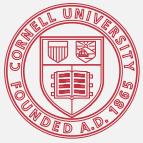
Identification and Estimation for Algorithmic Frontiers with Selective Labels

Yiqi Liu, Francesca Molinari, Amilcar Velez



ESIF-AIML, Ithaca, NY
June 16, 2026

Motivation

- Algorithms play an increasingly prevalent role in everyday life. (e.g., medical treatment, lending)

- In response, regulatory and oversight processes have emerged as part of *AI governance*.

Motivation

- Algorithms play an increasingly prevalent role in everyday life. (e.g., medical treatment, lending)
  - But their performance can vary substantially across subgroups.
- In response, regulatory and oversight processes have emerged as part of *AI governance*.
  - This includes the design, deployment, documentation and monitoring of AI systems.

Motivation

- Algorithms play an increasingly prevalent role in everyday life. (e.g., medical treatment, lending)
  - But their performance can vary substantially across subgroups.
- In response, regulatory and oversight processes have emerged as part of *AI governance*.
- Evaluation of fairness and accuracy plays a central role in AI governance.
- Regulators may ask:
  - For a class of algorithms: When does the **tradeoff** between fairness and accuracy arise?
  - For a specific algorithm: Is there a **less discriminatory alternatives** (LDA)?

Motivation

- Algorithms play an increasingly prevalent role in everyday life. (e.g., medical treatment, lending)
  - But their performance can vary substantially across subgroups.
- In response, regulatory and oversight processes have emerged as part of *AI governance*.
- Evaluation of fairness and accuracy plays a central role in AI governance.
- Regulators may ask:
  - For a class of algorithms: When does the **tradeoff** between fairness and accuracy arise?
  - For a specific algorithm: Is there a **less discriminatory alternatives** (LDA)?

How can we answer these questions with **selectively observed outcomes** and only **finite samples?**

What We Do

- Extend the theoretical framework of Liang, Lu, Mu & Okumura (2025) [LLMO]
  (by allowing for fairness and accuracy to be measured with different loss functions)
  - Characterize a **fairness-accuracy (FA) frontier** using support functions
  - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy

What We Do

- Extend the theoretical framework of Liang, Lu, Mu & Okumura (2025) [LLMO]
  (by allowing for fairness and accuracy to be measured with different loss functions)
  - Characterize a **fairness-accuracy (FA) frontier** using support functions
  - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy
- Extend the inference methods of Liu & Molinari (2025) to allow for selectively observed Y^*:

What We Do

- Extend the theoretical framework of Liang, Lu, Mu & Okumura (2025) [LLMO] (by allowing for fairness and accuracy to be measured with different loss functions)
  - Characterize a **fairness-accuracy (FA) frontier** using support functions
  - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy
- Extend the inference methods of Liu & Molinari (2025) to allow for selectively observed Y^*:
  - Without restrictions on the selection process (binary Y^* and specific loss functions):
    - Derive the **sharp identified set** for the FA frontier
    - Characterize when an LDA exists

What We Do

- Extend the theoretical framework of Liang, Lu, Mu & Okumura (2025) [LLMO]
  (by allowing for fairness and accuracy to be measured with different loss functions)
  - Characterize a **fairness-accuracy (FA) frontier** using support functions
  - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy
- Extend the inference methods of Liu & Molinari (2025) to allow for selectively observed Y^*:
  - Without restrictions on the selection process (binary Y^* and specific loss functions):
    - Derive the **sharp identified set** for the FA frontier
    - Characterize when an LDA exists
  - When labels are missing at random:
    - **Point-identify** the FA frontier using inverse propensity score weighting
    - Construct a DML estimator of the FA frontier
    - Provide a method to test whether an LDA exists for a given algorithm

What We Do

- Extend the theoretical framework of Liang, Lu, Mu & Okumura (2025) [LLMO] (by allowing for fairness and accuracy to be measured with different loss functions)
  - Characterize a **fairness-accuracy (FA) frontier** using support functions
  - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy
- Extend the inference methods of Liu & Molinari (2025) to allow for selectively observed Y^*:
  - Without restrictions on the selection process (binary Y^* and specific loss functions):
    - Derive the **sharp identified set** for the FA frontier
    - Characterize when an LDA exists
  - When labels are missing at random:
    - **Point-identify** the FA frontier using inverse propensity score weighting
    - Construct a DML estimator of the FA frontier
    - Provide a method to test whether an LDA exists for a given algorithm

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by:

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: $(Y^* \quad)$
  - Outcome $Y^* \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$ (number of active chronic conditions and medical costs)

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: $(Y^*, G \quad)$
  - Group $G \in \{r, b\}$ (race)

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: $(Y^*, G \quad)$
  - Covariate inputs $X \in \mathcal{X}$ (age, gender, comorbidity and medication variables, biomarkers, ...)

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: (Y^*, G, X)
- Each individual receives a binary **decision** $d \in \{0, 1\}$.
  (whether automatically enrolled in high-risk care management program)

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: (Y^*, G, X)
- Each individual receives a binary **decision** $d \in \{0, 1\}$.
- **Algorithm** $a : \mathcal{X} \to [0, 1]$ predicts the probability of an event.
  (whether number of active chronic conditions in the subsequent year $\geq 97^{\text{th}}$ percentile)

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: (Y^*, G, X)
- Each individual receives a binary **decision** $d \in \{0, 1\}$.
- **Algorithm** $a : \mathcal{X} \to [0, 1]$ predicts the probability of an event.
  - This prediction determines the probability with which the decision equals 1.

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: (Y^*, G, X)
- Each individual receives a binary **decision** $d \in \{0, 1\}$.
- **Algorithm** $a : \mathcal{X} \to [0, 1]$ predicts the probability of an event.
  - This prediction determines the probability with which the decision equals 1.
- Evaluate the **accuracy** of a through $\ell^A : \{0, 1\} \times \mathcal{Y} \to \mathbb{R}$
  - Example: classification loss $\ell^A(d, Y^*) = \mathbb{1}\{Y_1^* \neq d\}$.

Setup and Notation

Ideal case: suppose Y^* is observed

- A population of individuals is described by: (Y^*, G, X)
- Each individual receives a binary **decision** $d \in \{0, 1\}$.
- **Algorithm** $a : \mathcal{X} \to [0, 1]$ predicts the probability of an event.
  - This prediction determines the probability with which the decision equals 1.
- Evaluate the **accuracy** of a through $\ell^A : \{0, 1\} \times \mathcal{Y} \to \mathbb{R}$
  - Example: classification loss $\ell^A(d, Y^*) = \mathbb{1}\{Y_1^* \neq d\}$.
- Evaluate the **fairness** of a through $\ell^F : \{0, 1\} \times \mathcal{Y} \to \mathbb{R}$
  - Example: statistical parity $\ell^F(d, Y^*) = \mathbb{1}\{d = 1\}$.

Feasible Set

From algorithms to vectors in $\mathbb{R}^3$

- For any algorithm a define its **fairness-accuracy** allocation

$$\varepsilon(a) := \Big(\underbrace{e_r^A(a), e_b^A(a)}_{\textbf{accuracy}}, \underbrace{e_r^F(a) - e_b^F(a)}_{\textbf{fairness}} \Big) \in \mathbb{R}^3.$$

where e_g^ι are group-expected losses induced by algorithm a for $\iota = F, A$

$$e_g^\iota(a) := \mathbb{E}\left[a(X) \cdot \ell^\iota(1, Y^*) + (1 - a(X)) \cdot \ell^\iota(0, Y^*) | G = g\right]$$

Feasible Set

From algorithms to vectors in $\mathbb{R}^3$

- For any algorithm a define its **fairness-accuracy** allocation

$$\varepsilon(a) := \Big(\underbrace{e_r^A(a), e_b^A(a)}_{\textbf{accuracy}}, \underbrace{e_r^F(a) - e_b^F(a)}_{\textbf{fairness}} \Big) \in \mathbb{R}^3.$$

  where e_g^ι are group-expected losses induced by algorithm a for $\iota = F, A$

$$e_g^\iota(a) := \mathbb{E}\left[a(X) \cdot \ell^\iota(1, Y^*) + (1 - a(X)) \cdot \ell^\iota(0, Y^*) | G = g\right]$$

- Given $\mathcal{A}(\mathcal{X}) := \{a : \mathcal{X} \to [0, 1]\}$, the **feasible set** of fairness-accuracy allocations is defined as

$$\mathcal{E} := \{\varepsilon(a) : a \in \mathcal{A}(\mathcal{X})\}$$

- $\mathcal{E}$ is a **closed and convex** set in $\mathbb{R}^3$.

Preferences & FA-Frontier

Specify preferences over fairness-accuracy allocations.

- $\varepsilon = \left(e_r^A, e_b^A, e_r^F - e_b^F\right)$ and $\tilde{\varepsilon} = \left(\tilde{\varepsilon}_r^A, \tilde{\varepsilon}_b^A, \tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F\right)$

Preferences & FA-Frontier

Specify preferences over fairness-accuracy allocations.

- $\varepsilon = \left(e_r^A, e_b^A, e_r^F - e_b^F\right)$ and $\tilde{\varepsilon} = \left(\tilde{\varepsilon}_r^A, \tilde{\varepsilon}_b^A, \tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F\right)$

- ε is FA-preferred over $\tilde{\varepsilon}$ if

$$\underbrace{\varepsilon_r^A \leq \tilde{\varepsilon}_r^A,\ \varepsilon_b^A \leq \tilde{\varepsilon}_b^A}_{\varepsilon \text{ is more accurate}} \quad \text{and} \quad \underbrace{\left|\varepsilon_r^F - \varepsilon_b^F\right| \leq \left|\tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F\right|}_{\varepsilon \text{ is fairer}}.$$

  with at least one inequality strict.

Preferences & FA-Frontier

Specify preferences over fairness-accuracy allocations.

- $\varepsilon = \left(e_r^A, e_b^A, e_r^F - e_b^F\right)$ and $\tilde{\varepsilon} = \left(\tilde{\varepsilon}_r^A, \tilde{\varepsilon}_b^A, \tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F\right)$

- ε is FA-preferred over $\tilde{\varepsilon}$ if

$$\underbrace{\varepsilon_r^A \leq \tilde{\varepsilon}_r^A,\ \varepsilon_b^A \leq \tilde{\varepsilon}_b^A}_{\varepsilon \text{ is more accurate}} \quad \text{and} \quad \underbrace{\left|\varepsilon_r^F - \varepsilon_b^F\right| \leq \left|\tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F\right|}_{\varepsilon \text{ is fairer}}.$$

  with at least one inequality strict.

- The **Fairness-Accuracy Frontier** is defined as

$$\mathcal{F} := \{\varepsilon \in \mathcal{E} : \nexists \tilde{\varepsilon} \in \mathcal{E} \ s.t.\ \tilde{\varepsilon} \succ_{FA} \varepsilon\}$$

Preferences & Pareto Frontier

Specify preferences over accuracy allocations.

- $\varepsilon = (e_r^A, e_b^A, e_r^F - e_b^F)$ and $\tilde{\varepsilon} = (\tilde{\varepsilon}_r^A, \tilde{\varepsilon}_b^A, \tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F)$
- ε is preferred over $\tilde{\varepsilon}$ if

$$\underbrace{\varepsilon_r^A \leq \tilde{\varepsilon}_r^A,\ \varepsilon_b^A \leq \tilde{\varepsilon}_b^A}_{\varepsilon \text{ is more accurate}} \quad \text{and} \quad \underbrace{\left|\varepsilon_r^F - \varepsilon_b^F\right| \leq \left|\tilde{\varepsilon}_r^F - \tilde{\varepsilon}_b^F\right|}_{\varepsilon \text{ is fairer}}.$$

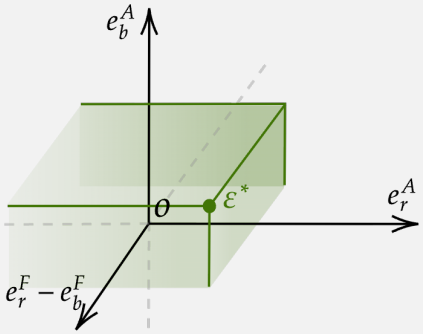
  with at least one inequality strict.

- The **Pareto Frontier** is defined as

$$\mathcal{PF} := \{\varepsilon \in \mathcal{E} : \nexists\, \tilde{\varepsilon} \in \mathcal{E} \ s.t.\ \tilde{\varepsilon} \succ_{PD} \varepsilon\}$$

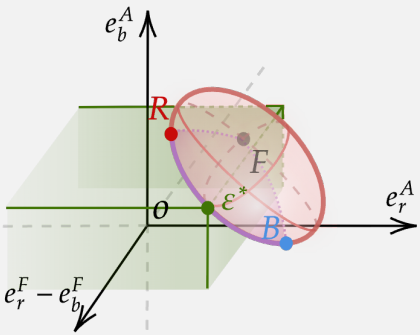
Expressing the FA-Frontier $\mathcal{F}$ Through the Support Function $h_{\mathcal{E}}(q)$

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

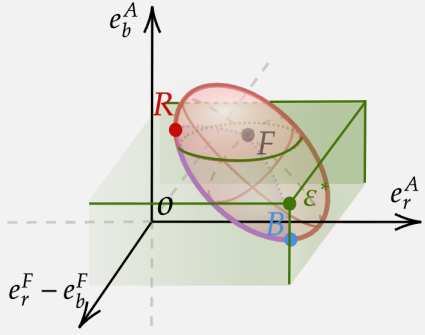


Expressing the FA-Frontier $\mathcal{F}$ Through the Support Function $h_{\mathcal{E}}(q)$

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



(a) ε^* is on $\mathcal{F}$



(b) ε^* is off $\mathcal{F}$

We characterize the set of points on $\mathcal{F}$ by judicious use of the **separating hyperplane theorem**.

Expressing the FA-Frontier $\mathcal{F}$ Through the Support Function $h_{\mathcal{E}}(q)$

- Let $h_{\mathcal{C}(\varepsilon^*)}(q) = \max_{\varepsilon \in \mathcal{C}(\varepsilon^*)} q^{\mathsf{T}}\varepsilon$ be the support function of $\mathcal{C}(\varepsilon^*)$
- Let $h_{\mathcal{E}}(q) = \max_{\varepsilon \in \mathcal{E}} q^{\mathsf{T}}\varepsilon$ be the support function of $\mathcal{E}$

Theorem

When $\mathcal{E}$ is strictly convex, $\varepsilon^* \in \mathcal{F}$ if and only if there exists a hyperplane that properly separates $\mathcal{C}(\varepsilon^*)$ and $\mathcal{E}$, i.e., there exists $q \in \mathbb{S}^2$ such that

$$h_{\mathcal{C}(\varepsilon^*)}(q) = -h_{\mathcal{E}}(-q).$$

Expressing the FA-Frontier $\mathcal{F}$ Through the Support Function $h_{\mathcal{E}}(q)$

- Let $h_{\mathcal{C}(\varepsilon^*)}(q) = \max_{\varepsilon \in \mathcal{C}(\varepsilon^*)} q^{\mathsf{T}}\varepsilon$ be the support function of $\mathcal{C}(\varepsilon^*)$
- Let $h_{\mathcal{E}}(q) = \max_{\varepsilon \in \mathcal{E}} q^{\mathsf{T}}\varepsilon$ be the support function of $\mathcal{E}$

Theorem

When $\mathcal{E}$ is strictly convex, $\varepsilon^* \in \mathcal{F}$ if and only if there exists a hyperplane that properly separates $\mathcal{C}(\varepsilon^*)$ and $\mathcal{E}$, i.e., there exists $q \in \mathbb{S}^2$ such that

$$h_{\mathcal{C}(\varepsilon^*)}(q) = -h_{\mathcal{E}}(-q).$$

We then have that

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

What We Do

- Extend the theoretical framework of Liang, Lu, Mu & Okumura (2025) [LLMO]
  (by allowing for fairness and accuracy to be measured with different loss functions)

  ✓ Characterize a **fairness-accuracy (FA) frontier** using support functions

  - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy

- Extend the inference methods of Liu & Molinari (2025) to allow for selectively observed Y^*:

  - Without restrictions on the selection process (binary Y^* and specific loss functions):

    - Derive the **sharp identified set** for the FA frontier
    - Characterize when an LDA exists

  - When labels are missing at random:

    - **Point-identify** the FA frontier using inverse propensity score weighting
    - Construct a DML estimator of the FA frontier
    - Provide a method to test whether an LDA exists for a given algorithm

Selectively Observed Y^*

Observe $(Y, G, X, Z) \sim \mathbb{P}$ with $Y \equiv ZY^*$, $Z \in \{0, 1\}$

For simplicity, let:

$$Y^* \in \{0, 1\}$$

$$\ell^A(d, Y^*) = \mathbb{1}\{d \neq Y^*\} \quad \text{(classification loss)} \rightarrow \text{selectively observed!}$$

$$\ell^F(d, Y^*) = \mathbb{1}\{1 = d\} \quad \text{(statistical disparity loss)}$$

Selectively Observed Y^*

Observe $(Y, G, X, Z) \sim \mathbb{P}$ with $Y \equiv ZY^*$, $Z \in \{0, 1\}$

For simplicity, let:

$$Y^* \in \{0, 1\}$$

$$\ell^A(d, Y^*) = \mathbb{1}\{d \neq Y^*\} \quad \text{(classification loss)} \rightarrow \text{selectively observed!}$$

$$\ell^F(d, Y^*) = \mathbb{1}\{1 = d\} \quad \text{(statistical disparity loss)}$$

- The following objects are *not* point identified:

  $\Rightarrow$ $\lambda_g(X) := \mathbb{P}[\, Y^* = 1 \,|\, G = g, Z = 0, X] \in [0, 1]$.

  $\Rightarrow$ The feasible allocation $\varepsilon^*(a^*; \lambda) = \left(e_r^A(a^*; \lambda), e_b^A(a^*; \lambda), e_r^F(a^*) - e_b^F(a^*)\right)$

  $\Rightarrow$ The support function $h_{\mathcal{E}}(q; \lambda)$

  $\Rightarrow$ The FA-frontier $\mathcal{F}(\lambda)$.

Selectively Observed Y^*: The Fundamental Challenge

Observe $Y \equiv ZY^*$ for $Z \in \{0, 1\}$, assume $Y^* \in \{0, 1\}$

The goal of the analysis is to ascertain properties of *algorithms*.

Selectively Observed Y^*: The Fundamental Challenge

Observe $Y \equiv ZY^*$ for $Z \in \{0,1\}$, assume $Y^* \in \{0,1\}$

The goal of the analysis is to ascertain properties of *algorithms*.

- For given $a^* \in \mathcal{A}(\mathcal{X})$, the empirical evidence is contained in $\varepsilon(a^*)$ and $\mathcal{F}$.
- Neither is point identified, but they are linked by the distribution of the missing data, λ.

Selectively Observed Y^*: The Fundamental Challenge

Observe $Y \equiv ZY^*$ for $Z \in \{0,1\}$, assume $Y^* \in \{0,1\}$

The goal of the analysis is to ascertain properties of *algorithms*.

- For given $a^* \in \mathcal{A}(\mathcal{X})$, the empirical evidence is contained in $\varepsilon(a^*)$ and $\mathcal{F}$.
- Neither is point identified, but they are linked by the distribution of the missing data, λ.

To determine if a^* yields a fairness-accuracy allocation $\varepsilon(a^*;\lambda)$ on a candidate FA frontier $\mathcal{F}(\lambda')$,

- Must use the *same* selection mechanism/completion ($\lambda = \lambda'$),

$$\lambda_g(X) := \mathbb{P}[Y^* = 1 | G = g, Z = 0, X] \in [0,1], \quad g \in \{r, b\}$$

- Must loop over all possible $\lambda_g : \mathcal{X} \mapsto [0,1]$, $g \in \{r, b\}$.

Selectively Observed Y^*: Algorithms on the FA Frontier

We reduce the task to a finite dimensional optimization problem

Theorem

$\varepsilon^* \equiv \varepsilon(a^*) \in \mathcal{F}$ *for some admissible distribution of* Y^* *if and only if* $\exists\, \lambda : \mathcal{X} \to [0,1]^2$ *such that*

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

Selectively Observed Y^*: Algorithms on the FA Frontier

We reduce the task to a finite dimensional optimization problem

Theorem

$\varepsilon^* \equiv \varepsilon(a^*) \in \mathcal{F}$ *for some admissible distribution of* Y^* *if and only if* $\exists\, \lambda : \mathcal{X} \to [0,1]^2$ *such that*

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

This occurs if and only if the next finite dimensional optimization problem equals zero

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

for $\lambda_0 = (0,0)^\intercal$; $\lambda_1 = (1,1)^\intercal$; $J_d(\lambda, q, a)$, $d \in \{0,1\}$, *known/estimable functions of the arguments.*

Selectively Observed Y^*: Algorithms on the FA Frontier

We reduce the task to a finite dimensional optimization problem

Theorem

$\varepsilon^* \equiv \varepsilon(a^*) \in \mathcal{F}$ *for some admissible distribution of* Y^* *if and only if* $\exists\, \lambda : \mathcal{X} \to [0,1]^2$ *such that*

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

This occurs if and only if the next finite dimensional optimization problem equals zero

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

for $\lambda_0 = (0,0)^\intercal$; $\lambda_1 = (1,1)^\intercal$; $J_d(\lambda, q, a)$, $d \in \{0,1\}$, *known/estimable functions of the arguments.*

- The paper derives the sharp-identified set of $\mathcal{F}$ free of λ

Conclusion

- This paper allows for fairness and accuracy to be measured with different loss functions
  - ✓ Characterize a **fairness-accuracy (FA) frontier** using support functions
    - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy
- Without restrictions on the selection process (binary Y^* and specific loss functions):
  - Derive the **sharp identified set** for the FA frontier
  - ✓ Characterize when an LDA exists
- When labels are missing at random:
  - **Point-identify** and estimate the FA frontier using IPW and DML approach
  - Provide a method to test whether an LDA exists for a given algorithm
- In Progress:
  - Test for $\mathcal{PF} = \mathcal{F}$ under partial identification or MAR.
  - Empirical application + package.

Conclusion

- This paper allows for fairness and accuracy to be measured with different loss functions
  - ✓ Characterize a **fairness-accuracy (FA) frontier** using support functions
    - Provide an **IFF** condition for when improving fairness does **not** come at the cost of reducing accuracy
- Without restrictions on the selection process (binary Y^* and specific loss functions):
  - Derive the **sharp identified set** for the FA frontier
  - ✓ Characterize when an LDA exists
- When labels are missing at random:
  - **Point-identify** and estimate the FA frontier using IPW and DML approach
  - Provide a method to test whether an LDA exists for a given algorithm
- In Progress:
  - Test for $\mathcal{PF} = \mathcal{F}$ under partial identification or MAR.
  - Empirical application + package.

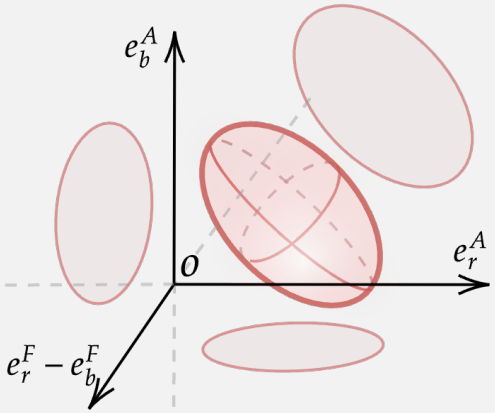
Thanks for listening!

Paper & slides here

Comments/questions are welcome!

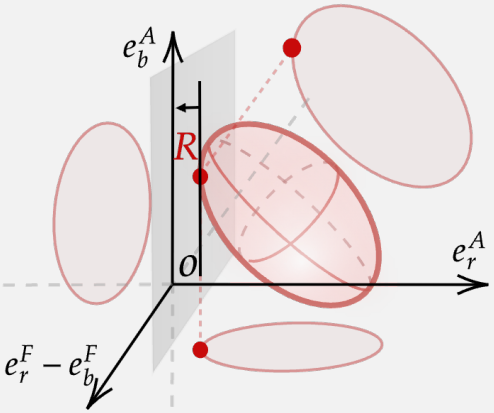
Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)

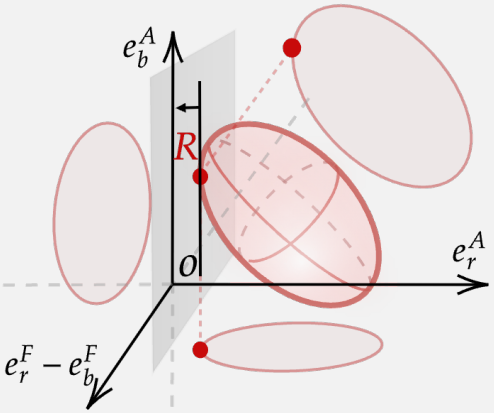


Best point for group r:

$$R := \arg\min_{\varepsilon \in \mathcal{E}} \varepsilon_1$$

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



Best point for group r:

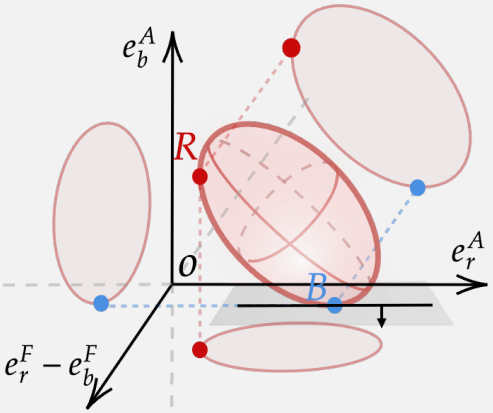
$$R := \arg\min_{\varepsilon \in \mathcal{E}} \varepsilon_1$$

R is the **support set** of $\mathcal{E}$ in direction $q = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$:

$$\mathcal{S}_{\mathcal{E}}(q) = \arg\max_{\varepsilon \in \mathcal{E}} q^{\mathsf{T}} \varepsilon$$

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



Best point for group b:

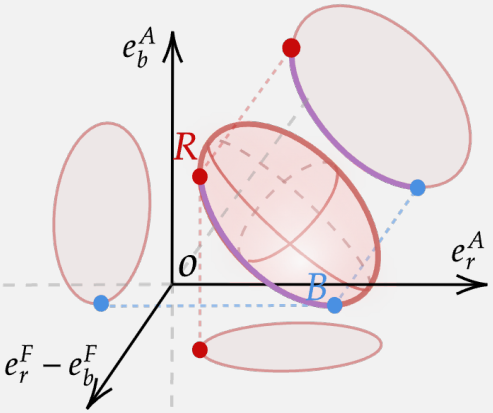
$$B := \arg\min_{\varepsilon \in \mathcal{E}} \varepsilon_2$$

B is the **support set** of $\mathcal{E}$ in direction $q = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$:

$$\mathcal{S}_{\mathcal{E}}(q) := \arg\max_{\varepsilon \in \mathcal{E}} q^{\mathsf{T}} \varepsilon$$

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)

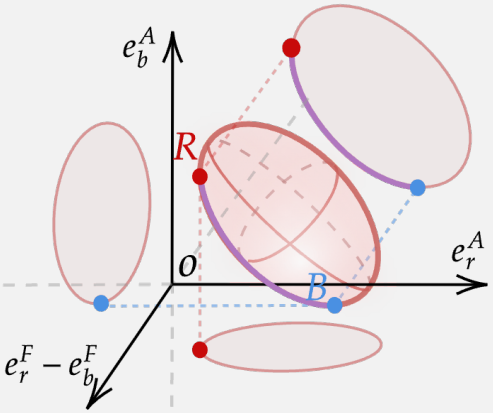


The **Pareto Frontier** is given by the segment connecting R and B

$$\mathcal{PF} = \{\mathcal{S}_{\mathcal{E}}(q) : q \in \mathbb{S}^2, q_1 \leq 0, q_2 \leq 0, \; q_3 = 0\}$$

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



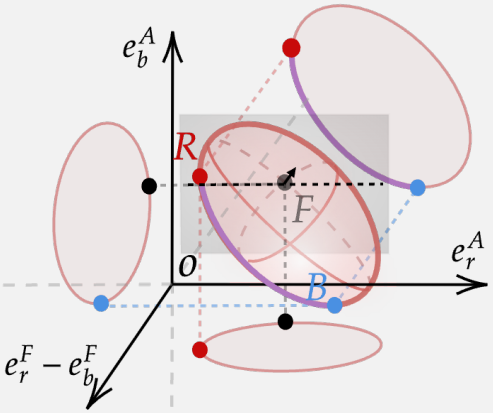
The **Pareto Frontier** is given by the segment connecting R and B

$$\mathcal{PF} = \left\{ \mathcal{S}_{\mathcal{E}}(q) : q \in \mathbb{S}^2, q_1 \leq 0, q_2 \leq 0, \; q_3 = 0 \right\}$$

But we also care about **Fairness**!

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



When $\mathcal{E} \subset \{e_r^F - e_b^F > 0\}$, the fairest point F:

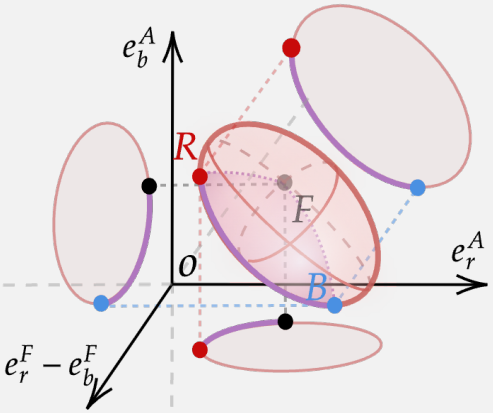
$$F := \arg\min_{\varepsilon \in \mathcal{E}} |\varepsilon_3|$$

F is the **support set** of $\mathcal{E}$ in direction $q = [0, 0, -1]^\intercal$:

$$\mathcal{S}_{\mathcal{E}}(q) := \arg\max_{\varepsilon \in \mathcal{E}} q^\intercal \varepsilon$$

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



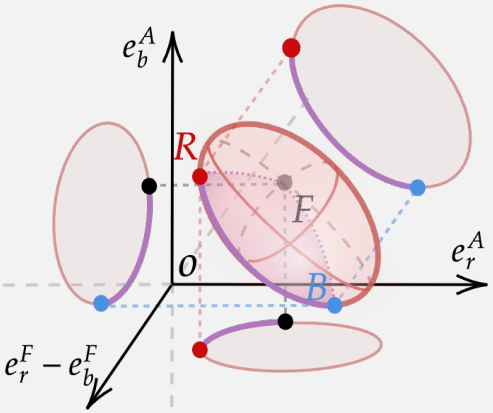
The **FA Frontier** is given by

- "usual" Pareto frontier connecting R and B
- "segment" connecting R and F
- "segment" connecting B and F

and everything in between...equivalently

Feasible Set $\mathcal{E} \subset \mathbb{R}^3$ and Points of Interest

Assume $\mathcal{E}$ is strictly convex (back1,back2,back3)



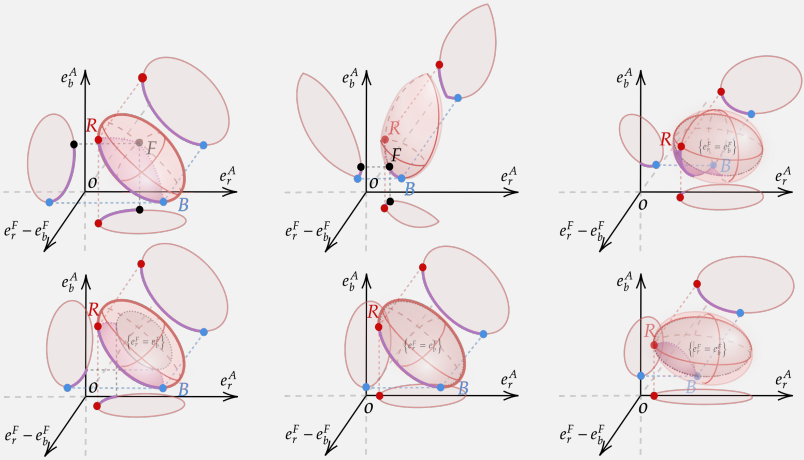
The **FA Frontier** is given by

- "usual" Pareto frontier connecting R and B
- "segment" connecting R and F
- "segment" connecting B and F

and everything in between...equivalently

But...This is only when $\mathcal{E} \subset \{e_r^F - e_b^F > 0\}$

Geometry of $\mathcal{F}$ in $\mathbb{R}^3$ Is...Complicated



We provide a simple characterization of $\mathcal{F}$ **without** pre-testing which case we are in!

And an **if and only if** condition for $\mathcal{PF} = \mathcal{F}$ (not today).

Support Function when Y^* is Perfectly Observed

By definition, $h_{\mathcal{E}}(q) = \max_{e \in \mathcal{E}} q^\intercal e$, $q \in \mathbb{S}^2 = \{q \in \mathbb{R}^3 : ||q|| = 1\}$.

For $g \in \{r, b\}$ and $d \in \{0, 1\}$, let:

$$
\begin{aligned}
\mu_g &\equiv \mathbb{P}(G = g) \\
L_d^F &\equiv \ell^F(d, Y^*) \left(\frac{\mathbb{1}\{G=r\}}{\mu_r} - \frac{\mathbb{1}\{G=b\}}{\mu_b} \right) \\
L_d^{g,A} &\equiv \ell^A(d, Y^*) \frac{\mathbb{1}\{G=g\}}{\mu_g} \\
\theta_d^F(X) &\equiv \mathbb{E}[L_d^F | X] \\
\theta_d^{g,A}(X) &\equiv \mathbb{E}[L_d^{g,A} | X]
\end{aligned}
$$

Denoting $\mathbf{L}_d := \left[L_d^{r,A},\ L_d^{b,A},\ L_d^F \right]^\intercal$, algebraic manipulations yield

$$
h_{\mathcal{E}}(q) = \mathbb{E}\big[q^\intercal \mathbf{L}_0 + q^\intercal (\mathbf{L}_1 - \mathbf{L}_0) \underbrace{\mathbb{1}\{q^\intercal(\theta_1(X) - \theta_0(X)) > 0\}}_{a_q^{\text{opt}}(X)} \big]
$$

Missing at Random (MAR)

Recall: $\ell^{\iota}(d, Y^*)\mathbb{1}\{G = g\}/\mu_g$ is selectively observed, for $\iota = A, F$

If one is willing to assume for **generic** $Y^* \in \mathcal{Y}$ and loss functions:

Assumption (MAR)

$(Y^*, G) \perp Z \mid X$ *and* $\pi(X) := \mathbb{E}[Z|X] \in (0, 1)$.

Then, the following objects are point identified (after adjusting for $\pi(X)$):

- The feasible allocation $\varepsilon^*(a^*) = (e_r^A(a^*), e_b^A(a^*), e_r^F(a^*) - e_b^F(a^*))$
- The support function $h_{\mathcal{E}}(q)$ and also the FA-frontier $\mathcal{F}$

The paper provides:

- a DML estimator for the FA-frontier using on inverse propensity score weighting, $\hat{h}_{\mathcal{E}}(q)$
- a test for LDA existence based on $\mathrm{H}_0 : \min_{q \in \mathbb{S}^2}(h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q)) = 0$

Are There Less Discriminatory Alternatives to a Given Algorithm?

- Given an existing algorithm with $\varepsilon^* \in \mathcal{E}$, **we would like to test**:

$$\begin{aligned} \mathrm{H}_0 &: \varepsilon^* \in \mathcal{F} \\ \mathrm{H}_1 &: \varepsilon^* \notin \mathcal{F} \end{aligned}$$

- Based on previous derivation, we can express this hypothesis as

$$\begin{aligned} \mathrm{H}_0 &: \min_{q \in \mathbb{S}^2} \left(h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q) \right) = 0 \\ \mathrm{H}_1 &: \min_{q \in \mathbb{S}^2} \left(h_{\mathcal{C}(\varepsilon^*)}(q) + h_{\mathcal{E}}(-q) \right) \neq 0 \end{aligned}$$

- We build on
  - Liu & Molinari (2025) to derive a statistic for testing H_0.
  - Fang & Santos (2019) to prove size control and propose a consistent bootstrap critical value.

Asymptotic Theory: Main Assumptions

Assumption (Margin Condition)

$\exists\ 0 < m \leq 1$ *such that* $\forall\ \delta > 0$, *we have* $\sup_{q \in \mathbb{S}^2} \mathbb{P}\big(|q^\intercal(\boldsymbol{\theta}_1(X) - \boldsymbol{\theta}_0(X))| < \delta\big) \lesssim \delta^m$.

Assumption (Behavior of Nuisance Parameters)

We can partition $X = (X_1, X_2)$ *such that (1) the density of* $|q^\intercal \Delta\boldsymbol{\theta}(X)|$ *conditional on* X_2 *is uniformly bounded in* q*; the variance of* $|q^\intercal \Delta\boldsymbol{\theta}(X)|$ *conditional on* X_2 *is uniformly (in* q*) bounded away from zero; (3)* $(\pi(X), \Delta\boldsymbol{\theta}(X)) = F(\nu(X_1), \gamma(X_2))$ *for unknown functions* ν *and* γ *and known link function* F*, where the derivatives of* F *with respective to* (ν, γ) *are uniformly bounded.*

Assumption (Rate Requirement)

The k*-th fold estimators satisfy* $\|\widehat{\nu}_k(X_1) - \nu(X_1)\|_\infty = o_p(n^{-1/4})$, $\|\widehat{\gamma}_k(X_2) - \gamma(X_2)\|_{L^2} = o_p(n^{-1/4})$.