) is due to Theorem 2.2 (iii).

Part 2: We use (68) and the support set $\mathcal{S}_{\mathcal{E}}$ to note that (69) can be written as follows

$$\min_{q \in \tilde{\mathbb{S}}^2, q_3=0} (q_1 \varepsilon_1^* + q_2 \varepsilon_2^* - q^\top \mathcal{S}_{\mathcal{E}}(-q)) = 0.$$

Therefore, any $\varepsilon^* = \mathcal{S}_{\mathcal{E}}(-q)$ for $q \in \tilde{\mathbb{S}}^2$ and $q_3 = 0$ satisfies the previous equation, implying that $\{\varepsilon = \mathcal{S}_{\mathcal{E}}(-q) : q \in \tilde{\mathbb{S}}^2, q_3 = 0\} \subset \mathcal{PF}$. Similarly, for any ε^* that satisfies the previous

equation there exists $q \in \tilde{\mathbb{S}}^2$ such that $q_3 = 0$ and $q_1\varepsilon_1^* + q_2\varepsilon_2^* - q^\top \mathcal{S}_\mathcal{E}(-q) = 0$, which implies $\varepsilon^* = \mathcal{S}_\mathcal{E}(-q)$ and $\mathcal{PF} \subset \{\varepsilon = \mathcal{S}_\mathcal{E}(-q) : q \in \tilde{\mathbb{S}}^2, q_3 = 0\}$. This proves Lemma B.3(1).

We derive a convenient characterization of \mathcal{F} to prove Lemma B.3(2). Fix $\varepsilon^* \in \mathcal{E}$ and $q \in \tilde{\mathbb{S}}^2$. Since $h_{\mathcal{C}(\varepsilon^*)}(q) = q_1\varepsilon_1^* + q_2\varepsilon_2^* + |q_3||\varepsilon_3^*|$, and $h_\mathcal{E}(-q) = -\min_{\varepsilon \in \mathcal{E}} q^\top \varepsilon$, we obtain

$$h_{\mathcal{C}(\varepsilon^*)}(q) + h_\mathcal{E}(-q) = q_1\varepsilon_1^* + q_2\varepsilon_2^* + |q_3||\varepsilon_3^*| - \min_{\varepsilon \in \mathcal{E}} q^\top \varepsilon.$$

Since $\min_{\varepsilon \in \mathcal{E}} q^\top \varepsilon \leq q^\top \varepsilon^* = q_1\varepsilon_1^* + q_2\varepsilon_2^* + q_3\varepsilon_3^*$, it follows that

$$h_{\mathcal{C}(\varepsilon^*)}(q) + h_\mathcal{E}(-q) \geq |q_3||\varepsilon_3^*| - q_3\varepsilon_3^* \geq 0.$$

Therefore, $\varepsilon^* \in \mathcal{F} \iff \exists q \in \tilde{\mathbb{S}}^2$ such that $\varepsilon^* = \mathcal{S}_\mathcal{E}(-q)$ and $q_3\varepsilon_3^* \geq 0$, proving Lemma B.3(2). *Q.E.D.*

Lemma B.3 has two immediate implications. First, $\mathcal{PF} \subseteq \mathcal{F}$. Second, the Pareto frontier \mathcal{PF} is a one-dimensional object, while the FA-frontier can be a two-dimensional object.

Let $\tilde{\mathbb{S}}_+^2 \equiv \{q \in \tilde{\mathbb{S}}^2 : q_3 \geq 0\}$. Define the coordinate-edge image

$$\mathcal{IM}_{ij} \equiv \left\{ \arg \min_{\varepsilon \in \mathcal{E}} \{\alpha\varepsilon_i + (1-\alpha)\varepsilon_j\} : \alpha \in [0, 1] \right\}, \quad 1 \leq i < j \leq 3.$$

COROLLARY B.1: *Suppose Assumptions 1-2 hold and $\mathcal{E} \subset \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 \geq 0\}$. Then*

$$\mathcal{F} = \{\mathcal{S}_\mathcal{E}(-q) : q \in \tilde{\mathbb{S}}_+^2\}. \quad (71)$$

The images under $q \mapsto \mathcal{S}_\mathcal{E}(-q)$ of the three boundary edges of $\tilde{\mathbb{S}}_+^2$ are

$$\mathcal{IM}_{12} = \{\mathcal{S}_\mathcal{E}(-q) : q \in \tilde{\mathbb{S}}_+^2, q_3 = 0\} = \mathcal{PF},$$

$$\mathcal{IM}_{13} = \{\mathcal{S}_\mathcal{E}(-q) : q \in \tilde{\mathbb{S}}_+^2, q_2 = 0\}, \quad \mathcal{IM}_{23} = \{\mathcal{S}_\mathcal{E}(-q) : q \in \tilde{\mathbb{S}}_+^2, q_1 = 0\}.$$

If Assumption 3 also holds, then $\mathcal{IM}_{12}, \mathcal{IM}_{13}, \mathcal{IM}_{23}$ are the relative boundary edge curves of \mathcal{F} and $\mathcal{IM}_{12} = \mathcal{PF} \subsetneq \mathcal{F}$. The case $\mathcal{E} \subset \{\varepsilon_3 \leq 0\}$ follows by the reflection $(\varepsilon_1, \varepsilon_2, \varepsilon_3) \mapsto (\varepsilon_1, \varepsilon_2, -\varepsilon_3)$.

PROOF: If $q \in \tilde{\mathbb{S}}_+^2$, then for $\varepsilon = \mathcal{S}_{\mathcal{E}}(-q)$ we have $\varepsilon_3 \geq 0$ and $q_3 \geq 0$, so $q_3\varepsilon_3 \geq 0$. Lemma B.3 yields $\varepsilon \in \mathcal{F}$.

Conversely, let $\varepsilon = \mathcal{S}_{\mathcal{E}}(-q) \in \mathcal{F}$ for some $q \in \tilde{\mathbb{S}}^2$, so that $q_3\varepsilon_3 \geq 0$. If $q_3 \geq 0$, then $q \in \tilde{\mathbb{S}}_+^2$. If $q_3 < 0$, then $\varepsilon_3 = 0$ and for any $\eta \in \mathcal{E}$,

$$q_1\varepsilon_1 + q_2\varepsilon_2 = q^\top \varepsilon \leq q^\top \eta = q_1\eta_1 + q_2\eta_2 + q_3\eta_3 \leq q_1\eta_1 + q_2\eta_2,$$

where the last inequality uses $q_3 < 0$ and $\eta_3 \geq 0$. The case $q_1 = q_2 = 0$ would force $\eta_3 = 0$ for all $\eta \in \mathcal{E}$, contradicting that \mathcal{E} has nonempty interior. Hence $q_1 + q_2 > 0$. After normalizing $[q_1, q_2, 0]^\top$, we obtain a vector in $\tilde{\mathbb{S}}_+^2$ exposing the same point ε . This proves (71).

The identities for \mathcal{IM}_{12} , \mathcal{IM}_{13} , and \mathcal{IM}_{23} follow because positive rescaling of a normal vector does not change the minimizers. For example, $\{q \in \tilde{\mathbb{S}}_+^2 : q_3 = 0\}$ consists exactly of positive normalizations of $[\alpha, 1 - \alpha, 0]^\top$ for $\alpha \in [0, 1]$. Thus its support image is \mathcal{IM}_{12} , which equals \mathcal{PF} by Lemma B.3. The other two edges follow by the same argument.

Under Assumption 3, the support map is injective. Hence the map $T : \tilde{\mathbb{S}}_+^2 \rightarrow \mathcal{F}$, $T(q) = \mathcal{S}_{\mathcal{E}}(-q)$, is injective. It is also continuous by Berge's maximum theorem, and it is surjective by the representation in (71). Therefore T is a continuous bijection from the compact space $\tilde{\mathbb{S}}_+^2$ onto the Hausdorff space $\mathcal{F} \subset \mathbb{R}^3$, and hence T is a homeomorphism. Thus, the images of the three boundary edges of $\tilde{\mathbb{S}}_+^2$ are the relative boundary edge curves of \mathcal{F} . In particular, $\mathcal{IM}_{12} = T\left(\{q \in \tilde{\mathbb{S}}_+^2 : q_3 = 0\}\right) = \mathcal{PF}$. Since $\{q \in \tilde{\mathbb{S}}_+^2 : q_3 = 0\}$ is a proper subset of $\tilde{\mathbb{S}}_+^2$, injectivity of T implies $\mathcal{IM}_{12} = \mathcal{PF} \subsetneq \mathcal{F}$. *Q.E.D.*

COROLLARY B.2: *Let Assumptions 1-2 hold. Suppose that $\mathcal{E} \cap \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 > 0\} \neq \emptyset$, $\mathcal{E} \cap \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 < 0\} \neq \emptyset$, and that either $\mathcal{PF} \subset \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 > 0\}$ or $\mathcal{PF} \subset \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 < 0\}$. Recall $\mathcal{E}^A \equiv \pi(\mathcal{E})$, and let*

$$\mathcal{PF}^A \equiv \left\{ c \in \mathcal{E}^A : \exists c' \in \mathcal{E}^A \text{ such that } c'_1 \leq c_1, c'_2 \leq c_2, \text{ with at least one strict inequality} \right\}.$$

Then $\mathcal{PF}^A \subsetneq \mathcal{F}^A$ and $\mathcal{PF} \subsetneq \mathcal{F}$.

PROOF: It is enough to prove the result when $\mathcal{PF} \subset \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 > 0\}$; the other case follows by applying the reflection $(\varepsilon_1, \varepsilon_2, \varepsilon_3) \mapsto (\varepsilon_1, \varepsilon_2, -\varepsilon_3)$.

Since \mathcal{E} is convex and contains points with positive and negative third coordinate, $\mathcal{E} \cap \{\varepsilon \in \mathbb{R}^3 : \varepsilon_3 = 0\} \neq \emptyset$. Hence, $\pi(\mathcal{E} \cap \{\varepsilon_3 = 0\})$ is a nonempty compact subset of \mathcal{E}^A . Choose $c^0 \in \arg \min_{c \in \pi(\mathcal{E} \cap \{\varepsilon_3 = 0\})} (c_1 + c_2)$. Then $c^0 \in \mathcal{E}^A$ and $d(c^0) = 0$ for $d(\cdot)$ defined in (5), because c^0 has a feasible lift with third coordinate equal to zero.

We first show that $c^0 \in \mathcal{F}^A$. Suppose not. Then there exists $c' \in \mathcal{E}^A$ such that

$$c'_1 \leq c_1^0, \quad c'_2 \leq c_2^0, \quad d(c') \leq d(c^0) = 0,$$

with at least one strict inequality. Since $d(c') \geq 0$, we have $d(c') = 0$. Therefore $c' \in \pi(\mathcal{E} \cap \{\varepsilon_3 = 0\})$. Moreover, the strict inequality cannot occur only in the fairness coordinate, because $d(c') = d(c^0) = 0$. Hence at least one of $c'_1 < c_1^0$ or $c'_2 < c_2^0$ holds, so $c'_1 + c'_2 < c_1^0 + c_2^0$, contradicting the definition of c^0 . Thus $c^0 \in \mathcal{F}^A$.

Next, $c^0 \notin \mathcal{P}\mathcal{F}^A$. Indeed, since $c^0 \in \pi(\mathcal{E} \cap \{\varepsilon_3 = 0\})$, the point $\varepsilon^0 \equiv (c_1^0, c_2^0, 0)$ belongs to \mathcal{E} . If $c^0 \in \mathcal{P}\mathcal{F}^A$, then $\varepsilon^0 \in \mathcal{P}\mathcal{F}$: otherwise there would exist $\eta \in \mathcal{E}$ with $\eta_1 \leq \varepsilon_1^0$, $\eta_2 \leq \varepsilon_2^0$, and at least one strict inequality, and then $\pi(\eta)$ would Pareto-dominate c^0 in \mathcal{E}^A . This contradicts $c^0 \in \mathcal{P}\mathcal{F}^A$. Therefore $\varepsilon^0 \in \mathcal{P}\mathcal{F}$, contradicting the maintained case $\mathcal{P}\mathcal{F} \subset \{\varepsilon_3 > 0\}$. Hence $c^0 \notin \mathcal{P}\mathcal{F}^A$.

We have shown that $c^0 \in \mathcal{F}^A \setminus \mathcal{P}\mathcal{F}^A$. It remains to check that $\mathcal{P}\mathcal{F}^A \subseteq \mathcal{F}^A$. If $c \in \mathcal{P}\mathcal{F}^A$ but $c \notin \mathcal{F}^A$, then some $c' \in \mathcal{E}^A$ satisfies $c'_1 \leq c_1$, $c'_2 \leq c_2$, $d(c') \leq d(c)$, with at least one strict inequality. The strict inequality cannot occur only in the fairness coordinate: if $c'_1 = c_1$ and $c'_2 = c_2$, then $c' = c$, hence $d(c') = d(c)$. Thus at least one of $c'_1 < c_1$ or $c'_2 < c_2$ holds, contradicting $c \in \mathcal{P}\mathcal{F}^A$. Therefore $\mathcal{P}\mathcal{F}^A \subseteq \mathcal{F}^A$. Together with $c^0 \in \mathcal{F}^A \setminus \mathcal{P}\mathcal{F}^A$, this gives

$$\mathcal{P}\mathcal{F}^A \subsetneq \mathcal{F}^A.$$

Finally, because $c^0 \in \mathcal{F}^A$, $d(c^0) = 0$, and $\varepsilon^0 = (c_1^0, c_2^0, 0) \in \mathcal{E}$, Theorem 2.2 (iv) implies $\varepsilon^0 \in \mathcal{F}$. But $\varepsilon^0 \notin \mathcal{P}\mathcal{F}$, since $\varepsilon_3^0 = 0$ while $\mathcal{P}\mathcal{F} \subset \{\varepsilon_3 > 0\}$. Since Lemma B.3 gives $\mathcal{P}\mathcal{F} \subseteq \mathcal{F}$, we obtain $\mathcal{P}\mathcal{F} \subsetneq \mathcal{F}$. *Q.E.D.*

REMARK B.1—Implications for Figure 1: Lemma B.3 identifies $\mathcal{P}\mathcal{F}$ with support directions in $\tilde{\mathbb{S}}^2$ satisfying $q_3 = 0$, and identifies \mathcal{F} with support directions in $\tilde{\mathbb{S}}^2$ satisfying $q_3 \varepsilon_3 \geq 0$. Combining this characterization with Theorem 2.3 gives the classification used

in Figure 1. Under Assumption 3, $\mathcal{PF} = \mathcal{F} \iff \mathcal{PF} \subset \{\varepsilon \in \mathcal{E} : \varepsilon_3 = 0\}$. Thus, under Assumption 3, whenever $\mathcal{PF} \not\subset \{\varepsilon_3 = 0\}$, one has $\mathcal{PF} \subsetneq \mathcal{F}$, while if $\mathcal{PF} \subset \{\varepsilon_3 = 0\}$, the two frontiers coincide. In particular, a single intersection of \mathcal{PF} with the plane $\{\varepsilon_3 = 0\}$, or the fact that only the endpoints R and B lie in that plane, is not sufficient for $\mathcal{PF} = \mathcal{F}$. Corollary B.2 further shows that in the crossing case with \mathcal{PF} strictly on one side of $\{\varepsilon_3 = 0\}$, the strict inclusions $\mathcal{PF}^A \subsetneq \mathcal{F}^A$ and $\mathcal{PF} \subsetneq \mathcal{F}$ hold already under Assumptions 1-2 alone.

B.2. Lemmas, Auxiliary Results, and Notation for Section 3

Let $\Delta A(x) \equiv A_1(x) - A_0(x)$, so that $\Delta \theta(x; \lambda) \equiv \theta_1(x; \lambda) - \theta_0(x; \lambda) = \Delta A(x) - 2B(x)\lambda(x)$, and $\tau_q(x) \equiv \frac{1}{2}q^\top \Delta A(x)$. For each $x \in \mathcal{X}$ and for $q \in \mathbb{S}^2$, define

$$d_q(x) \equiv \inf_{s \in [0,1]^2} |q^\top \{\Delta A(x) - 2B(x)s\}|, \quad \mathcal{U}(x) \equiv \{B(x)s : s \in [0,1]^2\},$$

$$K_q^0(x) \equiv \{A_0(x) + u : u \in \mathcal{U}(x), q^\top u \leq \tau_q(x)\}, \quad K_q^1(x) \equiv \{A_1(x) - u : u \in \mathcal{U}(x), q^\top u \geq \tau_q(x)\},$$

$$K_q(x) \equiv K_q^0(x) \cup K_q^1(x).$$

Define the finite constants $M_0 \equiv \text{ess sup}_x \sup_{u \in \mathcal{U}(x)} \|\Delta A(x) - 2u\| < \infty$ and

$$H_0 \equiv \text{ess sup}_x \sup_{u \in \mathcal{U}(x)} \max\{\|A_0(x) + u\|, \|A_1(x) - u\|\} < \infty,$$

where the bounds follow from the binary-loss setup and the lower bound on μ_g , $g \in \{r, b\}$.

With the convention that $\sup\{\emptyset\} = -\infty$, define

$$\kappa_0(v, q; x) \equiv \sup_{u \in \mathcal{U}(x): q^\top u \leq \tau_q(x)} v^\top (A_0(x) + u), \quad (72)$$

$$\kappa_1(v, q; x) \equiv \sup_{u \in \mathcal{U}(x): q^\top u \geq \tau_q(x)} v^\top (A_1(x) - u). \quad (73)$$

Recall the fixed-completion feasible set $\mathcal{E}(\lambda)$ in (19) and its associated support set,

$$\mathcal{S}_{\mathcal{E}(\lambda)}(v) = \arg \max_{\eta \in \mathcal{E}(\lambda)} v^\top \eta.$$

Thus $\mathcal{S}_{\mathcal{E}(\lambda)}(-q)$ is the set of minimizers of $q^\top \eta$ over $\mathcal{E}(\lambda)$.

LEMMA B.4: *Let Assumptions 1 and 4 hold. Then, for every $\lambda \in \Lambda$, the set $\mathcal{E}(\lambda)$ is nonempty, compact, convex, has nonempty interior, and $\mathcal{S}_{\mathcal{E}(\lambda)}(v)$ is a singleton for every $v \in \mathbb{S}^2$. Moreover, $\mathcal{F}(\lambda) = \left\{ \varepsilon = \mathcal{S}_{\mathcal{E}(\lambda)}(-q) : q \in \tilde{\mathbb{S}}^2, q_3 \varepsilon_3 \geq 0 \right\}$.*

PROOF: By Assumption 4, $\mathbb{P}_X\{|q^\top \Delta \theta(X; \lambda)| \leq \delta\} \leq \mathbb{P}_X\{d_q(X) \leq \delta\} \lesssim \delta^m$ and $\mathbb{P}_X\{d_q(X) = 0\} = 0$ for all $q \in \mathbb{S}^2$. Hence, Assumptions 1-2 hold for the distribution of (Y^*, G, X, Z) implied by each $\lambda \in \Lambda$. By Lemma B.2 and Theorem 2.1, the set $\mathcal{E}(\lambda)$ has nonempty interior, is compact and convex, and $\mathcal{S}_{\mathcal{E}(\lambda)}(p)$ is a singleton for every $p \in \mathbb{S}^2$. By Lemma B.3, for every $\lambda \in \Lambda$, $\mathcal{F}(\lambda) = \left\{ \varepsilon = \mathcal{S}_{\mathcal{E}(\lambda)}(-q) : q \in \tilde{\mathbb{S}}^2, q_3 \varepsilon_3 \geq 0 \right\}$. *Q.E.D.*

LEMMA B.5: *Let Assumption 1 hold. Then, for every $q \in \tilde{\mathbb{S}}^2$, the set \mathcal{M}_q is nonempty, compact, and convex.*

PROOF: Since $u(x) \in \mathcal{U}(x) = B(x)[0, 1]^2$, for $j = 1, 2$, $u_j(x) = 0$ whenever $B_{jj}(x) = 0$, and $0 \leq u_j(x) \leq B_{jj}(x)$ whenever $B_{jj}(x) > 0$. Define

$$\lambda_j(x) = \begin{cases} u_j(x)/B_{jj}(x), & B_{jj}(x) > 0, \\ 0, & B_{jj}(x) = 0, \end{cases} \quad j = 1, 2.$$

Then $\lambda_j(x) \in [0, 1]$. Hence, for $u(x) \in \mathcal{U}(x)$, \mathbb{P}_X -a.s., with u measurable, there exists a measurable $\lambda \in \Lambda$, such that

$$B(x)\lambda(x) = u(x) \quad \mathbb{P}_X\text{-a.s.} \quad (74)$$

By measurability of $A_0, A_1, B, \Delta A$ and $\tau_q(x) = \frac{1}{2}q^\top \Delta A(x)$ are measurable. Since $B(x)$ has diagonal form, with nonnegative entries $B_{11}(x)$ and $B_{22}(x)$, the graph

$$\text{Gr}(\mathcal{U}) = \{(x, u) \in \mathcal{X} \times \mathbb{R}^3 : u_3 = 0, 0 \leq u_1 \leq B_{11}(x), 0 \leq u_2 \leq B_{22}(x)\}$$

is measurable. Hence, for fixed q ,

$$\text{Gr}(K_q^0) = \{(x, y) : (x, y - A_0(x)) \in \text{Gr}(\mathcal{U}), q^\top(y - A_0(x)) \leq \tau_q(x)\}$$

is measurable, and, similarly $\text{Gr}(K_q^1)$ is measurable. Therefore $K_q = K_q^0 \cup K_q^1$ has a measurable graph. For each x , $\mathcal{U}(x) \equiv B(x)[0, 1]^2$ is compact. Intersecting $\mathcal{U}(x)$ with a closed

half-space preserves compactness, and translating or reflecting preserves compactness. Hence $K_q^0(x)$ and $K_q^1(x)$ are compact, possibly empty sets, and $K_q(x) \equiv K_q^0(x) \cup K_q^1(x)$ is compact. Moreover, $K_q(x)$ is nonempty because $\mathcal{U}(x) \neq \emptyset$, and every $u \in \mathcal{U}(x)$ satisfies either $q^\top u \leq \tau_q(x)$ or $q^\top u \geq \tau_q(x)$. By [Molchanov \(2017, Thm. 1.3.3\)](#), K_q is a non-empty random compact set. For each fixed $v \in \mathbb{S}^2$, the support function of K_q in direction v , $\sup_{y \in K_q(x)} v^\top y$ is measurable ([Molchanov, 2017, Prop. 1.3.8](#)) and admits a measurable maximizer. The binary-loss bounds imply that K_q is integrably bounded. Consequently, the following Aumann integral is well defined:

$$\{\mathbb{E}[k(X)] : k(x) \in K_q(x) \text{ } \mathbb{P}_X\text{-a.s.}\}.$$

We next show that for every $q \in \mathbb{S}^2$, we can represent \mathcal{M}_q as an Aumann expectation ([Molchanov and Molinari, 2018, Def. 3.1](#)):

$$\mathcal{M}_q = \{\mathbb{E}[k(X)] : k(x) \in K_q(x) \text{ } \mathbb{P}_X\text{-a.s.}\}. \quad (75)$$

To see this, take an element of \mathcal{M}_q . Then for some $\lambda \in \Lambda$ and some q -optimal rule a , let

$$k(x) = a(x)\boldsymbol{\theta}_1(x; \lambda) + (1 - a(x))\boldsymbol{\theta}_0(x; \lambda).$$

Let $u(x) = B(x)\lambda(x)$. If $a(x) = 0$, the q -optimality condition gives

$$q^\top u(x) \leq \frac{1}{2}q^\top \Delta A(x) = \tau_q(x).$$

Thus $k(x) = A_0(x) + u(x) \in K_q^0(x)$. If $a(x) = 1$, the same argument gives $q^\top u(x) \geq \tau_q(x)$, so $k(x) = A_1(x) - u(x) \in K_q^1(x)$. Hence $k(x) \in K_q(x)$ a.s., and the first inclusion follows.

Conversely, let k be a measurable selector of K_q . Since K_q^0 and K_q^1 have measurable graphs and compact values, a measurable branch-selection argument gives measurable $a : \mathcal{X} \rightarrow \{0, 1\}$ and measurable $u(x) \in \mathcal{U}(x)$ such that

$$k(x) = (1 - a(x))(A_0(x) + u(x)) + a(x)(A_1(x) - u(x)),$$

with $q^\top u(x) \leq \tau_q(x)$ on $\{a(x) = 0\}$, and $q^\top u(x) \geq \tau_q(x)$ on $\{a(x) = 1\}$. By [\(74\)](#), there exists $\lambda \in \Lambda$ with $B(x)\lambda(x) = u(x)$ a.s., and $a(x) \in \arg \min_{d \in \{0, 1\}} q^\top \boldsymbol{\theta}_d(x; \lambda)$, so that

$\mathbb{E}[k(X)] \in \mathcal{M}_q$. It follows that \mathcal{M}_q is the Aumann integral of K_q . The correspondence K_q is nonempty, measurable, compact-valued, and integrably bounded because $\mathcal{U}(x)$ is compact and because A_0, A_1, B are bounded in the binary-loss setup. Hence, for each $q \in \mathbb{S}^2$, \mathcal{M}_q is convex and compact by [Molchanov \(2017, Thms. 2.1.26 and 2.1.38\)](#), where to apply Thm. 2.1.26 we used that Assumption 4 implies that \mathbb{P}_X is nonatomic. *Q.E.D.*

LEMMA B.6: *Let Assumptions 1 and 4 hold. Then, for every fixed $\varepsilon \in \mathbb{R}^3$ and $\rho_q(\varepsilon) \equiv \sup_{v \in \mathbb{S}^2} \{v^\top \varepsilon - h_{\mathcal{M}_q}(v)\}$, the mapping $q \mapsto \rho_q(\varepsilon)$ is continuous on $\tilde{\mathbb{S}}^2$. Moreover, the infimum defining $T_{\text{env}}(\varepsilon)$ in (29) is attained.*

PROOF: By Lemma B.7, for all $p, q \in \tilde{\mathbb{S}}^2$,

$$\sup_{v \in \mathbb{S}^2} |h_{\mathcal{M}_p}(v) - h_{\mathcal{M}_q}(v)| \leq 2H_0CM_0^m \|p - q\|^m.$$

Therefore, $|\rho_p(\varepsilon) - \rho_q(\varepsilon)| \leq \sup_{v \in \mathbb{S}^2} |h_{\mathcal{M}_p}(v) - h_{\mathcal{M}_q}(v)| \leq 2H_0CM_0^m \|p - q\|^m$. Hence $q \mapsto \rho_q(\varepsilon)$ is continuous.

Now define $Q(\varepsilon) = \{q \in \tilde{\mathbb{S}}^2 : q_3 \varepsilon_3 \geq 0\}$. This set is nonempty because $(1, 0, 0) \in Q(\varepsilon)$, and it is compact because it is a closed subset of \mathbb{S}^2 . Since $q \mapsto [\rho_q(\varepsilon)]_+$ is continuous on $Q(\varepsilon)$, it attains its minimum. Thus, $T_{\text{env}}(\varepsilon) = \min_{q \in Q(\varepsilon)} [\rho_q(\varepsilon)]_+$. *Q.E.D.*

LEMMA B.7: *Let Assumptions 1 and 4 hold. Then, for every $v \in \mathbb{S}^2$,*

$$h_{\mathcal{M}_q}(v) = \mathbb{E}[\max\{\kappa_0(v, q; X), \kappa_1(v, q; X)\}]. \quad (76)$$

Moreover, $q_n \rightarrow q$ implies $h_{\mathcal{M}_{q_n}}(v) \rightarrow h_{\mathcal{M}_q}(v)$, and for all $p, q \in \mathbb{S}^2$,

$$\sup_{v \in \mathbb{S}^2} |h_{\mathcal{M}_p}(v) - h_{\mathcal{M}_q}(v)| \leq 2H_0CM_0^m \|p - q\|^m. \quad (77)$$

PROOF: By [Molchanov \(2017, Thm. 2.1.35\)](#), the support function of \mathcal{M}_q is given by

$$h_{\mathcal{M}_q}(v) = \mathbb{E} \left[\sup_{y \in K_q(X)} v^\top y \right].$$

Using the definition of $K_q(X)$, $\kappa_0(v, q; X)$, $\kappa_1(v, q; X)$, we obtain (76). Next, fix $p, q \in \mathbb{S}^2$. For $u \in \mathcal{U}(x)$, write $z_x(u) = \Delta A(x) - 2u$. As $p^\top u \leq \frac{1}{2}p^\top \Delta A(x)$ if and only if $p^\top (\Delta A(x) -$

$2u) \geq 0$, the feasibility inequalities can be rewritten as $p^\top u \leq \tau_p(x) \iff p^\top z_x(u) \geq 0$ and $p^\top u \geq \tau_p(x) \iff p^\top z_x(u) \leq 0$. Suppose $d_q(x) > M_0 \|p - q\|$. Then for every $u \in \mathcal{U}(x)$,

$$|(p - q)^\top z_x(u)| \leq \|p - q\| \|z_x(u)\| \leq M_0 \|p - q\| < d_q(x) \leq |q^\top z_x(u)|.$$

Therefore $p^\top z_x(u)$ and $q^\top z_x(u)$ have the same sign for every $u \in \mathcal{U}(x)$. Hence

$$\begin{aligned} \{u \in \mathcal{U}(x) : p^\top u \leq \tau_p(x)\} &= \{u \in \mathcal{U}(x) : q^\top u \leq \tau_q(x)\}, \\ \{u \in \mathcal{U}(x) : p^\top u \geq \tau_p(x)\} &= \{u \in \mathcal{U}(x) : q^\top u \geq \tau_q(x)\}. \end{aligned}$$

The objectives in the definitions of κ_0 and κ_1 do not depend on the normal vector. Therefore, on the event $\{d_q(x) > M_0 \|p - q\|\}$, we have

$$\max\{\kappa_0(v, p; X), \kappa_1(v, p; X)\} = \max\{\kappa_0(v, q; X), \kappa_1(v, q; X)\}.$$

On the complementary event, the bound defining H_0 gives

$$|\max\{\kappa_0(v, p; X), \kappa_1(v, p; X)\}| \leq H_0, \quad |\max\{\kappa_0(v, q; X), \kappa_1(v, q; X)\}| \leq H_0.$$

Thus

$$|\max\{\kappa_0(v, p; X), \kappa_1(v, p; X)\} - \max\{\kappa_0(v, q; X), \kappa_1(v, q; X)\}| \leq 2H_0 \mathbf{1}\{d_q(x) \leq M_0 \|p - q\|\}.$$

Taking expectations, $|h_{\mathcal{M}_p}(v) - h_{\mathcal{M}_q}(v)| \leq 2H_0 \mathbb{P}_X\{d_q(X) \leq M_0 \|p - q\|\}$. By Assumption 4, $\mathbb{P}_X\{d_q(X) \leq M_0 \|p - q\|\} \lesssim M_0^m \|p - q\|^m$. Hence

$$|h_{\mathcal{M}_p}(v) - h_{\mathcal{M}_q}(v)| \lesssim 2H_0 M_0^m \|p - q\|^m,$$

establishing (77). The right-hand side does not depend on v , so taking $p = q_n$ establishes continuity of $h_{\mathcal{M}_q}$ in q , uniformly v . *Q.E.D.*

B.3. Lemmas and Auxiliary Results for Sections 4-5

Recall that \mathfrak{F}_k denotes the σ -algebra generated by all the data except the ones with indices in \mathcal{I}_k .

LEMMA B.8: Let Assumptions 6 and 7 hold. Then, there exist $\widehat{\delta}_k : \mathcal{X}_2 \rightarrow \mathbb{R}^+$ and an event $\widehat{E}_k \in \mathfrak{F}_k$ such that (i) $\widehat{\delta}_k$ conditional on \mathfrak{F}_k is nonstochastic, (ii) $\mathbb{P}(\widehat{E}_k^c) = o(1)$, (iii) conditional on \widehat{E}_k we have that $\widehat{\boldsymbol{\mu}}_k \in \mathbb{B}_\epsilon(\boldsymbol{\mu}_r, \boldsymbol{\mu}_b)$, $\|\Delta\widehat{\boldsymbol{\theta}}_k(X_i) - \Delta\boldsymbol{\theta}(X_i)\| \leq \widehat{\delta}_k(X_{2,i})$ and $\|\widehat{\pi}_k(X_i) - \pi(X_i)\| \leq \widehat{\delta}_k(X_{2,i})$, and (iv) $\mathbb{E}[\widehat{\delta}_k(X_{2,i})^2 | \mathfrak{F}_k] = o_p(n^{-1/2})$.

PROOF: Let $C_F = \sup_{\nu \in \mathbb{R}^{d_\nu}} \sup_{\gamma \in \mathbb{R}^{d_\gamma}} \sup_{\mu \in \mathbb{B}_\epsilon(\boldsymbol{\mu}_r, \boldsymbol{\mu}_b)} \|DF(\nu, \gamma, \mu)\|$. Consider the event $\widehat{E}_k = \{\|\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}\| < \epsilon\}$, which holds with probability approaching one by definition of $\widehat{\boldsymbol{\mu}}_k$. By Assumptions 6 and 7, the mean value inequality, and conditional on \widehat{E}_k , we have

$$\left\| \left(\widehat{\pi}(X_i) - \pi(X_i), \Delta\widehat{\boldsymbol{\theta}}(X_i) - \Delta\boldsymbol{\theta}(X_i) \right) \right\|$$

is lower than or equal to $C_F (\|\widehat{\nu}_k(X_{1,i}) - \nu(X_{1,i})\| + \|\widehat{\gamma}_k(X_{2,i}) - \gamma(X_{2,i})\| + \|\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}\|)$. By Assumption 7, the previous expression is lower than or equal to

$$\widehat{\delta}_k(X_{2,i}) \equiv C_F \left(\sup_{x_1 \in \mathcal{X}_1} \|\widehat{\nu}_k(x_1) - \nu(x_1)\| + \|\widehat{\gamma}_k(X_{2,i}) - \gamma(X_{2,i})\| + \|\widehat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}\| \right).$$

By Assumption 7 and Jensen's inequality, we have $\mathbb{E}[\widehat{\delta}_k(X_{2,i})^2 | \mathfrak{F}_k] = o_p(n^{-1/2})$. *Q.E.D.*

LEMMA B.9: Let Assumption 6 and 7 hold. Then, $n_k^{-1/2} \sum_{i \in \mathcal{I}_k} |\widehat{\alpha}_k^h(q; X_i) - \alpha^h(q; X_i)|^2 = o_p(1)$ uniformly on $q \in \mathbb{S}^2$.

PROOF: We can write $n_k^{-1/2} \sum_{i \in \mathcal{I}_k} |\widehat{\alpha}_k^h(q; X_i) - \alpha^h(q; X_i)|^2 \leq \widetilde{A}_k + B_k + C_k$, where

$$\widetilde{A}_k = n_k^{-1/2} \sum_{i \in \mathcal{I}_k} |q^\top \Delta\boldsymbol{\theta}(X_i)| \mathbb{1}\{|q^\top \Delta\boldsymbol{\theta}(X_i)| \leq |q^\top \Delta\widehat{\boldsymbol{\theta}}_k(X_i) - q^\top \Delta\boldsymbol{\theta}(X_i)|\}$$

$$B_k = n_k^{-1/2} \sum_{i \in \mathcal{I}_k} |q^\top \widehat{\boldsymbol{\theta}}_{0,k}(X_i) - q^\top \boldsymbol{\theta}_0(X_i)|^2, C_k = n_k^{-1/2} \sum_{i \in \mathcal{I}_k} |q^\top \Delta\widehat{\boldsymbol{\theta}}_k(X_i) - q^\top \Delta\boldsymbol{\theta}(X_i)|^2.$$

In the proof of Lemma B.11, we conclude $\widetilde{A}_k = o_p(1)$ uniformly on $q \in \mathbb{S}^2$, with \widetilde{A}_k defined in (82). Finally, note that B_k and C_k both are $o_p(1)$ uniformly on $q \in \mathbb{S}^2$ by Cauchy-Schwartz and Assumption 7. *Q.E.D.*

B.3.1. Auxiliary Results for Weighted Estimators for the Support Function

Let $\{\omega_i : 1 \leq i \leq n\}$ be a random sample of positive weights that are independent of the data such that $\mathbb{E}[\omega_i^2] \leq C_\omega$. Define

$$\widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta}) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}), \quad (78)$$

where $\bar{\omega} = n^{-1} \sum_{i=1}^n \omega_i$ and $\zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta})$ is defined in (40).

LEMMA B.10: *Let Assumptions 1–7 hold. Then,*

$$\sup_{q \in \mathbb{S}^2} n^{1/2} \left| \widehat{h}_{\mathcal{E}}^{\omega}(q; \widehat{\mathbf{L}}^{\omega}, \widehat{\boldsymbol{\eta}}) - \widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta}) \right| = o_p(1), \quad (79)$$

where $\widehat{h}_{\mathcal{E}}^{\omega}(q; \widehat{\mathbf{L}}^{\omega}, \widehat{\boldsymbol{\eta}})$ and $\widehat{h}_{\mathcal{E}}^{\omega,*}(q; \mathbf{L}, \boldsymbol{\eta})$ are defined in (55) and (78), respectively.

PROOF: Let $\widehat{\alpha}_k^h$ be the estimator of α^h defined in (35) using $\widehat{\boldsymbol{\eta}}_k$. Recall $n_k \equiv n/K$ and define $I_{j,n}(q) \equiv \frac{1}{K} \sum_{k=1}^K \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) \mathfrak{Z}_{k,i}^{(j)}(q)$ for $j \in \{0, 1, 2, 3\}$, where

$$\begin{aligned} \mathfrak{Z}_{k,i}^{(0)}(q) &\equiv \frac{q^\top \widehat{\mathbf{L}}_{0,i}^{\omega} Z_i}{\widehat{\pi}_k(X_i)} + \left(\frac{q^\top \Delta \widehat{\mathbf{L}}_i^{\omega} Z_i}{\widehat{\pi}_k(X_i)} \right) \mathbb{1}\{q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) > 0\} + \widehat{\alpha}_k^h(q; X_i) \left(1 - \frac{Z_i}{\widehat{\pi}_k(X_i)}\right) \\ \mathfrak{Z}_{k,i}^{(1)}(q) &\equiv \frac{q^\top \widehat{\mathbf{L}}_{0,i}^{\omega} Z_i}{\widehat{\pi}_k(X_i)} + \left(\frac{q^\top \Delta \widehat{\mathbf{L}}_i^{\omega} Z_i}{\widehat{\pi}_k(X_i)} \right) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} + \widehat{\alpha}_k^h(q; X_i) \left(1 - \frac{Z_i}{\widehat{\pi}_k(X_i)}\right) \\ \mathfrak{Z}_{k,i}^{(2)}(q) &\equiv \frac{q^\top \mathbf{L}_{0,i} Z_i}{\widehat{\pi}_k(X_i)} + \left(\frac{q^\top \Delta \mathbf{L}_i Z_i}{\widehat{\pi}_k(X_i)} \right) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} + \widehat{\alpha}_k^h(q; X_i) \left(1 - \frac{Z_i}{\widehat{\pi}_k(X_i)}\right) \\ &\quad + \sum_{g \in \{r, b\}} \Gamma_g^h(q) \left(1 - \frac{\mathbb{1}\{G_i = g\}}{\mu_g}\right) \\ \mathfrak{Z}_{k,i}^{(3)}(q) &\equiv \zeta_i^*(q; \mathbf{L}, \boldsymbol{\eta}). \end{aligned}$$

Our goal is to establish (79). We divide the proof in three claims.

Claim 1: $\sup_{q \in \mathbb{S}^2} \sqrt{n} |I_{0,n}(q) - I_{1,n}(q)| = o_p(1)$.

We write $I_{0,n}(q) - I_{1,n}(q) = \frac{1}{K} \sum_{k=1}^K (A_k + B_k)$, where

$$A_k = \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) \left(\frac{q^\top \Delta \mathbf{L}_i Z_i}{\pi(X_i)} \right) \left(\mathbb{1}\{q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) > 0\} - \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} \right) \quad (80)$$

$$B_k = \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) \left(\frac{q^\top \Delta \widehat{\mathbf{L}}_i^{\omega} Z_i}{\widehat{\pi}_k(X_i)} - \frac{q^\top \Delta \mathbf{L}_i Z_i}{\pi(X_i)} \right) \left(\mathbb{1}\{q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) > 0\} - \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} \right) \quad (81)$$

Lemma B.11 shows that $A_k = o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$ by relying on Assumptions 6 and 7 and the approach used in the proof of Theorem 4.1 in Liu and Molinari (2026). Using similar ideas, Lemma B.12 shows that $B_k = o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$. Since K is fixed as $n \rightarrow \infty$, we conclude the claim of step 1.

Claim 2: $\sup_{q \in \mathbb{S}^2} \sqrt{n} |I_{1,n}(q) - I_{2,n}(q)| = o_p(1)$.

Recall $\widehat{\mu}_g^\omega \equiv n^{-1} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \mathbb{1}\{G_i = g\}$. Note that $n^{-1} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \left(1 - \frac{\mathbb{1}\{G_i=g\}}{\widehat{\mu}_g^\omega}\right) = 0$, which implies

$$I_{1,n}(q) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \left[\zeta_i(q; \widehat{\mathbf{L}}^\omega, (\Delta\boldsymbol{\theta}, \widehat{\pi}, \widehat{\alpha}^h)) + \sum_{g \in \{r,b\}} \Gamma_g^h(q) \left(1 - \frac{\mathbb{1}\{G_i=g\}}{\widehat{\mu}_g^\omega}\right) \right].$$

Since

$$I_{2,n}(q) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \left[\zeta_i(q; \mathbf{L}, (\Delta\boldsymbol{\theta}, \widehat{\pi}, \widehat{\alpha}^h)) + \sum_{g \in \{r,b\}} \Gamma_g^h(q) \left(1 - \frac{\mathbb{1}\{G_i=g\}}{\mu_g}\right) \right],$$

we can write

$$I_{1,n}(q) - I_{2,n}(q) = \sum_{g \in \{r,b\}} \left(\frac{1}{\widehat{\mu}_g^\omega} - \frac{1}{\mu_g}\right) \left(\frac{1}{K} \sum_{k=1}^K A_k^g + B_k^g\right),$$

where

$$A_k^g = \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) \left\{ q^\top (\mathbf{L}_{0,i} + \Delta \mathbf{L}_i \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\}) \frac{Z_i \mu_g}{\pi(X_i)} - \Gamma_g^h(q) \right\} \mathbb{1}\{G_i = g\}$$

$$B_k^g = \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) \left(\frac{Z_i \mu_g}{\widehat{\pi}_k(X_i)} - \frac{Z_i \mu_g}{\pi(X_i)} \right) q^\top (\mathbf{L}_{0,i} + \Delta \mathbf{L}_i \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\}) \mathbb{1}\{G_i = g\}.$$

We conclude that $A_k^g = O_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$ by using that the weights ω_i are independent of the data, $\bar{\omega} = \mathbb{E}[\omega_i] + O_p(n^{-1/2})$, Assumptions 1 and 5, the definition of $\Gamma_g(q)$, and the Central Limit Theorem. We also conclude that $B_k^g = o_p(1)$ uniformly on $q \in \mathbb{S}^2$ by Cauchy-Schwartz, and Assumptions 1, 6, and 7. Finally, we conclude $\frac{1}{\mu_g} - \frac{1}{\widehat{\mu}_g^\omega} = O_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$ by the Central Limit Theorem, Assumption 1, and the definition of the weights ω_i .

Claim 3: $\sup_{q \in \mathbb{S}^2} \sqrt{n} |I_{2,n}(q) - I_{3,n}(q)| = o_p(1)$.

We write $\sqrt{n}(I_{2,n}(q) - I_{3,n}(q)) = \frac{1}{\sqrt{K}} \sum_{k=1}^K (A_k + B_k + C_k)$, where

$$A_k = \frac{1}{\sqrt{n_k}} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}} \right) \left(\frac{Z_i}{\hat{\pi}_k(X_i)} - \frac{Z_i}{\pi(X_i)} \right) \left(q^\top \mathbf{L}_{0,i} + (q^\top \Delta \mathbf{L}_i) \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} - \alpha^h(q; X_i) \right),$$

$$B_k = \frac{1}{\sqrt{n_k}} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}} \right) \left(\hat{\alpha}_k^h(q; X) - \alpha^h(q; X) \right) \left(1 - \frac{Z_i}{\pi(X_i)} \right),$$

$$C_k = \frac{1}{\sqrt{n_k}} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}} \right) \left(\hat{\alpha}_k^h(q; X) - \alpha^h(q; X) \right) \left(\frac{Z_i}{\pi(X_i)} - \frac{Z_i}{\hat{\pi}_k(X_i)} \right).$$

Recall that $(\Delta \hat{\boldsymbol{\theta}}_k, \hat{\pi}_k, \hat{\boldsymbol{\theta}}_{0,k})$ was estimated using all the data except the ones with indices in \mathcal{I}_k , and \mathfrak{F}_k is the σ -algebra generated by all the data except the ones with indices on \mathcal{I}_k . The definition of the weights ω_i , Assumptions 6 and 7, the definition of $\alpha^h(q; X)$ in (35), and algebra show $\mathbb{E}[A_k | \mathfrak{F}_k] = 0$ and $\sup_{q \in \mathbb{S}^2} \mathbb{E}[A_k^2 | \mathfrak{F}_k] = o(1)$. Similarly, we can show $\mathbb{E}[B_k | \mathfrak{F}_k] = 0$ and $\sup_{q \in \mathbb{S}^2} \mathbb{E}[B_k^2 | \mathfrak{F}_k] = o(1)$. By Chernozhukov et al. (2018, Lemma 6.1), we conclude A_k and B_k both are $o_p(1)$ uniformly on $q \in \mathbb{S}^2$. Finally, we conclude that $C_k = o_p(1)$ uniformly on $q \in \mathbb{S}^2$ by Cauchy-Schwartz, Assumptions 6 and 7, Lemma B.9, and the definition of the weights ω_i .

The triangle inequality and Claims 1, 2, and 3, complete the proof of the lemma. *Q.E.D.*

LEMMA B.11: Let Assumptions 1–7 hold. Then, $A_k = o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$, with A_k defined in (80).

PROOF: It is sufficient to prove that $\mathbb{E}[A_k | \mathfrak{F}_k, (X_i)_{i \in \mathcal{I}_k}] = o_p(n^{-1/2})$ to conclude that A_k is $o_p(n^{-1/2})$, by Chernozhukov et al. (2018, Lemma 6.1). Define

$$\tilde{A}_k \equiv \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} |q^\top \Delta \boldsymbol{\theta}(X_i)| \mathbb{1}\{|q^\top \Delta \boldsymbol{\theta}(X_i)| \leq |q^\top \Delta \hat{\boldsymbol{\theta}}_k(X_i) - q^\top \Delta \boldsymbol{\theta}(X_i)|\}. \quad (82)$$

Note that Law of Iterated Expectations, Assumption 1, and the definitions of the weights ω_i imply $|\mathbb{E}[A_k | \mathfrak{F}_k, (X_i)_{i \in \mathcal{I}_k}]| \leq \tilde{A}_k$; therefore, it is sufficient to show that \tilde{A}_k is $o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$. By Cauchy-Schwartz and Lemma B.8, we have $|q^\top \Delta \hat{\boldsymbol{\theta}}(X_i) - q^\top \Delta \boldsymbol{\theta}(X_i)| \leq \hat{\delta}_k(X_{2,i})$ conditional on \hat{E}_k , where $\hat{\delta}_k$ and \hat{E}_k are as in Lemma B.8. Since $\mathbb{P}(\hat{E}_k^c) = o(1)$, it is sufficient to show that $\mathbb{E}[\tilde{A}_k \mathbb{1}\{\hat{E}_k\} | \mathfrak{F}_k] = o_p(n^{-1/2})$ to conclude \tilde{A}_k is

$o_p(n^{-1/2})$. We verify this next:

$$\begin{aligned}
\mathbb{E}[\tilde{A}_k \mathbf{1}\{\widehat{E}_k\} | \mathfrak{F}_k] &= \mathbb{E} \left[|q^\top \Delta \boldsymbol{\theta}(X_i)| \mathbf{1}\{|q^\top \Delta \boldsymbol{\theta}(X_i)| \leq |q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) - q^\top \Delta \boldsymbol{\theta}(X_i)|\} \mathbf{1}\{\widehat{E}_k\} | \mathfrak{F}_k \right] \\
&\stackrel{(1)}{\leq} \mathbb{E} \left[|q^\top \Delta \boldsymbol{\theta}(X_i)| \mathbf{1}\{|q^\top \Delta \boldsymbol{\theta}(X_i)| \leq \widehat{\delta}_k(X_{2,i})\} | \mathfrak{F}_k \right] \mathbf{1}\{\widehat{E}_k\} \\
&\stackrel{(2)}{\leq} \mathbb{E} \left[\widehat{\delta}_k(X_{2,i}) \mathbb{E} \left[\mathbf{1}\{|q^\top \Delta \boldsymbol{\theta}(X_i)| \leq \widehat{\delta}_k(X_{2,i})\} | \mathfrak{F}_k, X_{2,i} \right] | \mathfrak{F}_k \right] \mathbf{1}\{\widehat{E}_k\} \\
&\stackrel{(3)}{\lesssim} \mathbb{E} \left[\widehat{\delta}_k(X_{2,i})^2 | \mathfrak{F}_k \right] \stackrel{(4)}{=} o_p(n^{-1/2})
\end{aligned}$$

where (1) holds by Lemma B.8 and since $\mathbf{1}\{\widehat{E}_k\} \in \mathfrak{F}_k$, (2) holds since $\widehat{\delta}_k(X_{2,i})$ is non-stochastic conditional on \mathfrak{F}_k and $X_{2,i}$, (3) holds by Assumption 6 and since $\mathbf{1}\{\widehat{E}_k\} \leq 1$, and (4) holds by Lemma B.8. Finally, note that $\mathbb{P}(\widehat{E}_k^c) = o(1)$ and (4) hold uniformly on $q \in \mathbb{S}^2$ since these objects do not depend on q . *Q.E.D.*

LEMMA B.12: Let Assumptions 1–7 hold. Then, $B_k = o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$, with B_k defined in (81).

PROOF: We can write $B_k = B_{k,1} + B_{k,2}$, where

$$\begin{aligned}
B_{k,1} &= \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}} \right) q^\top \Delta \mathbf{L}_i Z_i \left(\frac{1}{\widehat{\pi}_k(X_i)} - \frac{1}{\pi(X_i)} \right) \left(\mathbf{1}\{q^\top \Delta \widehat{\boldsymbol{\theta}}(X_i) > 0\} - \mathbf{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} \right) \\
B_{k,2} &= \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}} \right) \left(\frac{q^\top \Delta \widehat{\mathbf{L}}_i^\omega Z_i - q^\top \Delta \mathbf{L}_i Z_i}{\widehat{\pi}_k(X_i)} \right) \left(\mathbf{1}\{q^\top \Delta \widehat{\boldsymbol{\theta}}(X_i) > 0\} - \mathbf{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} \right)
\end{aligned}$$

As in the proof of Lemma B.11, it is sufficient to show that $\mathbb{E}[B_{k,1} | \mathfrak{F}_k, (X_i)_{i \in \mathcal{I}_k}]$ and $B_{k,2}$ are both $o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$. Define

$$\widetilde{B}_{k,1} = \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} |\widehat{\pi}_k(X_i) - \pi(X_i)| \mathbf{1}\{|q^\top \Delta \boldsymbol{\theta}(X_i)| \leq |q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) - q^\top \Delta \boldsymbol{\theta}(X_i)|\}$$

Note that Law of Iterated Expectations, Assumptions 1 and 6, and the definitions of the weights ω_i imply that $|\mathbb{E}[B_{k,1} | \mathfrak{F}_k, (X_i)_{i \in \mathcal{I}_k}]| \leq (\frac{1}{\epsilon}) \widetilde{B}_{k,1}$; therefore, it is sufficient to show that $\widetilde{B}_{k,1}$ is $o_p(n^{-1/2})$ uniformly on $q \in \mathbb{S}^2$ to conclude the same for $\mathbb{E}[B_{k,1} | \mathfrak{F}_k, (X_i)_{i \in \mathcal{I}_k}]$.

By similar arguments used in the proof of Lemma B.11 and by relying on Lemma B.8,

$$\mathbb{E}[\tilde{B}_{k,1} \mathbb{1}\{\widehat{E}_k\} | \mathfrak{F}_k] \lesssim \mathbb{E} \left[\widehat{\delta}_k(X_{2,i})^2 | \mathfrak{F}_k \right] = o_p(n^{-1/2}),$$

where \widehat{E}_k and $\widehat{\delta}_k$ are as in Lemma B.8. By the next identity

$$q^\top \Delta \widehat{\mathbf{L}}_i^\omega - q^\top \Delta \mathbf{L}_i = \sum_{g \in \{r,b\}} \mu_g \left(\frac{1}{\widehat{\mu}_g^\omega} - \frac{1}{\mu_g} \right) (q^\top \Delta \mathbf{L}_i) \mathbb{1}\{G_i = g\} \implies B_{k,2} = \sum_{g \in \{r,b\}} \mu_g \left(\frac{1}{\widehat{\mu}_g^\omega} - \frac{1}{\mu_g} \right) B_{k,2}^g,$$

where

$$B_{k,2}^g = \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}} \right) \left(\frac{\Delta \mathbf{L}_i \mathbb{1}\{G_i=g\} Z_i}{\widehat{\pi}_k(X_i)} \right) \left(\mathbb{1}\{q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) > 0\} - \mathbb{1}\{q^\top \Delta \boldsymbol{\theta}(X_i) > 0\} \right), \quad g \in \{r,b\}$$

We conclude that $\mu_g \left(\frac{1}{\widehat{\mu}_g^\omega} - \frac{1}{\mu_g} \right) = O_p(n^{-1/2})$ uniformly on $g \in \mathbb{S}^2$ by the Central Limit Theorem, Assumption 1, and the definition of the weights ω_i . Therefore, it is sufficient to show that $B_{k,2}^g = o_p(1)$ uniformly on $g \in \mathbb{S}^2$ to conclude that $B_{k,2} = o_p(n^{-1/2})$ uniformly on $g \in \mathbb{S}^2$. Define $\tilde{B}_{k,2}^g \equiv \frac{1}{n_k} \sum_{i \in \mathcal{I}_k} \mathbb{1}\{|q^\top \Delta \boldsymbol{\theta}(X_i)| \leq |q^\top \Delta \widehat{\boldsymbol{\theta}}_k(X_i) - q^\top \Delta \boldsymbol{\theta}(X_i)|\}$. Note that Law of Iterated Expectations, Assumptions 1 and 6, and the definitions of the weights imply $|\mathbb{E}[B_{k,2}^g | \mathfrak{F}_k]| \leq C \tilde{B}_{k,2}^g$, where C is a constant based only on c_1 , c_2 , and ϵ (constants from Assumptions 1 and 6). By similar arguments used in the proof of Lemma B.11 and by relying on Lemma B.8, we conclude

$$\mathbb{E}[\tilde{B}_{k,2}^g \mathbb{1}\{\widehat{E}_k\} | \mathfrak{F}_k] \lesssim \mathbb{E} \left[\widehat{\delta}_k(X_{2,i}) | \mathfrak{F}_k \right] = o_p(n^{-1/4}),$$

which is sufficient to complete the proof of the lemma. Q.E.D.

B.3.2. Auxiliary Results for Weighted Estimator of ε^*

Let $\{\omega_i : 1 \leq i \leq n\}$ be a random sample of positive weights that are independent of the data such that $\mathbb{E}[\omega_i^2] \leq C_\omega$. Define

$$\widehat{\varepsilon}^\omega(a^*; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}} \right) \xi_i(a^*; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) \quad (83)$$

$$\widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \boldsymbol{\eta}) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}} \right) \xi_i^*(a^*; \mathbf{L}, \boldsymbol{\eta}), \quad (84)$$

where $\bar{\omega} = n^{-1} \sum_{i=1}^n \omega_i$, $\xi_i(a^*; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}})$, $\widehat{\mathbf{L}}^\omega$ is defined as in (1)-(2) with μ_g replaced by $\widehat{\mu}_g^\omega = n^{-1} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \mathbb{1}\{G_i = g\}$, and

$$\xi_i^*(a^*; \mathbf{L}, \boldsymbol{\eta}) = \xi_i(a^*; \mathbf{L}, \boldsymbol{\eta}) + \sum_{g \in \{r, b\}} \Gamma_g^e \left(1 - \frac{\mathbb{1}\{G_i = g\}}{\mu_g}\right)$$

and $\Gamma_g^e \equiv \mathbb{E}[\mathbb{1}\{G_i = g\} (\mathbf{L}_{0,i} + a(X_i) \Delta \mathbf{L}_i) \frac{Z_i}{\pi(X_i)}]$.

LEMMA B.13: *Let Assumptions 1–7 hold. Then,*

$$\sup_{q \in \mathbb{S}^2} n^{1/2} \left| \widehat{\varepsilon}^\omega(a^*; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) - \widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \boldsymbol{\eta}) \right| = o_p(1) \quad (85)$$

PROOF: Let $\widehat{\boldsymbol{\alpha}}_k^e$ be the estimator of $\boldsymbol{\alpha}^e$ defined in (38) using $\widehat{\boldsymbol{\eta}}_k$. Define

$$\widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \widehat{\boldsymbol{\eta}}) \equiv \frac{1}{n} \sum_{i=1}^n \left(\frac{\omega_i}{\bar{\omega}}\right) \xi_i^*(a^*; \mathbf{L}, \widehat{\boldsymbol{\eta}}).$$

Our goal is to establish (85). We divide the proofs in two claims.

Claim 1: $\sup_{q \in \mathbb{S}^2} \sqrt{n} |\widehat{\varepsilon}^\omega(a^*; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) - \widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \widehat{\boldsymbol{\eta}})| = o_p(1)$. This claim is similar to Claim 2 in the proof of Lemma B.10 and also its proof; therefore, it is omitted.

Claim 2: $\sup_{q \in \mathbb{S}^2} \sqrt{n} |\widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \widehat{\boldsymbol{\eta}}) - \widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \boldsymbol{\eta})| = o_p(1)$.

We write $\sqrt{n} (\widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \widehat{\boldsymbol{\eta}}) - \widehat{\varepsilon}^{\omega,*}(a^*; \mathbf{L}, \boldsymbol{\eta})) = \frac{1}{\sqrt{K}} \sum_{k=1}^K (A_k + B_k + C_k)$, where

$$A_k = \frac{1}{\sqrt{n_k}} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) \left(\frac{Z_i}{\widehat{\pi}_k(X_i)} - \frac{Z_i}{\pi(X_i)}\right) (\mathbf{L}_{0,i} + a^*(X_i) \Delta \mathbf{L}_i - \boldsymbol{\alpha}^e(a^*; X_i)),$$

$$B_k = \frac{1}{\sqrt{n_k}} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) (\widehat{\boldsymbol{\alpha}}_k^e(a^*; X) - \boldsymbol{\alpha}^e(a^*; X)) \left(1 - \frac{Z_i}{\pi(X_i)}\right),$$

$$C_k = \frac{1}{\sqrt{n_k}} \sum_{i \in \mathcal{I}_k} \left(\frac{\omega_i}{\bar{\omega}}\right) (\widehat{\boldsymbol{\alpha}}_k^e(a^*; X) - \boldsymbol{\alpha}^e(a^*; X)) \left(\frac{Z_i}{\pi(X_i)} - \frac{Z_i}{\widehat{\pi}_k(X_i)}\right).$$

Recall that $\widehat{\boldsymbol{\eta}}_k \equiv (\Delta \widehat{\boldsymbol{\theta}}_k, \widehat{\pi}_k, \widehat{\boldsymbol{\theta}}_{0,k})$ was estimated using all the data except the ones with indices in \mathcal{I}_k , and \mathfrak{F}_k is the σ -algebra generated by all the data except the ones with indices on \mathcal{I}_k . The definition of the weights ω_i , Assumptions 6 and 7, the definition of $\boldsymbol{\alpha}^e(a^*; X)$ in (38), and algebra show $\mathbb{E}[A_k | \mathfrak{F}_k] = 0$ and $\sup_{q \in \mathbb{S}^2} \mathbb{E}[A_k^2 | \mathfrak{F}_k] = o(1)$. Similarly, we can show $\mathbb{E}[B_k | \mathfrak{F}_k] = 0$ and $\sup_{q \in \mathbb{S}^2} \mathbb{E}[B_k^2 | \mathfrak{F}_k] = o(1)$. By Chernozhukov et al. (2018, Lemma

6.1), we conclude A_k and B_k both are $o_p(1)$ uniformly on $q \in \mathbb{S}^2$. Finally, we conclude that $C_k = o_p(1)$ uniformly on $q \in \mathbb{S}^2$ by Cauchy-Schwartz, Assumptions 6 and 7, and the definition of the weights ω_i .

The triangle inequality and Claims 1 and 2 complete the proof of the lemma. *Q.E.D.*

THEOREM B.1: *Let Assumptions 1–7 hold and let $\{\omega_i\}_{i=1}^n$ be random positive weights independent of the data such that $E[\omega_i] = 1$. Then,*

$$\sqrt{n} \left(\widehat{\varepsilon}^\omega(a^*; \widehat{\mathbf{L}}^\omega, \widehat{\boldsymbol{\eta}}) - \varepsilon^* \right) = \mathbb{G}[\omega_i \{ \xi_i^*(a^*; \mathbf{L}, \boldsymbol{\eta}) - \varepsilon^* \}] + o_p(1),$$

where $\xi_i^*(a^*; \mathbf{L}, \boldsymbol{\eta}) = \xi_i(a^*; \mathbf{L}, \boldsymbol{\eta}) + \sum_{g \in \{r, b\}} \Gamma_g^e \left(1 - \frac{\mathbb{1}\{G_i = g\}}{\mu_g} \right)$ and $\Gamma_g^e = \mathbb{E}[\mathbb{1}\{G_i = g\} (\mathbf{L}_{0,i} + a(X_i) \Delta \mathbf{L}_i) \frac{Z_i}{\pi(X_i)}]$.

PROOF: We use Lemma B.13, the Central Limit Theorem, and similar arguments as in the proof of Theorem A.1; therefore, it is omitted. *Q.E.D.*

LEMMA B.14: Let Assumptions 1–7 hold. Then,

$$\left\| \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^\omega - \widehat{\boldsymbol{\beta}}^\omega \right) - \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^{\omega,*} - \widehat{\boldsymbol{\beta}}^{\omega,*} \right) \right\| = o_p(1),$$

where $\widehat{\boldsymbol{\beta}}^{\omega,*}$ and $\widetilde{\boldsymbol{\beta}}^{\omega,*}$ are defined in (62)–(63), respectively.

PROOF: The norm $\| \cdot \|$ on $\ell^\infty(\mathbb{S}^2) \times \mathbb{R}^3$ is the product norm, i.e., $\|(f, \varepsilon)\| = \|f\|_\infty + \|\varepsilon\|_2$, where $\| \cdot \|_2$ is the euclidean norm on \mathbb{R}^3 . By the triangular inequality $\left\| \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^\omega - \widehat{\boldsymbol{\beta}}^\omega \right) - \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^{\omega,*} - \widehat{\boldsymbol{\beta}}^{\omega,*} \right) \right\|$ is lower or equal than

$$\left\| \sqrt{n} \left(\widehat{\boldsymbol{\beta}}^\omega - \widehat{\boldsymbol{\beta}}^{\omega,*} \right) \right\| + \left\| \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^\omega - \widetilde{\boldsymbol{\beta}}^{\omega,*} \right) \right\|. \quad (86)$$

Therefore, it is sufficient to show that each term above is $o_p(1)$.

We use Lemmas B.10 and B.13 to conclude $\left\| \sqrt{n} \left(\widehat{\boldsymbol{\beta}}^{\omega,*} - \widehat{\boldsymbol{\beta}}^\omega \right) \right\| = o_p(1)$. Similarly, we can conclude $\left\| \sqrt{n} \left(\widetilde{\boldsymbol{\beta}}^{\omega,*} - \widetilde{\boldsymbol{\beta}}^\omega \right) \right\| = o_p(1)$ by using the weights $W_i \omega_i^h$ and $W_i \omega_i^\varepsilon$ in Lemmas B.10 and B.13, respectively. This completes the proof. *Q.E.D.*