Appendix for **Delayed Impact of Fair Machine Learning**

A Further Examples

We present two more examples of the problem formulation in Sec. 2, showing its general applicability to many domains.

Example A.1 (Advertising)**.** A second illustrative example is given by the case of advertising agencies making decisions about which groups to target. An individual with product interest score x responds positively to an ad with probability $\boldsymbol{\rho}(x)$. The ad agency experiences utility $\boldsymbol{u}(x)$ related to click-through rates, which increases with $\boldsymbol{\rho}(x)$. Individuals who see the ad but are uninterested may react negatively (becoming less interested in the product), and $\boldsymbol{\Delta}(x)$ encodes the interest change. If the product is a positive good like education or employment opportunities, interest can correspond to well-being. Thus the advertising agency's incentives to only show ads to individuals with extremely high interest may leave behind groups whose interest is lower on average. A related historical example occurred in advertisements for computers in the 1980s, where male consumers were targeted over female consumers, arguably contributing to the current gender gap in computing.

Example A.2 (College Admissions)**.** The scenario of college admissions or scholarship allotments can also be considered within our framework. Colleges may select certain applicants for acceptance according to a score x, which could be thought encode a "college preparedness" measure. The students who are admitted might "succeed" (this could be interpreted as graduating, graduating with honors, finding a job placement, etc.) with some probability $\boldsymbol{\rho}(x)$ depending on their preparedness. The college might experience a utility $\boldsymbol{u}(x)$ corresponding to alumni donations, or positive rating when a student succeeds; they might also show a drop in rating or a loss of invested scholarship money when a student is unsuccessful. The student's success in college will affect their later success, which could be modeled generally by $\boldsymbol{\Delta}(x)$. In this scenario, it is challenging to ensure that a single summary statistic x captures enough information about a student; it may be more appropriate to consider x as a vector as well as more complex forms of $\boldsymbol{\rho}(x)$.

While a variety of applications are modeled faithfully within our framework, there are limitations to the accuracy with which real-life phenomenon can be measured by strictly binary decisions and success probabilities. Such binary rules are necessary for the definition and execution of existing fairness criteria, (see Sec. 2.2) and as we will see, even modeling these facets of decision making as binary allows for complex and interesting behavior.

B Optimality of Threshold Policies

Next, we move towards statements of the main theorems underlying the results presented in Section 3. We begin by establishing notation which we shall use throughout. Recall that $\circ$ denotes the Hadamard product between vectors. We identify functions mapping $\mathcal{X} \to \mathbb{R}$ with vectors in $\mathbb{R}^C$. We also define the group-wise utilities

$$\mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathsf{j}}(x) \boldsymbol{u}(x) \ , \tag{1}$$

so that for $\boldsymbol{\tau} = (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$, $\mathcal{U}(\boldsymbol{\tau}) := g_{\mathsf{A}} \mathcal{U}_{\mathsf{A}}(\boldsymbol{\tau}_{\mathsf{A}}) + g_{\mathsf{B}} \mathcal{U}_{\mathsf{B}}(\boldsymbol{\tau}_{\mathsf{B}})$.

First, we formally describe threshold policies, and rigorously justify why we may always assume without loss of generality that the institution adopts policies of this form.

Definition B.1 (Threshold selection policy)**.** A single group selection policy $\boldsymbol{\tau} \in [0,1]^C$ is called a *threshold policy* if it has the form of a randomized threshold on score:

$$\boldsymbol{\tau}_{c,\gamma} = \begin{cases} 1, & x > c \\ \gamma, & x = c \\ 0, & x < c \end{cases} \ , \text{ for some } \ c \in [C] \text{ and } \gamma \in (0,1] \ . \tag{2}$$

As a technicality, if no members of a population have a given score $x \in \mathcal{X}$, there may be multiple threshold policies which yield equivalent selection rates for a given population. To avoid redundancy, we introduce the notation $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ to mean that the set of scores on which $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ differ has probability 0 under $\boldsymbol{\pi}_{\mathsf{j}}$; formally, $\sum_{x : \boldsymbol{\tau}_{\mathsf{j}}(x) \neq \boldsymbol{\tau}_{\mathsf{j}}(x)} \boldsymbol{\pi}_{\mathsf{j}}(x) = 0$. For any distribution $\boldsymbol{\pi}_{\mathsf{j}}$, $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$ is an equivalence relation. Moreover, we see that if $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$, then $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ both provide the same utility for the institution, induce the same outcomes for individuals in group j, and have the same selection and true positive rates. Hence, if $(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$ is an optimal solution to any of `MaxUtil`, `EqOpt`, or `DemParity`, so is any $(\boldsymbol{\tau}'_{\mathsf{A}}, \boldsymbol{\tau}'_{\mathsf{B}})$ for which $\boldsymbol{\tau}_{\mathsf{A}} \cong_{\boldsymbol{\pi}_{\mathsf{A}}} \boldsymbol{\tau}'_{\mathsf{A}}$ and $\boldsymbol{\tau}_{\mathsf{B}} \cong_{\boldsymbol{\pi}_{\mathsf{B}}} \boldsymbol{\tau}'_{\mathsf{B}}$.

For threshold policies in particular, their equivalence class under $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$ is uniquely determined by the selection rate function,

$$r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x) \boldsymbol{\tau}_{\mathsf{j}}(x) \ , \tag{3}$$

which denotes the fraction of group j which is selected. Indeed, we have the following lemma (proved in Appendix D.1):

Lemma B.1. *Let $\boldsymbol{\tau}_{\mathsf{j}}$ and $\boldsymbol{\tau}'_{\mathsf{j}}$ be threshold policies. Then $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ if and only if $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$. Further, $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$ is a bijection from $\mathcal{T}_{\mathrm{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ to $[0,1]$, where $\mathcal{T}_{\mathrm{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ is the set of equivalence classes between threshold policies under $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$. Finally, $\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta_{\mathsf{j}})$ is well defined.*

Remark that $r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta_{\mathsf{j}})$ is an equivalence class rather than a single policy. However, $\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\boldsymbol{\tau}_{\mathsf{j}})$ is well defined, meaning that $\boldsymbol{\pi}_{\mathsf{j}} \circ \boldsymbol{\tau}_{\mathsf{j}} = \boldsymbol{\pi}_{\mathsf{j}} \circ \boldsymbol{\tau}'_{\mathsf{j}}$ for any two policies in the same equivalence class. Since all quantities of interest will only depend on policies $\boldsymbol{\tau}_{\mathsf{j}}$ through $\boldsymbol{\pi}_{\mathsf{j}} \circ \boldsymbol{\tau}_{\mathsf{j}}$, it does not matter *which* representative of $r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta_{\mathsf{j}})$ we pick. Hence, abusing notation slightly, we shall represent $\mathcal{T}_{\mathrm{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ by choosing one representative from each equivalence class under $\cong_{\boldsymbol{\pi}_{\mathsf{j}}}$[1].

It turns out the policies which arise in this away are always optimal in the sense that, for a given loan rate β_j, the threshold policy $r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta_j)$ is the (essentially unique) policy which maximizes

[1] One way to do this is to consider the set of all threshold policies $\boldsymbol{\tau}_{c,\gamma}$ such that, $\gamma = 1$ if $\boldsymbol{\pi}_{\mathsf{j}}(c) = 0$ and $\boldsymbol{\pi}_{\mathsf{j}}(c-1) > 0$ if $\gamma = 1$ and $c > 1$.

both the institution's utility and the utility of the group. Defining the group-wise utility,

$$\mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \boldsymbol{u}(x)\boldsymbol{\pi}_{\mathsf{j}}(x)\boldsymbol{\tau}_{\mathsf{j}}(x) \ , \tag{4}$$

we have the following result:

Proposition B.2 (Threshold policies are preferable)**.** *Suppose that $\boldsymbol{u}(x)$ and $\boldsymbol{\Delta}(x)$ are strictly increasing in x. Given any loaning policy $\boldsymbol{\tau}_{\mathsf{j}}$ for population with distribution $\boldsymbol{\pi}_j$, then the policy $\boldsymbol{\tau}_{\mathsf{j}}^{\text{thresh}} := r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})) \in \mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_{\mathsf{j}})$ satisfies*

$$\Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}^{\text{thresh}}) \geq \Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) \ \textit{and} \ \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}^{\text{thresh}}) \geq \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) \ . \tag{5}$$

Moreover, both inequalities hold with equality if and only if $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}_{\mathsf{j}}^{\text{thresh}}$.

The map $\boldsymbol{\tau}_{\mathsf{j}} \mapsto r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$ can be thought of transforming an arbitrary policy $\boldsymbol{\tau}_{\mathsf{j}}$ into a threshold policy with the same selection rate. In this language, the above proposition states that this map never reduces institution utility or individual outcomes. We can also show that optimal `MaxUtil` and `DemParity` policies are threshold policies, as well as all `EqOpt` policies under an additional assumption:

Proposition B.3 (Existance of optimal threshold policies under fairness constraints)**.** *Suppose that $\boldsymbol{u}(x)$ is strictly increasing in x. Then all optimal* `MaxUtil` *policies $(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})$ satisfy $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}\left(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})\right)$ for $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$. The same holds for all optimal* `DemParity` *policies, and if in addition $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing, the same is true for all optimal* `EqOpt` *policies.*

To prove proposition B.2, we invoke the following general lemma which is proved using standard convex analysis arguments (in Appendix D.2):

Lemma B.4. *Let $\boldsymbol{v} \in \mathbb{R}^C$, and let $\boldsymbol{w} \in \mathbb{R}^C_{>0}$, and suppose either that $\boldsymbol{v}(x)$ is increasing in x, and $\boldsymbol{v}(x)/\boldsymbol{w}(x)$ is increasing or, $\forall x \in \mathcal{X}$, $\boldsymbol{w}(x) = 0$. Let $\boldsymbol{\pi} \in \text{Simplex}^{C-1}$ and fix $t \in [0, \sum_{x\in\mathcal{X}} \boldsymbol{\pi}(x) \cdot \boldsymbol{w}(x)]$. Then any*

$$\boldsymbol{\tau}^* \in \arg \max_{\boldsymbol{\tau} \in [0,1]^C} \langle \boldsymbol{v} \circ \boldsymbol{\pi}, \boldsymbol{\tau} \rangle \quad \text{s.t.} \quad \langle \boldsymbol{\pi} \circ \boldsymbol{w}, \boldsymbol{\tau} \rangle = t \tag{6}$$

satisfies $\boldsymbol{\tau}^* \cong_{\boldsymbol{\pi}} r_{\boldsymbol{\pi}}^{-1}(r_{\boldsymbol{\pi}}(\boldsymbol{\tau}^*))$. Moreover, at least one maximizer $\boldsymbol{\tau}^* \in \mathcal{T}_{\text{thresh}}(\boldsymbol{\pi})$ exists.

Proof of Proposition B.2. We will first prove Proposition B.2 for the function $\mathcal{U}_{\mathsf{j}}$. Given our nominal policy $\boldsymbol{\tau}_{\mathsf{j}}$, let $\beta_j = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$. We now apply Lemma B.4 with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = 1$. For this choice of $\boldsymbol{v}$ and $\boldsymbol{w}$, $\langle \boldsymbol{v}, \boldsymbol{\tau} \rangle = \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau})$ and that $\langle \boldsymbol{\pi}_j \circ \boldsymbol{w}, \boldsymbol{\tau} = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau})$. Then, if $\boldsymbol{\tau}_{\mathsf{j}} \in \arg\max_{\boldsymbol{\tau}} \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau})$ s.t. $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}) = \beta_{\mathsf{j}}$, Lemma 6 implies that $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$.

On the other hand, assume that $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}\left(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})\right)$. We show that $r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$ is a maximizer; which will imply that $\boldsymbol{\tau}_{\mathsf{j}}$ is a maximizer since $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$ implies that $\mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) = \boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$. By Lemma B.4 there exists a maximizer $\boldsymbol{\tau}_{\mathsf{j}}^* \in \mathcal{T}_{\text{thresh}}(\boldsymbol{\pi})$, which means that $\boldsymbol{\tau}_{\mathsf{j}}^* = r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}^*))$. Since $\boldsymbol{\tau}_{\mathsf{j}}^*$ is feasible, we must have $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}^*) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}})$, and thus $\boldsymbol{\tau}_{\mathsf{j}}^* = r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}))$, as needed. The same argument follows verbatim if we instead choose $\boldsymbol{v}(x) = \boldsymbol{\Delta}(x)$, and compute $\langle \boldsymbol{v}, \boldsymbol{\tau} \rangle = \Delta\boldsymbol{\mu}_{\mathsf{j}}(\boldsymbol{\tau})$. □

We now argue Proposition B.3 for `MaxUtil`, as it is a straightforward application of Lemma B.4. We will prove Proposition B.3 for `DemParity` and `EqOpt` separately in Sections C.1 and C.2.

Proof of Proposition B.3 for `MaxUtil`. `MaxUtil` follows from lemma B.4 with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$, and $t = 0$ and $\boldsymbol{w} = \mathbf{0}$.

□

B.1 Quantiles and Concavity of the Outcome Curve

To further our analysis, we now introduce left and right quantile functions, allowing us to specify thresholds in terms of both selection rate and score cutoffs.

Definition B.2 (Upper quantile function)**.** Define Q to be the upper quantile function corresponding to $\boldsymbol{\pi}$, i.e.

$$\mathrm{Q}_{\mathsf{j}}(\beta) = \mathrm{argmax}\{c : \sum_{x=c}^{C} \boldsymbol{\pi}_{\mathsf{j}}(x) > \beta\} \quad \text{and} \quad \mathrm{Q}_{\mathsf{j}}^{+}(\beta) := \mathrm{argmax}\{c : \sum_{x=c}^{C} \boldsymbol{\pi}_{\mathsf{j}}(x) \geq \beta\}\,. \tag{7}$$

Crucially $\mathrm{Q}(\beta)$ is continuous from the right, and $\mathrm{Q}^{+}(\beta)$ is continuous from the left. Further, $\mathrm{Q}(\cdot)$ and $\mathrm{Q}^{+}(\cdot)$ allow us to compute derivatives of key functions, like the mapping from selection rate β to the group outcome associated with a policy of that rate, $\Delta\boldsymbol{\mu}(r_{\pi}^{-1}(\beta))$. Because we take $\boldsymbol{\pi}$ to have discrete support, all functions in this work are *piecewise linear*, so we shall need to distinguish between the left and right derivatives, defined as follows

$$\partial_{-} f(x) := \lim_{t \to 0^{-}} \frac{f(x+t) - f(x)}{t} \quad \text{and} \quad \partial_{+} f(y) := \lim_{t \to 0^{+}} \frac{f(y+t) - f(y)}{t}\,. \tag{8}$$

For f supported on $[a, b]$, we say that f is left- (resp. right-) differentiable if $\partial_{-} f(x)$ exists for all $x \in (a, b]$ (resp. $\partial_{+} f(y)$ exists for all $y \in [a, b)$). We now state the fundamental derivative computation which underpins the results to follow:

Lemma B.5. *Let $\mathbf{e}_x$ denote the vector such that $\mathbf{e}_x(x) = 1$, and $\mathbf{e}_x(x') = 0$ for $x' \neq x$. Then $\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) : [0,1] \to [0,1]^{C}$ is continuous, and has left and right derivatives*

$$\partial_{+}\left(\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)\right) = \boldsymbol{e}_{\mathrm{Q}(\beta)} \quad \textit{and} \quad \partial_{-}\left(\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)\right) = \boldsymbol{e}_{\mathrm{Q}^{+}(\beta)}\,. \tag{9}$$

The above lemma is proved in Appendix D.3. Moreover, Lemma B.5 implies that the outcome curve is concave under the assumption that $\boldsymbol{\Delta}(x)$ is monotone:

Proposition B.6. *Let $\boldsymbol{\pi}$ be a distribution over C states. Then $\beta \mapsto \Delta\boldsymbol{\mu}(r_{\boldsymbol{\pi}}^{-1}(\beta))$ is concave. In fact, if $\boldsymbol{w}(x)$ is any non-decreasing map from $\mathcal{X} \to \mathbb{R}$, $\beta \mapsto \langle \boldsymbol{w}, r_{\boldsymbol{\pi}}^{-1}(\beta)\rangle$ is concave.*

Proof. Recall that a univariate function f is concave (and finite) on $[a, b]$ if and only if (a) f is left- and right-differentiable, (b) for all $x \in (a, b)$, $\partial_{-} f(x) \geq \partial_{+} f(x)$ and (c) for any $x > y$, $\partial_{-} f(x) \leq \partial_{+} f(y)$.

Observe that $\Delta\boldsymbol{\mu}(r_{\boldsymbol{\pi}}^{-1}(\beta)) = \langle \boldsymbol{\Delta}, \boldsymbol{\pi} \circ r_{\boldsymbol{\pi}}^{-1}(\beta)\rangle$. By Lemma B.5, $\boldsymbol{\pi} \circ r_{\boldsymbol{\pi}}^{-1}(\beta)$ has right and left derivatives $\boldsymbol{e}_{\mathrm{Q}(\beta)}$ and $\boldsymbol{e}_{\mathrm{Q}^{+}(\beta)}$. Hence, we have that

$$\partial_{+}\Delta\boldsymbol{\mu}(\beta_{\mathsf{B}}) = \boldsymbol{\Delta}(\mathrm{Q}(\beta_{\mathsf{B}})) \quad \text{and} \quad \partial_{-}\Delta\boldsymbol{\mu}(\beta_{\mathsf{B}}) = \boldsymbol{\Delta}(\mathrm{Q}^{+}(\beta_{\mathsf{B}}))\,. \tag{10}$$

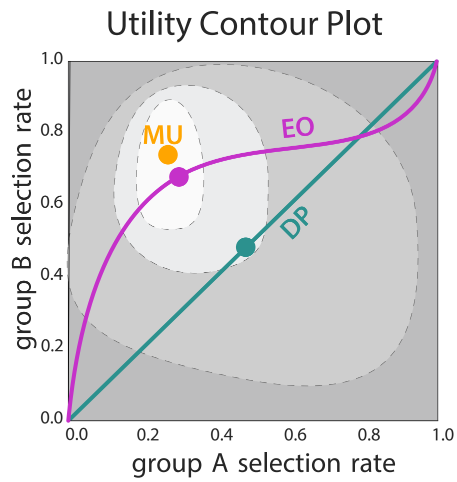

Figure 1: Considering the utility as a function of selection rates, fairness constraints correspond to restricting the optimization to one-dimensional curves. The DemParity (DP) constraint is a straight line with slope 1, while the EqOpt (EO) constraint is a curve given by the graph of $G^{(\mathsf{A}\to\mathsf{B})}$. The derivatives considered throughout Section C are taken with respect to the selection rate β_{A} (horizontal axis); projecting the EO and DP constraint curves to the horizontal axis recovers concave utility curves such as those shown in the lower panel of Figure 2 (where MaxUtil in is represented by a horizontal line through the MU optimal solution).

Using the fact that $\mathbf{\Delta}(x)$ is monotone, and that $\mathrm{Q} \leq \mathrm{Q}^+$, we see that $\partial_+ \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}})) \leq \partial_- \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$, and that $\partial \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$ and $\partial_+ \Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$ are non-increasing, from which it follows that $\Delta\boldsymbol{\mu}(f_{\boldsymbol{\pi}}^{-1}(\beta_{\mathsf{B}}))$ is concave. The general concavity result holds by replacing $\mathbf{\Delta}(x)$ with $\boldsymbol{w}(x)$. □

C Proofs of Characterization Theorems

We are now ready to present and prove theorems that characterize the selection rates under fairness constraints, namely DemParity and EqOpt. These characterizations are crucial for proving the results in Section 3. Our computations also generalize readily to other linear constraints, in a way that will become clear in Section C.2.

C.1 A Characterization Theorem for DemParity

In this section, we provide a theorem that gives an explicit characterization for the range of selection rates β_{A} for A when the bank loans according to DemParity. Observe that the DemParity objective corresponds to solving the following linear program:

$$\max_{\boldsymbol{\tau}=(\boldsymbol{\tau}_{\mathsf{A}},\boldsymbol{\tau}_{\mathsf{B}})\in[0,1]^{2C}} \mathcal{U}(\boldsymbol{\tau}) \quad \text{s.t.} \quad \langle \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}} \rangle = \langle \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{\tau}_{\mathsf{B}} \rangle \,.$$

Let us introduce the auxiliary variable $\beta := \langle \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}} \rangle = \langle \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{\tau}_{\mathsf{B}} \rangle$ corresponding to the selection rate which is held constant across groups, so that all feasible solutions lie on the green DP line in

Figure 1. We can then express the following equivalent linear program:

$$\max_{\boldsymbol{\tau}=(\boldsymbol{\tau}_{\mathsf{A}},\boldsymbol{\tau}_{\mathsf{B}})\in[0,1]^{2C},\beta\in[0,1]} \mathcal{U}(\boldsymbol{\tau}) \quad \text{s.t.} \quad \beta = \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{\tau}_{\mathsf{j}} \rangle,\ \mathsf{j} \in \{\mathsf{A},\mathsf{B}\}\ .$$

This is equivalent because, for a given β, Proposition B.3 says that the utility maximizing policies are of the form $\boldsymbol{\tau}_{\mathsf{j}} = r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)$. We now prove this:

Proof of Proposition B.3 for `DemParity`. Noting that $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{\tau}_{\mathsf{j}} \rangle$, we see that, by Lemma B.4, under the special case where $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = 1$, the optimal solution $(\boldsymbol{\tau}_{\mathsf{A}}^*(\beta), \boldsymbol{\tau}_{\mathsf{B}}^*(\beta))$ for fixed $r_{\boldsymbol{\pi}_{\mathsf{A}}}(\boldsymbol{\tau}_{\mathsf{A}}) = r_{\boldsymbol{\pi}_{\mathsf{B}}}(\boldsymbol{\tau}_{\mathsf{B}}) = \beta$ can be chosen to coincide with the threshold policies. Optimizing over β, the global optimal must coincide with thresholds.

□

Hence, any optimal policy is equivalent to the threshold policy $\boldsymbol{\tau} = (r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta))$, where β solves the following optimization:

$$\max_{\beta\in[0,1]} \mathcal{U}\left(\left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\right)\right)\ . \tag{11}$$

We shall show that the above expression is in fact a *concave* function in β, and hence the set of optimal selection rates can be characterized by first order conditions. This is presented formally in the following theorem:

Theorem C.1 (Selection rates for `DemParity`)**.** *The set of optimal selection rates* β^* *satisfying* (11) *forms a continuous interval* $[\beta^-_{\texttt{DemParity}}, \beta^+_{\texttt{DemParity}}]$*, such that for any* $\beta \in [0,1]$*, we have*

$$\begin{aligned} \beta < \beta^-_{\texttt{DemParity}} \quad &\text{if} \quad g_{\mathsf{A}}\boldsymbol{u}\left(\mathrm{Q}_{\mathsf{A}}(\beta)\right) + g_{\mathsf{B}}\boldsymbol{u}\left(\mathrm{Q}_{\mathsf{B}}(\beta)\right) > 0\ , \\ \beta > \beta^+_{\texttt{DemParity}} \quad &\text{if} \quad g_{\mathsf{A}}\boldsymbol{u}\left(\mathrm{Q}^+_{\mathsf{A}}(\beta)\right) + g_{\mathsf{B}}\boldsymbol{u}\left(\mathrm{Q}^+_{\mathsf{B}}(\beta)\right) < 0\ . \end{aligned}$$

Proof. Note that we can write

$$\mathcal{U}\left(\left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\right)\right) \quad = g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{A}} \circ r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta)\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{B}} \circ r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\rangle\ .$$

Since $\boldsymbol{u}(x)$ is non-decreasing in x, Proposition B.6 implies that $\beta \mapsto \mathcal{U}\left(\left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\right)\right)$ is concave in β. Hence, all optimal selection rates β^* lie in an interval $[\beta^-, \beta^+]$. To further characterize this interval, let us us compute left- and right-derivatives.

$$\begin{aligned} \partial_+ \mathcal{U}\left(\left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\right)\right) \quad &= \quad \partial_+ g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{A}} \circ r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta)\rangle + \partial_+ g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{\pi}_{\mathsf{B}} \circ r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\rangle \\ &= \quad g_{\mathsf{A}}\langle \boldsymbol{u}, \partial_+\left(\boldsymbol{\pi}_{\mathsf{A}} \circ r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta)\right)\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \partial_+\left(\boldsymbol{\pi}_{\mathsf{B}} \circ r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta)\right)\rangle \\ &\overset{\text{Lemma } B.5}{=} \quad g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{e}_{\mathrm{Q}_{\mathsf{A}}(\beta)}\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{e}_{\mathrm{Q}_{\mathsf{B}}(\beta)}\rangle \\ &= \quad g_{\mathsf{A}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) + g_{\mathsf{B}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))\ . \end{aligned}$$

The same argument shows that

$$\partial_- \mathcal{U}((r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta), r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(\beta))) = g_{\mathsf{A}}\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta)) + g_{\mathsf{B}}\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(\beta)).$$

By concavity of $\mathcal{U}\left((r^{-1}_{\boldsymbol{\pi}_\mathsf{A}}(\beta), r^{-1}_{\boldsymbol{\pi}_\mathsf{B}}(\beta))\right)$, a positive right derivative at β implies that $\beta < \beta^*$ for all β^* satisfying (11), and similarly, a negative left derivative at β implies that $\beta > \beta^*$ for all β^* satisfying (11).

□

The results in Section 3 pertaining to `DemParity` follow readily from Theorem C.1, and are proved in Section F. For example, we prove Theorem 3.3 by fixing a selection rate of interest (e.g. β_0) and inverting the inequalities in Theorem C.1 to find the exact population proportions under which, for example, `DemParity` results in a higher selection rate than β_0.

C.2 `EqOpt` and General Constraints

Next, we will provide a theorem that gives an explicit characterization for the range of selection rates β_A for A when the bank loans according to `EqOpt`. Observe that the `EqOpt` objective corresponds to solving the following linear program:

$$\max_{\boldsymbol{\tau}=(\boldsymbol{\tau}_\mathsf{A},\boldsymbol{\tau}_\mathsf{B})\in[0,1]^{2C}} \mathcal{U}(\boldsymbol{\tau}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_\mathsf{A} \circ \boldsymbol{\pi}_\mathsf{A}, \boldsymbol{\tau}_\mathsf{A}\rangle = \langle \boldsymbol{w}_\mathsf{B} \circ \boldsymbol{\pi}_\mathsf{B}, \boldsymbol{\tau}_\mathsf{B}\rangle \;, \tag{12}$$

where $\boldsymbol{w}_\mathsf{j} = \frac{\boldsymbol{\rho}}{\langle \boldsymbol{\rho}, \boldsymbol{\pi}_\mathsf{j}\rangle}$. This problem is similar to the demographic parity optimization in (11), except for the fact that the constraint includes the weights. Whereas we parameterized demographic parity solutions in terms of the acceptance rate β in equation (11), we will parameterize equation (12) in terms of the true positive rate (TPR), $t := \langle \boldsymbol{w}_\mathsf{A} \circ \boldsymbol{\pi}_\mathsf{A}, \boldsymbol{\tau}_\mathsf{A}\rangle$. Thus, (12) becomes

$$\max_{t\in[0,t_{\max}]} \max_{(\boldsymbol{\tau}_\mathsf{A},\boldsymbol{\tau}_\mathsf{B})\in[0,1]^{2C}} \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j} \mathcal{U}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_\mathsf{j} \circ \boldsymbol{\pi}_\mathsf{j}, \boldsymbol{\tau}_\mathsf{j}\rangle = t,\ \mathsf{j} \in \{\mathsf{A},\mathsf{B}\} \;, \tag{13}$$

where $t_{\max} = \min_{\mathsf{j}\in\{\mathsf{A},\mathsf{B}\}}\{\langle \boldsymbol{\pi}_\mathsf{j}, \boldsymbol{w}_\mathsf{j}\rangle\}$ is the largest possible TPR. The magenta EO curve in Figure 1 illustrates that feasible solutions to this optimization problem lie on a curve parametrized by t. Note that the objective function decouples for $\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}$ for the inner optimization problem,

$$\max_{\boldsymbol{\tau}_\mathsf{j}\in[0,1]^C} \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j} \mathcal{U}_\mathsf{j}(\boldsymbol{\tau}_\mathsf{j}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_\mathsf{j} \circ \boldsymbol{\pi}_\mathsf{j}, \boldsymbol{\tau}_\mathsf{j}\rangle = t \;. \tag{14}$$

We will now show that all optimal solutions for this inner optimization problem are $\boldsymbol{\pi}_\mathsf{j}$-a.e. equal to a policy in $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}_\mathsf{j})$, and thus can be written as $r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(\beta_\mathsf{j})$, depending only on the resulting selection rate.

Proof of Proposition B.3 for `EqOpt`. We apply Lemma B.4 to the inner optimization in (14) with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = \frac{\boldsymbol{\rho}(x)}{\langle \boldsymbol{\rho}, \boldsymbol{\pi}_\mathsf{j}\rangle}$. The claim follows from the assumption that $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing by optimizing over t. □

This selection rate β_j is uniquely determined by the TPR t (proof appears in Appendix E.1):

Lemma C.2. *Suppose that $\boldsymbol{w}(x) > 0$ for all x. Then the function*

$$T_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(\beta) := \langle r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(\beta), \boldsymbol{\pi}_\mathsf{j} \circ \boldsymbol{w}_\mathsf{j}\rangle$$

is a bijection from $[0,1]$ to $[0, \langle \boldsymbol{\pi}_\mathsf{j}, \boldsymbol{w}\rangle]$.

Hence, for any $t \in [0, t_{\max}]$, the mapping from TPR to acceptance rate, $T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)$, is well defined and any solution to (14) is $\boldsymbol{\pi}_\mathsf{j}$-a.e. equal to the policy $r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))$. Thus (13) reduces to

$$\max_{t\in[0,t_{\max}]} \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j}\mathcal{U}_\mathsf{j}\left(r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}\left(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t)\right)\right) . \tag{15}$$

The above expression parametrizes the optimization problem in terms of a single variable. We shall show that the above expression is in fact a *concave* function in t, and hence the set of optimal selection rates can be characterized by first order conditions. This is presented formally in the following theorem:

Theorem C.3 (Selection rates for `EqOpt`). *The set of optimal selection rates β^* for group A satsifying (13) forms a continuous interval $[\beta^-_{\mathtt{EqOpt}}, \beta^+_{\mathtt{EqOpt}}]$, such that for any $\beta \in [0,1]$, we have*

$$\beta < \beta^-_{\mathtt{EqOpt}} \quad \text{if} \quad g_\mathsf{A}\frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{A}(\beta))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}_\mathsf{A}(\beta))} + g_\mathsf{B}\frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta)))} > 0 ,$$

$$\beta > \beta^+_{\mathtt{EqOpt}} \quad \text{if} \quad g_\mathsf{A}\frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{A}(\beta))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}^+_\mathsf{A}(\beta))} + g_\mathsf{B}\frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}^+_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta)))} < 0 .$$

Here, $G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta) := T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(\beta))$ denotes the (well-defined) map from selection rates β_A for A to the selection rate β_B for B such that the policies $\boldsymbol{\tau}^*_\mathsf{A} := r^{-1}_{\boldsymbol{\pi}_\mathsf{A}}(\beta_\mathsf{A})$ and $\boldsymbol{\tau}^*_\mathsf{B} := r^{-1}_{\boldsymbol{\pi}_\mathsf{B}}(\beta_\mathsf{B})$ satisfy the constraint in (12).

Proof. Starting with the equivalent problem in (15), we use the concavity result of Lemma E.1. Because the objective function is the positive weighted sum of two concave functions, it is also concave. Hence, all optimal true positive rates t^* lie in an interval $[t^-, t^+]$. To further characterize $[t^-, t^+]$, we can compute left- and right-derivatives, again using the result of Lemma E.1.

$$\begin{aligned}\partial_+ \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j}\mathcal{U}_\mathsf{j}\left(r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))\right) &= g_\mathsf{A}\partial_+\mathcal{U}_\mathsf{A}\left(r^{-1}_{\boldsymbol{\pi}_\mathsf{A}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t))\right) + g_\mathsf{A}\partial_+\mathcal{U}_\mathsf{A}\left(r^{-1}_{\boldsymbol{\pi}_\mathsf{A}}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t))\right) \\ &= g_\mathsf{A}\frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{A}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t)))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}_\mathsf{A}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t)))} + g_\mathsf{B}\frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{B}(T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(t)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}_\mathsf{B}(T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(t)))}\end{aligned}$$

The same argument shows that

$$\partial_- \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j}\mathcal{U}_\mathsf{j}\left(r^{-1}_{\boldsymbol{\pi}_\mathsf{j}}(T^{-1}_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(t))\right) = g_\mathsf{A}\frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{A}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}^+_\mathsf{A}(T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t)))} + g_\mathsf{B}\frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{B}(T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(t)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}^+_\mathsf{B}(T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(t)))}.$$

By concavity, a positive right derivative at t implies that $t < t^*$ for all t^* satisfying (15), and similarly, a negative left derivative at t implies that $t > t^*$ for all t^* satisfying (15).

Finally, by Lemma C.2, this interval in t uniquely characterizes an interval of acceptance rates. Thus we translate directly into a statement about the selection rates β for group A by seeing that $T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t) = \beta$ and $T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(t) = G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta)$. □

Lastly, we remark that the results derived in this section go through verbatim for any linear constraint of the form $\langle \boldsymbol{w}, \boldsymbol{\pi}_{\mathsf{A}} \circ \boldsymbol{\tau}_{\mathsf{A}} \rangle = \langle \boldsymbol{w}, \boldsymbol{\pi}_{\mathsf{B}} \circ \boldsymbol{\tau}_{\mathsf{B}} \rangle$, as long as $\boldsymbol{u}(x)/\boldsymbol{w}(x)$ is increasing in x, and $\boldsymbol{w}(x) > 0$.

D Optimality of Threshold Policies

D.1 Proof of Lemma B.1

We begin with the first statement of the lemma. Suppose $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$. Then there exists a set $\mathcal{S} \subset \mathcal{X}$ such that $\boldsymbol{\pi}_{\mathsf{j}}(x) = 0$ for all $x \in \mathcal{S}$, and for all $x \notin \mathcal{S}$, $\boldsymbol{\tau}_{\mathsf{j}}(x) = \boldsymbol{\tau}'_{\mathsf{j}}(x)$. Thus,

$$\begin{aligned} r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}}) - r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}}) &= \sum_{x \in \mathcal{X}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) \\ &= \sum_{x \in \mathcal{S}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) = 0 \, . \end{aligned}$$

Conversely, suppose that $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{\mathsf{j}}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$. Let $\boldsymbol{\tau}_{\mathsf{j}} = \boldsymbol{\tau}_{c,\gamma}$ and $\boldsymbol{\tau}'_{\mathsf{j}} = \boldsymbol{\tau}_{c',\gamma'}$ as in Definition B.1. We now have the following cases:

1. Case 1: $c = c'$. Then $\boldsymbol{\tau}_{\mathsf{j}}(x) = \boldsymbol{\tau}'_{\mathsf{j}}(x)$ for all $x \in \mathcal{X} - \{c\}$. Hence,

   $$0 = r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}}) - r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}}) = \boldsymbol{\pi}(x)(\boldsymbol{\tau}_{\mathsf{j}}(c) - \boldsymbol{\tau}'_{\mathsf{j}}(c)) \, .$$

   This implies that either $\boldsymbol{\tau}_{\mathsf{j}}(c) = \boldsymbol{\tau}'_{\mathsf{j}}(c)$, and thus $\boldsymbol{\tau}_{\mathsf{j}}(x) = \boldsymbol{\tau}'_{\mathsf{j}}(x)$ for all $x \in \mathcal{X}$, or otherwise $\boldsymbol{\pi}(c) = 0$, in which case we still have $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$ (since the two policies agree every outside the set $\{c\}$).

2. Case 2: $c \neq c'$. We assume assume without loss of generality that $c' < c \leq C$. Since the policies $\boldsymbol{\tau}_{c',1}$ and $\boldsymbol{\tau}_{c'+1,0}$ are identity for $c' < C$, we may also assume without loss of generality that $\gamma' \in [0, 1)$. Thus for all $x \in \mathcal{S} := \{c', c'+1, \ldots, C\}$, we have $\boldsymbol{\tau}'_{\mathsf{j}}(x) < \boldsymbol{\tau}_{\mathsf{j}}(x)$. This implies that

   $$\begin{aligned} 0 &= r_{\boldsymbol{\pi}}(\boldsymbol{\tau}_{\mathsf{j}}) - r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}}) \\ &= \sum_{x \in \mathcal{S}} \boldsymbol{\pi}_{\mathsf{j}}(x)(\boldsymbol{\tau}_{\mathsf{j}}(x) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) \\ &\geq \min_{x \in \mathcal{S}}(\boldsymbol{\tau}_{\mathsf{j}}(c) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) \cdot \sum_{x \in \mathcal{S}} \boldsymbol{\pi}(x) \, . \end{aligned}$$

   Since $\min_{x \in \mathcal{S}}(\boldsymbol{\tau}_{\mathsf{j}}(c) - \boldsymbol{\tau}'_{\mathsf{j}}(x)) > 0$, it follows that $\sum_{x \in \mathcal{S}} \boldsymbol{\pi}_{\mathsf{j}}(x) = 0$, whence $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$.

Next, we show that $r_{\boldsymbol{\pi}}$ is a bijection from $\mathcal{T}_{\text{thresh}}(\boldsymbol{\pi}) \to [0, 1]$. That $r_{\boldsymbol{\pi}}$ is injective follows immediately from the fact if $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}) = r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}'_{\mathsf{j}})$, then $\boldsymbol{\tau}_{\mathsf{j}} \cong_{\boldsymbol{\pi}_{\mathsf{j}}} \boldsymbol{\tau}'_{\mathsf{j}}$. To show it is surjective, we exhibit for every $\beta \in [0, 1]$ a threshold policy $\boldsymbol{\tau}_{c,\gamma}$ for which $r_{\boldsymbol{\pi}_{\mathsf{j}}}(\boldsymbol{\tau}_{c,\gamma}) = \beta$. We may assume $\beta < 1$, since the all-ones policy has a selection rate of 1.

Recall the definition of the inverse CDF

$$\mathrm{Q}_{\mathsf{j}}(\beta) := \mathrm{argmax}\{c : \sum_{x=c}^{C} \boldsymbol{\pi}(x) > \beta\} \, .$$

Since $\beta < 1$, $\mathrm{Q_j}(\beta) \le C$. Let $\beta_+ = \sum_{x=\mathrm{Q_j}(\beta)}^{C} \boldsymbol{\pi}(x)$, and let $\beta_- = \sum_{x=\mathrm{Q_j}(\beta)+1}^{C} \boldsymbol{\pi}(x)$. Note that by definition, we have $\beta_- \le \beta < \beta_+$, and $\beta_+ - \beta_- = \boldsymbol{\pi}(\mathrm{Q_j}(\beta))$. Hence, if we define $\gamma = \frac{\beta-\beta_-}{\beta_+-\beta_-}$, we have

$$r_{\boldsymbol{\pi}_\mathrm{j}}(\boldsymbol{\tau}_{\mathrm{Q_j}(\beta),\gamma}) = \boldsymbol{\pi}(\mathrm{Q_j}(\beta))\gamma + \sum_{x=\mathrm{Q_j}(\beta)+1}^{C} \boldsymbol{\pi}(x) = \beta_- + (\beta_+ - \beta_-)\gamma = \beta_- + \beta - \beta_- = \beta \, .$$

D.2 Proof of Lemma B.4

Given $\boldsymbol{\tau} \in [0,1]^C$, we define the *normal cone* at $\boldsymbol{\tau}$ as $\mathrm{NC}(\boldsymbol{\tau}) := \mathrm{ConicalHull}\{\boldsymbol{z} : \boldsymbol{\tau} + \boldsymbol{z} \in [0,1]^C\}$. We can describe $\mathrm{NC}(\boldsymbol{\tau})$ explicitly as:

$$\mathrm{NC}(\boldsymbol{\tau}) := \{\boldsymbol{z} \in \mathbb{R}^C : \boldsymbol{z}_i \le 0 \text{ if } \boldsymbol{\tau}_i = 0, \boldsymbol{z}_i \ge 0 \text{ if } \boldsymbol{\tau}_i = 1\} \, .$$

Immediately from the above definition, we have the following useful identity, which is that for any vector $\boldsymbol{g} \in \mathbb{R}^C$,

$$\langle \boldsymbol{g}, \boldsymbol{z} \rangle \le 0 \; \forall \boldsymbol{z} \in \mathrm{NC}(\boldsymbol{\tau}), \quad \text{if and only if} \quad \forall x \in \mathcal{X}, \begin{cases} \boldsymbol{\tau}(x) = 0 & \boldsymbol{g}(x) < 0 \\ \boldsymbol{\tau}(x) = 1 & \boldsymbol{g}(x) > 0 \\ \boldsymbol{\tau}(x) \in [0,1] & \boldsymbol{g}(x) = 0 \end{cases} . \tag{16}$$

Now consider the optimization problem (6). By the first order KKT conditions, we know that for any optimizer $\boldsymbol{\tau}_*$ of the above objective, there exists some $\widehat{\lambda} \in \mathbb{R}$ such that, for all $\boldsymbol{z} \in \mathrm{NC}(\boldsymbol{\tau}_*)$

$$\langle \boldsymbol{z}, \boldsymbol{v} \circ \boldsymbol{\pi} + \widehat{\lambda} \boldsymbol{\pi} \circ \boldsymbol{w} \rangle \le 0 \, .$$

By (16), we must have that

$$\boldsymbol{\tau}_*(x) = \begin{cases} 0 & \boldsymbol{\pi}(x)(\boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x)) < 0 \\ 1 & \boldsymbol{\pi}(x)(\boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x)) > 0 \\ \in [0,1] & \boldsymbol{\pi}(x)(\boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x)) = 0 \end{cases} .$$

Now $\boldsymbol{\tau}_*(x)$ is not necessarily a threshold policy. To conclude the theorem, it suffices to exhibit a threshold policy $\widetilde{\boldsymbol{\tau}}_*$ such that $\boldsymbol{\tau}_*(x) \cong_{\boldsymbol{\pi}} \widetilde{\boldsymbol{\tau}}_*$. (Note that $\widetilde{\boldsymbol{\tau}}_*(x)$ will also be feasible for the constraint, and have the same objective value; hence $\widetilde{\boldsymbol{\tau}}_*$ will be optimal as well.)

Given $\boldsymbol{\tau}_*$ and $\widehat{\lambda}$, let $c_* = \min\{c \in \mathcal{X} : \boldsymbol{v}(x) + \widehat{\lambda}\boldsymbol{w}(x) \ge 0\}$. If either (a) $\boldsymbol{w}(x) = 0$ for all $x \in \mathcal{X}$ and $\boldsymbol{v}(x)$ is strictly increasing or (b) $\boldsymbol{v}(x)/\boldsymbol{w}(x)$ is strictly increasing, then the modified policy

$$\widetilde{\boldsymbol{\tau}}_*(x) = \begin{cases} 0 & x < c_* \\ \boldsymbol{\tau}_*(x) & x = c_* \\ 1 & x > c_* \end{cases} ,$$

is a threshold policy, and $\boldsymbol{\tau}_*(x) \cong_{\boldsymbol{\pi}} \widetilde{\boldsymbol{\tau}}_*$. Moreover, $\langle \boldsymbol{w}, \widetilde{\boldsymbol{\tau}}_* \rangle = \langle \boldsymbol{w}, \widetilde{\boldsymbol{\tau}}_* \rangle$ and $\langle \boldsymbol{\pi}, \widetilde{\boldsymbol{\tau}}_* \rangle = \langle \boldsymbol{\pi}, \widetilde{\boldsymbol{\tau}}_* \rangle$, which implies that $\widetilde{\boldsymbol{\tau}}_*$ is an optimal policy for the objective in Lemma B.4.

D.3 Proof of Lemma B.5

We shall prove

$$\partial_+ \left(\boldsymbol{\pi}_\mathsf{j} \circ r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta) \right) = \boldsymbol{e}_{\mathrm{Q}_\mathsf{j}(\beta)} \ , \tag{17}$$

where the derivative is with respect to β. The computation of the left-derivative is analogous. Since we are concerned with right-derivatives, we shall take $\beta \in [0,1)$. Since $\boldsymbol{\pi}_\mathsf{j} \circ r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta)$ does not depend on the choice of representative for $r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}$, we can choose a cannonical representation for $r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}$. In Section D.1, we saw that the threshold policy $\boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}$ had acceptance rate β, where we had defined

$$\beta_+ = \sum_{x=\mathrm{Q}_\mathsf{j}(\beta)}^{C} \boldsymbol{\pi}(x) \quad \text{and} \quad \beta_- = \sum_{x=\mathrm{Q}_\mathsf{j}(\beta)+1}^{C} \boldsymbol{\pi}(x) \ , \tag{18}$$

$$\gamma(\beta) = \frac{\beta - \beta_-}{\beta_+ - \beta_-} \ . \tag{19}$$

Note then that for each x, $\boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}(x)$ is piece-wise linear, and thus admits left and right derivatives. We first claim that

$$\forall x \in \mathcal{X} \setminus \{\mathrm{Q}_\mathsf{j}(\beta)\}, \quad \partial_+ \boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}(x) = 0 \ . \tag{20}$$

To see this, note that $\mathrm{Q}_\mathsf{j}(\beta)$ is right continuous, so for all ϵ sufficiently small, $\mathrm{Q}_\mathsf{j}(\beta+\epsilon) = \mathrm{Q}_\mathsf{j}(\beta)$. Hence, for all ϵ sufficiently small and all $x \neq \mathrm{Q}(\beta)$, we have $\boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta+\epsilon),\gamma(\beta+\epsilon)}(x) = \boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta+\epsilon),\gamma(\beta+\epsilon)}(x)$, as needed. Thus, Equation (20) implies that $\partial_+ \boldsymbol{\pi}_\mathsf{j} \circ r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta)$ is supported on $x = \mathrm{Q}_\mathsf{j}(\beta)$, and hence

$$\partial_+ \boldsymbol{\pi}_\mathsf{j} \circ r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta) = \partial_+ \boldsymbol{\pi}_\mathsf{j}(x) \boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}(x) \big|_{x=\mathrm{Q}_\mathsf{j}(\beta)} \cdot \boldsymbol{e}_{\mathrm{Q}_\mathsf{j}(\beta)} \ .$$

To conclude, we must show that $\partial_+ \boldsymbol{\pi}_\mathsf{j}(x) \boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}(x) \big|_{x=\mathrm{Q}_\mathsf{j}(\beta)} = 1$. To show this, we have

$$\begin{aligned} 1 &= \partial_+(\beta) \\ &= \partial_+ (r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)})) \quad \text{since} \qquad r_{\boldsymbol{\pi}_\mathsf{j}}(\boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}) = \beta \ \forall \beta \in [0,1) \\ &= \partial_+ \left(\sum_{x \in \mathcal{X}} \boldsymbol{\pi}(x) \cdot \boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}(x) \right) \\ &= \partial_+ \boldsymbol{\pi}(x) \cdot \boldsymbol{\tau}_{\mathrm{Q}_\mathsf{j}(\beta),\gamma(\beta)}(x) \big|_{x=\mathrm{Q}_\mathsf{j}(\beta)} \ , \text{ as needed.} \end{aligned}$$

E Technical Lemmas for Characterization of Fairness Solutions

E.1 Derivative Computation for `EqOpt`

In this section, we prove Lemma C.2, which we recall below.

Lemma C.2. *Suppose that $\boldsymbol{w}(x) > 0$ for all x. Then the function*

$$T_{\mathsf{j},\boldsymbol{w}_\mathsf{j}}(\beta) := \langle r_{\boldsymbol{\pi}_\mathsf{j}}^{-1}(\beta), \boldsymbol{\pi}_\mathsf{j} \circ \boldsymbol{w}_\mathsf{j} \rangle$$

is a bijection from $[0,1]$ *to* $[0, \langle \boldsymbol{\pi}_{\mathsf{j}}, \boldsymbol{w} \rangle]$.

We will prove Lemma C.2 in tandem with the following derivative computation which we applied in the proof of Theorem C.3.

Lemma E.1. *The function*

$$\mathcal{U}_{\mathsf{j}}(t; \boldsymbol{w}_{\mathsf{j}}) := \mathcal{U}_{\mathsf{j}}\left(r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}\left(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)\right)\right)$$

is concave in t *and has left and right derivatives*

$$\partial_+ \mathcal{U}_{\mathsf{j}}(t; \boldsymbol{w}_{\mathsf{j}}) = \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}_{\mathsf{j}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))} \quad \textit{and} \quad \partial_- \mathcal{U}_{\mathsf{j}}(t; \boldsymbol{w}_{\mathsf{j}}) = \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}^+(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}_{\mathsf{j}}^+(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))} .$$

Proof of Lemmas C.2 and E.1. Consider a $\beta \in [0,1]$. Then, $\boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta)$ is continuous and left and right differentiable by Lemma B.5, and its left and right derivatives are indicator vectors $\boldsymbol{e}_{\mathrm{Q}_{\mathsf{j}}(\beta)}$ and $\boldsymbol{e}_{\mathrm{Q}_{\mathsf{j}}^+(\beta)}$, respectively. Consequently, $\beta \mapsto \langle \boldsymbol{w}_{\mathsf{j}}, \boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) \rangle$ has left and right derivatives $\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}(\beta))$ and $\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}^+(\beta))$, respectively; both of which are both strictly positive by the assumption $\boldsymbol{w}(x) > 0$. Hence, $T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(\beta) = \langle \boldsymbol{w}_{\mathsf{j}}, \boldsymbol{\pi}_{\mathsf{j}} \circ r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(\beta) \rangle$ is strictly increasing in β, and so the map is injective. It is also surjective because $\beta = 0$ induces the policy $\boldsymbol{\tau}_{\mathsf{j}} = \mathbf{0}$ and $\beta = 1$ induces the policy $\boldsymbol{\tau}_{\mathsf{j}} = \mathbf{1}$ (up to $\boldsymbol{\pi}_{\mathsf{j}}$-measure zero). Hence, $T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(\beta)$ is an order preserving bijection with left- and right-derivatives, and we can compute the left and right derivatives of its inverse as follows:

$$\partial_+ T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t) = \frac{1}{\partial_+ T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(\beta)\big|_{\beta = T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)}} = \frac{1}{\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}_{\mathsf{j}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))} ,$$

and similarly, $\partial_- T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t) = \frac{1}{\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}^+(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))}$. Then we can compute that

$$\begin{aligned} \partial_+ \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t))) &= \partial_+ \mathcal{U}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta))\big|_{\beta = T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)} \cdot \partial_+ T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(\sup(t)) \\ &= \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}_{\mathsf{j}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))} . \end{aligned}$$

and similarly $\partial_- \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(t))) = \frac{\mathcal{U}(\mathrm{Q}_{\mathsf{j}}^+(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathsf{j}}(\mathrm{Q}_{\mathsf{j}}^+(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))}$. One can verify that for all $t_1 < t_2$, one has that $\partial_+ \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t_1))) \geq \partial_- \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t_2)))$, and that for all t, $\partial_+ \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t))) \leq \partial_- \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))$. These facts establish that the mapping $t \mapsto \mathcal{U}_{\mathsf{j}}(r_{\boldsymbol{\pi}_{\mathsf{j}}}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t)))$ is concave. □

E.2 Characterizations Under Soft Constraints

Given a convex penalty $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$, and $\lambda \in \mathbb{R}_{\geq 0}$, one can write down the general form for soft constrained utility optimization

$$\max_{\boldsymbol{\tau} = (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}})} \mathcal{U}(\boldsymbol{\tau}) - \lambda \Phi(\langle \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{A}} \rangle - \langle \boldsymbol{w}_{\mathsf{B}} \circ \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{\tau}_{\mathsf{B}} \rangle) , \tag{21}$$

where $\boldsymbol{w}_{\mathsf{A}}$ and $\boldsymbol{w}_{\mathsf{B}}$ represent generic constraints. Again, we shall assume that for $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$, $\boldsymbol{u}(x)/\boldsymbol{w}_{\mathsf{j}}(x)$ is non-decreasing. Recall that for $\boldsymbol{w}_{\mathsf{j}} = (1, 1, \ldots, 1)$, one recovers the soft version of

DemParity, whereas for $\boldsymbol{w}_{\mathsf{j}} = \frac{\boldsymbol{\rho}}{\langle \boldsymbol{\rho}, \boldsymbol{\pi}_{\mathsf{j}} \rangle}$, one recovers the soft constrained version of EqOpt.

The same argument presented in Section C.2 shows that the optimal policies are of the form

$$\boldsymbol{\tau}_{\mathsf{j}} = r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1}(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t_{\mathsf{j}})) \ ,$$

where $(t_{\mathsf{A}}, t_{\mathsf{B}})$ are solutions to the following optimization problem:

$$\max_{t_{\mathsf{A}} \in [0, \langle \boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{w}_{\mathsf{A}} \rangle], t_{\mathsf{B}} \in [0, \langle \boldsymbol{\pi}_{\mathsf{B}}, \boldsymbol{w}_{\mathsf{B}} \rangle]} \ g_{\mathsf{A}} \mathcal{U}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}}))) + g_{\mathsf{B}} \mathcal{U}_{\mathsf{B}}(r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}}))) - \lambda \Phi(t_{\mathsf{A}} - t_{\mathsf{B}}) \ .$$

The following lemma gives us a first order characterization of these optimal TPRs, $(t_{\mathsf{A}}, t_{\mathsf{B}})$.

Lemma E.2. *All optimal policies are equivalent to threshold policies with selection rate* $(\beta_{\mathsf{A}}, \beta_{\mathsf{B}})$ *which satisfy*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \left[\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{A}}))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{B}}))} - \lambda \partial_+ \Phi(\Delta), \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^+(\beta_{\mathsf{A}}))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}^+(\beta_{\mathsf{A}}))} - \lambda \partial_- \Phi(\Delta) \right] \\ \left[\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta_{\mathsf{B}}))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}(\beta_{\mathsf{B}}))} + \lambda \partial_- \Phi(\Delta), \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^+(\beta))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}^+(\beta_{\mathsf{B}}))} + \lambda \partial_+ \Phi(\Delta) \right] \end{bmatrix} \ , \tag{22}$$

where $\Delta = t_{\mathsf{A}} - t_{\mathsf{B}} = T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(\beta_{\mathsf{A}}) - T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(\beta_{\mathsf{B}})$.

Proof. Let $\partial(\cdot)$ denote the super-gradient set of a concave function. Note that if F is left-and-right differentiable and concave, then $\partial F(x) = [\partial_+ F(x), \partial_- F(x)]$. By concavity of $\mathcal{U}_{\mathsf{j}}$ and convexity of Φ, we must have that

$$\begin{aligned}
\begin{bmatrix} 0 \\ 0 \end{bmatrix} &\in \partial \sum_{\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}} \mathcal{U}_{\mathsf{j}} \left(r_{\boldsymbol{\pi}_{\mathsf{j}}}^{-1} \left(T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}^{-1}(t_{\mathsf{j}}) \right) \right) - \lambda \Phi(t_{\mathsf{A}} - t_{\mathsf{B}}) \\
&= \begin{bmatrix} \partial \mathcal{U}_{\mathsf{A}} \left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})) \right) + \partial_{t_{\mathsf{A}}} \{ -\lambda \Phi(t_{\mathsf{A}} - t_{\mathsf{B}}) \} \\ \partial \mathcal{U}_{\mathsf{A}} \left(r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})) \right) + \partial_{t_{\mathsf{B}}} \{ -\lambda \Phi(t_{\mathsf{A}} - t_{\mathsf{B}}) \} \end{bmatrix} \\
&= \begin{bmatrix} \partial \mathcal{U}_{\mathsf{A}} \left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})) \right) - \lambda \partial \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}} \\ \partial \mathcal{U}_{\mathsf{B}} \left(r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})) \right) + \lambda \partial \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}} \end{bmatrix} \\
&= \begin{bmatrix} [\partial_+ \mathcal{U}_{\mathsf{A}} \left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})) \right) - \lambda \partial_+ \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}, \partial_- \mathcal{U}_{\mathsf{A}} \left(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})) \right) - \lambda \partial_- \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}] \\ [\partial_+ \mathcal{U}_{\mathsf{B}} \left(r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})) \right) + \lambda \partial_- \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}, \partial_- \mathcal{U}_{\mathsf{B}} \left(r_{\boldsymbol{\pi}_{\mathsf{B}}}^{-1}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{B}})) \right) + \lambda \partial_+ \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}] \end{bmatrix} \\
&= \begin{bmatrix} \left[\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})))} - \lambda \partial_+ \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}, \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^+(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}^+(T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}^{-1}(t_{\mathsf{A}})))} - \lambda \partial_- \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}} \right] \\ \left[\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})))} + \lambda \partial_- \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}, \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^+(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}^+(T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}^{-1}(t_{\mathsf{B}})))} + \lambda \partial_+ \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}} \right] \end{bmatrix} \\
&= \begin{bmatrix} \left[\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{A}}))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{A}}))} - \lambda \partial_+ \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}, \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^+(\beta_{\mathsf{A}}))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}^+(\beta_{\mathsf{B}}))} - \lambda \partial_- \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}} \right] \\ \left[\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}(\beta_{\mathsf{B}}))} + \lambda \partial_- \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}}, \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^+(\beta_{\mathsf{B}}))}{\boldsymbol{w}_{\mathsf{B}}(\mathrm{Q}_{\mathsf{B}}^+(\beta_{\mathsf{B}}))} + \lambda \partial_+ \Phi(t) \big|_{t = t_{\mathsf{A}} - t_{\mathsf{B}}} \right] \end{bmatrix} \ .
\end{aligned}$$

Substituting $\Delta = t_{\mathsf{A}} - t_{\mathsf{B}} = T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(\beta_{\mathsf{A}}) - T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(\beta_{\mathsf{B}})$ concludes the proof. □

In general, a closed form solution for the soft constrained problem may be difficult to state. However, for the case of $\Phi(t) = |t|$, we can state an explicit closed form solution:

Proposition E.3 (Special case of $\Phi(t) = |t|$). *Let $\Phi(t) = |t|$, fix λ, and let $[\beta_{\mathsf{A}}^{\lambda,-}, \beta_{\mathsf{A}}^{\lambda,+}]$ denote the interval of optimal selection rates for Equation (21) with regularization λ. Finally, suppose that for any optimal* MaxUtil *selection rates $(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}, \beta_{\mathsf{B}}^{\texttt{MaxUtil}})$, one has $T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}) < T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(\beta_{\mathsf{B}}^{\texttt{MaxUtil}})$. Let $[\beta_{\mathsf{A}}^{-}, \beta_{\mathsf{A}}^{+}]$ denote the optimal loan rates in (21). Then there exists a λ_* such that, for $\lambda \geq \lambda_*$, $[\beta_{\mathsf{A}}^{-}, \beta_{\mathsf{A}}^{+}]$ coincides with the hard constrained solution. Moreover, for $\lambda < \lambda_*$, any $\beta \in [0,1]$ satifies*

$$\begin{aligned} \beta < \beta_{\mathsf{A}}^{\lambda,-} \quad &\textit{if} \quad g_{\mathsf{A}} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(\beta))} + \sigma_* \lambda > 0 \\ \beta > \beta_{\mathsf{A}}^{\lambda,+} \quad &\textit{if} \quad g_{\mathsf{A}} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta))} + \sigma_* \lambda < 0 \ . \end{aligned}$$

Proof. Given a set of optimal constraint values $(t_{\mathsf{A}}, t_{\mathsf{B}}) = (T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(\beta_{\mathsf{A}}), T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(\beta_{\mathsf{B}}))$ for optimal selection rates $(\beta_{\mathsf{A}}, \beta_{\mathsf{B}})$ for a given parameter λ. By Proposition E.4 below, it follows that if $t_{\mathsf{A}} = t_{\mathsf{B}}$ for all optimal solutions, then for all $\lambda' \geq \lambda$, all optimal solutions must also have $t_{\mathsf{A}} = t_{\mathsf{B}}$.

Hence, it suffices to show that (a) there exists a finite λ such that all solutions must have $t_{\mathsf{A}} = t_{\mathsf{B}}$, and (b) if $t_{\mathsf{A}} \neq t_{\mathsf{B}}$, then the display in (E.3) holds.

To prove (a) and (b), suppose $t_{\mathsf{A}} \neq t_{\mathsf{B}}$. By Proposition E.4 below and the fact that $T_{\mathsf{A},\boldsymbol{w}_{\mathsf{A}}}(\beta^{\texttt{MaxUtil}}) < T_{\mathsf{B},\boldsymbol{w}_{\mathsf{B}}}(\beta_{\mathsf{B}}^{\texttt{MaxUtil}})$, we have $t_{\mathsf{A}} < t_{\mathsf{B}}$. Moreover we can compute that

$$\partial \Phi(t) = \begin{cases} \{1\} & t > 0 \\ [-1, 1] & t = 0 \\ \{-1\} & t < 0 \end{cases}$$

it follows from the first order condition in Lemma E.2 that, if $t_{\mathsf{A}} \neq t_{\mathsf{B}}$

$$0 \in [\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta_{\mathsf{A}}))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta_{\mathsf{A}}))} + \lambda, \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{A}}))}{\boldsymbol{w}_{\mathsf{A}}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{B}}))} + \lambda] \ , \tag{23}$$

which immediately implies point (b). Point (a) follows from the above display by noting that, since $\boldsymbol{w}_{\mathsf{j}}(x) > 0$ and $\boldsymbol{u}(x) < \infty$ for all x, where exists a λ sufficiently large such that (23) cannot hold for any β_{A}. □

E.3 Qualitative Behavior of Soft Constraints

We now present a proposition which formalizes the intuition that soft constraints interpolate between MaxUtil and the general hard constraint (12) in Section C.2 (for arbitrary $\boldsymbol{w}$, not just for EqOpt). Because optimal policies may not be unique, we define the solution sets

$$\mathbf{P}(\lambda) \ := \ \{(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) : (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \text{ solves (21) with parameter } \lambda\} \ ,$$

with the set $\mathbf{P}(\infty)$ denoting the set of solutions to (12).

At a high level, we parameterize the soft constrained solution in terms of the value of the constraint $t_{\mathsf{A}} = \langle \boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}} \rangle$ for A and the difference in constraint values $\Delta = \langle \boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}} \rangle - \langle \boldsymbol{\tau}_{\mathsf{B}}, \boldsymbol{w}_{\mathsf{B}} \circ \boldsymbol{\pi}_{\mathsf{B}} \rangle$, where $(\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in \mathbf{P}(\lambda)$. We show that t_{A} interpolates between the value of the constraint on A at $\lambda = 0$ and at $\lambda = \infty$, and that Δ interpolates between the difference at $\lambda = 0$ (MaxUtil) and at $\Delta = 0$ at $\lambda = \infty$. To be rigorous, we note that the possible values for t_{A} and

Δ for each λ are actually contiguous intervals. Hence, to make the interpolation precise, we define the following partial order on such intervals:

Definition E.1 (Interval order)**.** Let $\mathcal{S}_1, \mathcal{S}_2$ be two intervals. We say that $\mathcal{S}_1 \prec \mathcal{S}_2$ if $\max \{x \in \mathcal{S}_1 \} < \min \{x \in \mathcal{S}_2\}$ and $\mathcal{S}_1 \preceq \mathcal{S}_2$ if both $\max \{x \in \mathcal{S}_1\} \leq \max \{x \in \mathcal{S}_2\}$ and $\min\{x \in \mathcal{S}_1\} \leq \min \{x \in \mathcal{S}_2\}$. We say that an interval-valued function $\mathcal{S}(\lambda)$ is *non-decreasing* (resp. *non increasing*) in λ if $\mathcal{S}(\lambda) \preceq \mathcal{S}(\lambda')$ (resp $\mathcal{S}(\lambda') \preceq \mathcal{S}(\lambda')$ for $\lambda \leq \lambda'$).

In these terms, the interpolation of the soft constraints can be stated as follows:

Proposition E.4 (Soft constraints interpolate between `MaxUtil` and hard constrained solution)**.** *Let $\Phi(t)$ be a convex, symmetric convex function with $\Phi(t) > 0$ for $t > 0$. Then the sets*

$$\begin{aligned} \mathcal{D}(\lambda) &:= \{\Delta := \langle \boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}} \rangle - \langle \boldsymbol{\tau}_{\mathsf{B}}, \boldsymbol{w}_{\mathsf{B}} \circ \boldsymbol{\pi}_{\mathsf{B}} \rangle : (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in \mathbf{P}(\lambda)\} \\ \mathcal{T}_{\mathsf{A}}(\lambda) &:= \{t_{\mathsf{A}} := \langle \boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{w}_{\mathsf{A}} \circ \boldsymbol{\pi}_{\mathsf{A}} \rangle | \exists \boldsymbol{\tau}_{\mathsf{B}} : (\boldsymbol{\tau}_{\mathsf{A}}, \boldsymbol{\tau}_{\mathsf{B}}) \in \mathbf{P}(\lambda)\} \end{aligned}$$

are closed intervals. Moreover,

1. *In all cases,* $\lim_{\lambda \to \infty} \max\{|\Delta| \in \mathcal{D}(\lambda)\} = 0$.
2. *If $0 \in \mathcal{D}(\lambda)$, then there exists a* `MaxUtil` *solution satisfying* (12)*. Thus, for all $\lambda > 0$, $\mathbf{P}(\lambda) = \mathbf{P}(\infty)$.*
3. *If $\mathcal{D}(\lambda) \prec \{0\}$, then $\mathcal{D}(\lambda)$ and $\mathcal{T}_{\mathsf{A}}(\lambda)$ are non-decreasing on $\lambda \in (0, \infty]$, and vice versa if $\mathcal{D}(\lambda) \succ \{0\}$.*
4. *If $\mathcal{D}(\lambda) \prec \{0\}$, then $\{0\} = \mathcal{D}(\infty) \succeq \mathcal{D}(\lambda) \succeq \{\min : \Delta \in \mathcal{D}(0)\}$, and $\mathcal{T}_{\mathsf{A}}(\infty) \succeq \mathcal{T}_{\mathsf{A}}(\lambda) \succeq \{\min : \Delta \in \mathcal{T}_{\mathsf{A}}(\lambda)\}$, and vice versa if $\mathcal{D}(\lambda) \succ \{0\}$.*

E.3.1 Proof of Proposition E.4

Again, we parameterize all solutions to the soft-constrained problem as in correspondence with solutions $(t_{\mathsf{A}}, t_{\mathsf{B}})$ to

$$\min_{(t_{\mathsf{A}}, t_{\mathsf{B}})} g_{\mathsf{A}} \mathcal{U}_{\mathsf{A}}(t_{\mathsf{A}}; \boldsymbol{w}_{\mathsf{A}}) + g_{\mathsf{B}} \mathcal{U}_{\mathsf{B}}(t_{\mathsf{B}}; \boldsymbol{w}_{\mathsf{B}}) + \lambda \Phi(t_{\mathsf{A}} - t_{\mathsf{B}})\,.$$

Letting $\Delta := t_{\mathsf{B}} - t_{\mathsf{A}}$, we can reparameterize the above as

$$\min_{(t_{\mathsf{A}}, \Delta)} g_{\mathsf{A}} \mathcal{U}_{\mathsf{A}}(t_{\mathsf{A}}; \boldsymbol{w}_{\mathsf{A}}) + g_{\mathsf{B}} \mathcal{U}_{\mathsf{B}}(t_{\mathsf{A}} + \Delta; \boldsymbol{w}_{\mathsf{B}}) - \lambda \Phi(\Delta)\,.$$

Note then that $\mathcal{D}(\lambda)$ denotes the set of Δ which are partial maximimizers of the above display. If $0 \in \{\mathcal{D}(\lambda)\}$, this implies that there exists a `MaxUtil` solution for which $\Delta = 0$, therefore, for all $\lambda > 0$, all solutions will be `MaxUtil` solutions for which $\mathcal{D}(\lambda) = 0$. Otherwise assume without loss of generality that $\mathcal{D}(\lambda) < \{0\}$.

First, the statement $\{0\} = \mathcal{D}(\infty) \succeq \mathcal{D}(\lambda) \succeq \{\min : \Delta \in \mathcal{D}(0)\}$, and $\mathcal{T}_{\mathsf{A}}(\infty) \succeq \mathcal{T}_{\mathsf{A}}(\lambda) \succeq \{\min : \Delta \in \mathcal{T}_{\mathsf{A}}(\lambda)\}$, and vice versa if $\mathcal{D}(\lambda) \succ \{0\}$ can be solved by on a case-by-case basis. The strategy is to show that if any of these inequalities are violated, then the associated values of Δ and t_{A} are not partial maximizers of the soft constraint objective. In particular, $\mathcal{T}_{\mathsf{A}}(\lambda) \subset [T_-, T_+]$ for some appropriate T_-, T_+.

We now show that $\mathcal{D}(\lambda)$ and $\mathcal{T}_{\mathsf{A}}(\lambda)$ are non-increasing and non-decreasing, respectively. We shall do so invoking the following technical lemma.

Lemma E.5. *Let $G_1(t)$ be concave and let $G_2(t;\lambda)$ be concave in t. Let $\partial G_2(t;\lambda)$ denote the super-gradient of G_2, that is*

$$\partial G_2(t;\lambda) := \mathrm{Conv}(\{\partial_- G_2(t;\lambda)\} \cup \{\partial_- G_2(t;\lambda)\})$$

denotes the super-gradient set of the concave mapping $t \mapsto \partial G_2(t;\lambda)$.

Then if $\lambda \mapsto \partial G_2(t;\lambda)$ is non-increasing (resp. non-decreasing) in λ, the interval valued function defined below is non-increasing (resp. non-decreasing) in λ

$$\mathrm{MAX}(\lambda) := \lambda \mapsto \arg\max_{t\in[a,b]} G_1(t) + G_2(t;\lambda)\ .$$

For $\mathcal{D}(\lambda)$, one can write any partial maximizer Δ as

$$\max_{\Delta\geq 0} G_1(\Delta) + G_2(\Delta;\lambda)$$

with $G_1(\Delta) = \max_{t_{\mathsf{A}}} g_{\mathsf{A}}\mathcal{U}_{\mathsf{A}}(t_{\mathsf{A}};\boldsymbol{w}_{\mathsf{A}}) + g_{\mathsf{B}}\mathcal{U}_{\mathsf{B}}(t_{\mathsf{A}}+\Delta;\boldsymbol{w}_{\mathsf{B}})$ and $G_2(\Delta;\lambda) = \lambda\Phi(\Delta)$. Note that $G_1(\Delta)$ is concave, being the partial maximization of a concave function, and $\partial G_2(\Delta;\lambda) = -t\partial\Phi(\Delta)$. Since $\partial\Phi(\Delta) \succeq \{0\}$ for $\Delta \geq 0$ (by convexity of ϕ) , we have that $\partial G_2(\Delta;\lambda) = -t\partial\Phi(\Delta)$ is non-increasing in λ. Hence Lemma E.5 implies that interval valued function $\mathcal{D}(\lambda)$ is non-increasing.

To show that $\mathcal{T}_{\mathsf{A}}(\lambda)$ is non-decreasing, we have that any maximizer t_{A} can be written as

$$\max_{t_{\mathsf{A}}\in[T_-,T_+]} G_1(t_{\mathsf{A}}) + G_2(t_{\mathsf{A}};\lambda)$$

where $G_1(t_{\mathsf{A}}) = g_{\mathsf{A}}\mathcal{U}_{\mathsf{A}}(t_{\mathsf{A}};\boldsymbol{w}_{\mathsf{A}})$ and $G_2(t_{\mathsf{A}};\lambda) = \max_{\Delta\geq 0} g_{\mathsf{B}}\mathcal{U}_{\mathsf{B}}(t_{\mathsf{A}}+\Delta;\boldsymbol{w}_{\mathsf{B}}) + \lambda\Phi(\Delta)$. By Danskin's theorem,

$$\partial G_2(t_{\mathsf{A}};\lambda) = \{\partial\mathcal{U}_{\mathsf{B}}(t_{\mathsf{A}}+\Delta;\boldsymbol{w}_{\mathsf{B}}) : \Delta \in \arg\max G_2(t_{\mathsf{A}};\lambda)\}\ .$$

Note that $\{\Delta \in \arg\max G_2(t_{\mathsf{A}};\lambda)\}$ is non-increasing in λ for a fixed t_{A}, since the contribution of the regularizer increases. Since the sets $\partial\mathcal{U}_{\mathsf{B}}(t_{\mathsf{A}}+\Delta;\boldsymbol{w}_{\mathsf{B}})$ are themselves non-increasing in Δ by concavity, we conclude that $\partial G_2(t_{\mathsf{A}};\lambda)$ is non-decreasing in λ. Hence, Lemma E.5 implies that $\mathcal{T}_{\mathsf{A}}(\lambda)$ is non-decreasing in λ.

Finally, to show that $\max\{|\Delta| : \Delta \in \mathcal{D}(\lambda)|\} \to 0$, Note that the left and right derivatives of $g_{\mathsf{A}}\mathcal{U}_{\mathsf{A}}(t;\boldsymbol{w}_{\mathsf{A}})$ and $g_{\mathsf{B}}\mathcal{U}_{\mathsf{B}}(t;\boldsymbol{w}_{\mathsf{B}})$ are upper bounded by M whereas, since Φ is strictly convex, we know that for every $\epsilon > 0$, $\min\{|\partial_+\Phi(\Delta)|, |\partial_-\Phi(\Delta)|\} > m(\epsilon)$ for all $\Delta : |\Delta| > \epsilon$. Hence, the first order optimality conditions cannot be satisfied for $|\Delta| > \epsilon$, and $\lambda > \frac{M}{m(\epsilon)}$, so as $\lambda \to \infty$, $|\Delta| \to 0$.

Proof of Lemma E.5. We prove the case where $\partial G_2(t;\lambda)$ is non-increasing. The first order conditions requires that at an optimal t, one has

$$\partial_- G_1(t) + \partial G_2(t;\lambda)_- \geq 0 \geq \partial_+ G_1(t) + \partial G_2(t;\lambda)_+$$

where the super-gradients are amended to take into account boundary conditions. Suppose that for the sake of contradiction that for $\lambda' > \lambda$, $\mathrm{MAX}(\lambda') \preceq \mathrm{MAX}(\lambda)$ fails. Then, there (a) exists a

$t \in \text{MAX}(\lambda)$ such that $\{t\} \prec \text{MAX}(\lambda')$, or (b) $t \in \text{MAX}(\lambda')$ such that $\{t\} \succ \text{MAX}(\lambda')$. Note that if $\{t\} \prec \text{MAX}(\lambda')$, it must be the case that

$$\partial_+ G_1(t) + \partial G_2(t;\lambda')_+ > 0\,.$$

By assumption, $\partial_- G_2(t;\lambda')_+ \le \partial_0 G_2(t;\lambda)_+$, which implies

$$\partial_+ G_1(t) + \partial G_2(t;\lambda')_+ \le \partial_+ G_1(t) + \partial_+ G_2(t;\lambda)_- 0 \le 0\,,$$

a contradiction. ☐

F Proofs of Main Results

We remark that the proofs in this section rely crucially on the characterizations of the optimal fairness-constrained policies developed in Section C. We first define the notion of CDF domination, which is referred to in a few of the proofs. Intuitively, it means that for any score, the fraction of group B above this is higher than that for group A. It is realistic to assume this if we keep with our convention that group A is the disadvantaged group relative to group B.

Definition F.1 (CDF domination)**.** $\boldsymbol{\pi}_{\mathsf{A}}$ is said to be *dominated by* $\boldsymbol{\pi}_{\mathsf{B}}$ if $\forall a \ge 1, \sum_{x>a} \boldsymbol{\pi}_{\mathsf{A}} < \sum_{x>a} \boldsymbol{\pi}_{\mathsf{B}}$. We denote this as $\boldsymbol{\pi}_{\mathsf{A}} \prec \boldsymbol{\pi}_{\mathsf{B}}$.

We remark that the $\prec$ notation in this section is entirely unrelated to the the partial order on intervals from Section E.3. Frequently, we shall use the following lemma:

Lemma F.1. *Suppose that $\boldsymbol{\pi}_{\mathsf{A}} \prec \boldsymbol{\pi}_{\mathsf{B}}$. Then, for all $\beta > 0$, it holds that $\mathrm{Q}_{\mathsf{A}}(\beta) \le \mathrm{Q}_{\mathsf{B}}(\beta)$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) \le \boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))$*

Proof. The fact that $\mathrm{Q}_{\mathsf{A}}(\beta) \le \mathrm{Q}_{\mathsf{B}}(\beta)$ follows directly from the definition of monotonicty of $\boldsymbol{u}$ implies that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) \le \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))$. ☐

F.1 Proof of Proposition 3.1

The `MaxUtil` policy for group j solves the optimization

$$\max_{\boldsymbol{\tau}_{\mathsf{j}} \in [0,1]^C} \mathcal{U}_{\mathsf{j}}(\boldsymbol{\tau}_{\mathsf{j}}) = \max_{\beta_{\mathsf{j}} \in [0,1]} \mathcal{U}_{\mathsf{j}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta_{\mathsf{j}}))\,.$$

Computing left and right derivatives of this objective yields

$$\partial_+ \mathcal{U}_{\mathsf{j}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta_{\mathsf{j}})) = \boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(\beta)), \qquad \partial_- \mathcal{U}_{\mathsf{j}}(r^{-1}_{\boldsymbol{\pi}_{\mathsf{j}}}(\beta_{\mathsf{j}})) = \boldsymbol{u}(\mathrm{Q}^+_{\mathsf{j}}(\beta))\,.$$

By concavity, solutions β^* satisfy

$$\begin{aligned} \beta < \beta^* \quad &\text{if} \quad \boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(\beta)) > 0\,, \\ \beta > \beta^* \quad &\text{if} \quad \boldsymbol{u}(\mathrm{Q}^+_{\mathsf{j}}(\beta)) < 0\,. \end{aligned} \tag{24}$$

Therefore, we conclude that the `MaxUtil` policy loans only to scores x s.t. $\boldsymbol{u}(x) > 0$, which implies $\boldsymbol{\Delta}(x) > 0$ for all scores loaned to. Therefore we must have that $0 \le \Delta\boldsymbol{\mu}^{\texttt{MaxUtil}}$. By definition $\Delta\boldsymbol{\mu}^{\texttt{MaxUtil}} \le \Delta\boldsymbol{\mu}^*$.

F.2 Proof of Theorem 3.2

We begin with proving part (a), which gives conditions under which DemParity cases relative improvement. Recall that $\overline{\beta}$ is the largest selection rate for which $\mathcal{U}(\overline{\beta}) = \mathcal{U}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})$. First, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{DemParity}}$ from below. Fix an acceptance rate β such that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta < \min\{\beta_{\mathsf{B}}^{\texttt{MaxUtil}}, \overline{\beta}\}$. By Theorem C.1, we have that DemParity selects to group A with rate higher than β as long as

$$g_{\mathsf{A}} \leq g_1 := \frac{1}{1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}} .$$

By (24) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) > 0$, so $0 < g_1 < 1$.

Next, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{DemParity}}$ from above. First, consider the case that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < \overline{\beta}$, and fix β' such that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < \beta' < \overline{\beta}$. Then DemParity selects group A at a rate $\beta_{\mathsf{A}} < \beta'$ for any proportion g_{A}. This follows from applying Theorem C.1 since we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta')) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(\beta')) < 0$ by (24) and the monotonicity of $\boldsymbol{u}$.

Instead, in the case that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \overline{\beta}$, fix β' such that $\overline{\beta} < \beta' < \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$. Then DemParity selects group A at a rate less than β' as long as

$$g_{\mathsf{A}} \geq g_0 := \frac{1}{1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta'))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(\beta'))}} .$$

By (24) and the monotonicity of $\boldsymbol{u}$, $0 < g_0 < g_1$. Thus for $g_{\mathsf{A}} \in [g_0, g_1]$, the DemParity selection rate for group A is bounded between β and β', and thus DemParity results in relative improvement.

Next, we prove part (b), which gives conditions under which EqOpt cases relative improvement. First, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{EqOpt}}$ from below. Fix an acceptance rate β such that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$. By Theorem C.3, EqOpt selects group A at a rate higher than β as long as

$$g_{\mathsf{A}} > g_3 := \frac{1}{1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}} .$$

By (24) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) > 0$, so $g_3 > 0$.

Next, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{EqOpt}}$ from above. First, consider the case that there exists β' such that $\beta' < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < G^{(\mathsf{A}\to\mathsf{B})}(\beta')$. Then EqOpt selects group A at a rate less than this β' for any g_{A}. This follows from Theorem C.3 since we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta')) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta'))) < 0$ by (24) and the monotonicity of $\boldsymbol{u}$.

In the other case, fix β' such that $\beta < \beta' < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta')$. By Theorem C.3, EqOpt selects group A at a rate lower than β' as long as

$$g_{\mathsf{A}} > g_2 := \frac{1}{1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta')))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta')))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta'))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta'))}} .$$

By (24) and the monotonicity of $\boldsymbol{u}$, $0 < g_2 < g_3$. Thus for $g_{\mathsf{A}} \in [g_2, g_3]$, the EqOpt selection rate for group A is bounded between β and β', and thus EqOpt results in relative improvement.

F.3 Proof of Theorem 3.3

Recall our assumption that $\beta > \beta_{\mathsf{A}}^{\mathtt{MaxUtil}}$ and $\beta_{\mathsf{B}}^{\mathtt{MaxUtil}} > \beta$. As argued in the above proof of Theorem 3.2, by (24) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) > 0$. Applying Theorem C.1, `DemParity` selects at a higher rate than β for any population proportion $g_{\mathsf{A}} \leq g_0$, where $g_0 = 1/(1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}) \in (0,1)$. In particular, if $\beta = \beta_0$, which we defined as the harm threshold (i.e. $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_0)) = 0$ and $\Delta\boldsymbol{\mu}_{\mathsf{A}}$ is decreasing at β_0), then by the concavity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$, we have that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\mathtt{DemParity}})) < 0$, that is, `DemParity` causes active harm.

F.4 Proof of Theorem 3.4

By Theorem C.3, `EqOpt` selects at a higher rate than β for any population proportion $g_{\mathsf{A}} \leq g_0$, where $g_0 = 1/(1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))})$. Using our assumptions $\beta_{\mathsf{B}}^{\mathtt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$ and $\beta > \beta_{\mathsf{A}}^{\mathtt{MaxUtil}}$, we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) > 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$, by (24) and the monotonicity of $\boldsymbol{u}$. This verifies that $g_0 \in (0,1)$. In particular, if $\beta = \beta_0$, then by the concavity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$, we have that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\mathtt{EqOpt}})) < 0$, that is, `EqOpt` causes active harm.

F.5 Proof of Theorem 3.5

Applying Theorem C.1, we have

$$-\frac{1-g_{\mathsf{A}}}{g_{\mathsf{A}}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) \implies \beta_{\mathtt{DemParity}} > \beta\,.$$

Applying Theorem C.3, we have:

$$\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) \cdot \frac{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{B}}\rangle}{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{A}}\rangle} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} < -\frac{1-g_{\mathsf{A}}}{g_{\mathsf{A}}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta)) \implies \beta_{\mathtt{EqOpt}} < \beta\,.$$

By Theorems 3.3 and 3.4, choosing $g_{\mathsf{A}} < g_2 := 1/(1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))})$ and $g_{\mathsf{A}} > g_1 := 1/(1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta))})$ satisfies the above.

It remains to check that $g_1 < g_2$. Since we assumed $\beta > \sum_{x>\boldsymbol{\mu}_{\mathsf{A}}} \boldsymbol{\pi}_{\mathsf{A}}$, we may apply Lemma *F*.2 to verify this.

Thus we indeed have sufficient conditions for $\beta_{\mathtt{DemParity}} > \beta > \beta_{\mathtt{EqOpt}}$. In particular, if $\beta = \beta_0$, then by the concavity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$, we have that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\mathtt{EqOpt}})) > 0$, that is, `EqOpt` causes improvement, and $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\mathtt{DemParity}})) < 0$, that is, `DemParity` causes active harm.

Lastly, because $\beta_{\mathtt{DemParity}} > \beta_{\mathtt{EqOpt}}$, it is always true that $\Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\mathtt{DemParity}})) > 0 \implies \Delta\boldsymbol{\mu}_{\mathsf{A}}(r_{\boldsymbol{\pi}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\mathtt{EqOpt}})) > 0$, using the concavity of the outcome curve.

Lemma F.2 (Comparison of `DemParity` and `EqOpt` selection rates)**.** *Fix* $\beta \in [0,1]$. *Suppose* $\boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\pi}_{\mathsf{B}}$ *are identical up to a translation with* $\boldsymbol{\mu}_{\mathsf{A}} < \boldsymbol{\mu}_{\mathsf{B}}$. *Also assume* $\boldsymbol{\rho}(x)$ *is affine in* x. *Denote* $\kappa = \frac{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{B}}\rangle}{\langle\boldsymbol{\rho},\boldsymbol{\pi}_{\mathsf{A}}\rangle}$. *Then,*

$$\beta > \sum_{x>\boldsymbol{\mu}_{\mathsf{A}}} \boldsymbol{\pi}_{\mathsf{A}}$$

implies $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) \cdot \kappa \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} < \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))$.

Proof. If we have $\beta > \sum_{x>\boldsymbol{\mu}_{\mathsf{A}}} \boldsymbol{\pi}_{\mathsf{A}}$, by lemma F.3, we must also have $\frac{\boldsymbol{\mu}_{\mathsf{B}}}{\boldsymbol{\mu}_{\mathsf{A}}} < \frac{\mathrm{Q}_{\mathsf{B}}(\beta_0)}{\mathrm{Q}_{\mathsf{A}}(\beta_0)}$. This implies $\kappa = \frac{\sum_x \boldsymbol{\pi}_{\mathsf{B}}(x)\boldsymbol{\rho}(x)}{\sum_x \boldsymbol{\pi}_{\mathsf{A}}(x)\boldsymbol{\rho}(x)} < \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta_0))}$ by linearity of expectation and linearity of $\boldsymbol{\rho}$. Therefore,

$$\kappa \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta_0))} < 1 \tag{25}$$

Further, using $G^{(\mathsf{A}\to\mathsf{B})}(\beta) > \beta$ from lemma F.3 and the fact that $\frac{\boldsymbol{u}(x)}{\boldsymbol{\rho}(x)}$ is increasing in x, we have $\frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} < \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta))}$. Therefore, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) \cdot \kappa \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta_0))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta_0)))} < \kappa \cdot \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(\beta))} \cdot \boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta)) < \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))$ where the last inequality follows from (25). □

We use the following technical lemma in the proof of the above lemma.

Lemma F.3. *If $\boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\pi}_{\mathsf{B}}$ that are identical up to a translation with $\boldsymbol{\mu}_{\mathsf{A}} < \boldsymbol{\mu}_{\mathsf{B}}$, then*

$$G(\beta) > \beta \;\; \forall\, \beta\,, \tag{26}$$

$$\beta > \sum_{x>\boldsymbol{\mu}} \boldsymbol{\pi}_{\mathsf{A}} \implies \frac{\boldsymbol{\mu}_{\mathsf{B}}}{\boldsymbol{\mu}_{\mathsf{A}}} < \frac{\mathrm{Q}_{\mathsf{B}}(\beta)}{\mathrm{Q}_{\mathsf{A}}(\beta)}\,. \tag{27}$$

Proof. For (26), observe that $\mathrm{TPR}_{\mathsf{A}} = \boldsymbol{\rho}(\boldsymbol{\mu}_{\mathsf{A}}) < \mathrm{TPR}_{\mathsf{B}} = \boldsymbol{\rho}(\boldsymbol{\mu}_{\mathsf{B}})$. For any β, we can write $\mathrm{Q}_{\mathsf{B}}(\beta) = \boldsymbol{\mu}_{\mathsf{B}} + c$ and $\mathrm{Q}_{\mathsf{A}}(\beta) = \boldsymbol{\mu}_{\mathsf{A}} + c$ for some c, since $\boldsymbol{\pi}_{\mathsf{A}}, \boldsymbol{\pi}_{\mathsf{B}}$ that are identical up to translation by $\boldsymbol{\mu}_{\mathsf{A}} - \boldsymbol{\mu}_{\mathsf{B}}$. Thus, by computation, we can see that for $\mathrm{Q}(\beta) < \boldsymbol{\mu}$, $\partial_+ G^{(\mathsf{A}\to\mathsf{B})}(\beta) > 1$ and for $\mathrm{Q}(\beta) < \boldsymbol{\mu}$, $\partial_+ G^{(\mathsf{A}\to\mathsf{B})}(\beta) < 1$. Since $G^{(\mathsf{A}\to\mathsf{B})}$ is monotonically increasing on $[0,1]$, we must have $G^{(\mathsf{A}\to\mathsf{B})}(\beta) > \beta$ for every $\beta \in [0,1]$.

For (27), we have $\beta > \sum_{x>\boldsymbol{\mu}} \boldsymbol{\pi}_{\mathsf{A}}$, we can again write $\mathrm{Q}_{\mathsf{B}}(\beta) = \boldsymbol{\mu}_{\mathsf{B}} - c$ and $\mathrm{Q}_{\mathsf{A}}(\beta) = \boldsymbol{\mu}_{\mathsf{A}} - c$, for some $c > 0$. Then it is clear than we have $\frac{\boldsymbol{\mu}_{\mathsf{B}}}{\boldsymbol{\mu}_{\mathsf{A}}} < \frac{\mathrm{Q}_{\mathsf{B}}(\beta)}{\mathrm{Q}_{\mathsf{A}}(\beta)}$. □

F.6 Proof of Theorem 3.6

Proof. $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ implies $g_{\mathsf{A}} \cdot \boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})) + g_{\mathsf{B}} \cdot \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})) > 0$, which by Theorem C.1, implies $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{DemParity}}$.

$\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) > \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$ implies $G^{(\mathsf{A}\to\mathsf{B})}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}) > \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ and so $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}))) < 0$. Therefore by Theorem C.3, we have that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} > \beta_{\mathsf{A}}^{\texttt{EqOpt}}$. □

We now give a very simple example of $\boldsymbol{\pi}_{\mathsf{A}} \prec \boldsymbol{\pi}_{\mathsf{B}}$ where Theorem 3.5 holds. The construction of the example exemplifies the more general idea of using large in-group inequality in group A to skew the true positive rate at `MaxUtil`, making $\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) > \mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}})$.

Example F.1 (`EqOpt` causes relative harm)**.** Let $C = 6$, and let the utility function be such that $\boldsymbol{u}(4) = 0$. Suppose $\boldsymbol{\pi}_{\mathsf{A}}(5) = 1 - 2\epsilon, \boldsymbol{\pi}_{\mathsf{A}}(1) = 2\epsilon$ and $\boldsymbol{\pi}_{\mathsf{B}}(5) = 1 - \epsilon, \boldsymbol{\pi}_{\mathsf{B}}(3) = \epsilon$.

We can easily check that $\boldsymbol{\pi}_{\mathsf{A}} \prec \boldsymbol{\pi}_{\mathsf{B}}$. However, for any $\epsilon \in (0, 1/4)$, we have that $\mathrm{TPR}_{\mathsf{B}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) = \frac{5(1-\epsilon)}{5(1-\epsilon)+3\epsilon} < \mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}^{\texttt{MaxUtil}}) = \frac{5(1-2\epsilon)}{5(1-2\epsilon)+2\epsilon}$.

F.7 Proof of Proposition 4.1

Denote the upper quantile function under $\widehat{\boldsymbol{\pi}}$ as $\widehat{\mathrm{Q}}$. Since $\widehat{\boldsymbol{\pi}} \prec \boldsymbol{\pi}$, we have $\widehat{\mathrm{Q}}(\beta) \leq \mathrm{Q}(\beta)$. The conclusion follows for `MaxUtil` and `DemParity` from Theorem C.1 by the monotonicity of $\boldsymbol{u}$.

If we have that $\mathrm{TPR}_{\mathsf{A}}(\boldsymbol{\tau}) > \widehat{\mathrm{TPR}_{\mathsf{A}}}(\boldsymbol{\tau})\ \forall\ \tau$, that is, the true TPR dominates estimated TPR, the conclusion for `EqOpt` follows from Theorem C.3, by the same argument as in the proof of Theorem 3.6.

F.8 Proof of Proposition 4.2

By Proposition B.6, $\beta^* = \operatorname{argmax}_\beta \Delta\boldsymbol{\mu}_{\mathsf{A}}(\beta)$ exists and is unique. $\beta_0 = \max\{\beta \in [\beta_{\mathsf{A}}^{\mathtt{MaxUtil}}, 1] : \mathcal{U}(\beta_{\mathsf{A}}^{\mathtt{MaxUtil}}) - \mathcal{U}_{\mathsf{A}}(\beta) \leq \delta\}$ which exists and is unique, by the continuity of $\Delta\boldsymbol{\mu}_{\mathsf{A}}$ and Proposition B.6.