Fairness, Welfare, and Equity in Personalized Pricing

Nathan Kallus
Cornell University and Cornell Tech
kallus@cornell.edu

Angela Zhou*
Cornell University and Cornell Tech
az434@cornell.edu

ABSTRACT

We study the interplay of fairness, welfare, and equity considerations in personalized pricing based on customer features. Sellers are increasingly able to conduct *price personalization* based on predictive modeling of demand conditional on covariates: setting customized interest rates, targeted discounts of consumer goods, and personalized subsidies of scarce resources with positive externalities like vaccines and bed nets. These different application areas may lead to *different* concerns around fairness, welfare, and equity on *different* objectives: price burdens on consumers, price envy, firm revenue, access to a good, equal access, and distributional consequences when the good in question further impacts downstream outcomes of interest. We conduct a comprehensive literature review in order to disentangle these different normative considerations and propose a taxonomy of different objectives with mathematical definitions. We focus on observational metrics that do not assume access to an underlying valuation distribution which is either unobserved due to binary feedback or ill-defined due to overriding behavioral concerns regarding interpreting revealed preferences. In the setting of personalized pricing for the provision of goods with positive benefits, we discuss how price optimization may provide unambiguous benefit by achieving a "triple bottom line": personalized pricing enables *expanding access*, which in turn may lead to *gains in welfare* due to heterogeneous utility, and improve *revenue or budget utilization*. We empirically demonstrate the potential benefits of personalized pricing in two settings: pricing subsidies for an elective vaccine, and the effects of personalized interest rates on downstream outcomes in microcredit.

1 INTRODUCTION

Personalized pricing, once restricted to the idealized construction of economic theory, is now squarely within the realm of possibility for firms newly equipped with a deluge of fine-grained information about individuals and prediction modeling of demand or willingness to pay based on this information. Given both the ubiquity of prices and their relevance in important domains such as hiring, lending, and credit subject to antidiscrimination regulation, price personalization remains an area of increasing scrutiny and caution, as well as tentative optimism due to competitive considerations [31, 34].

The potential of expanded reliance on predictive models in domains affecting individuals is cause for concern. After all, the extensive study of fairness considerations in predictive models highlights how the joint structure of protected attributes and other information can lead algorithmic decisions, even based only on non-attribute information, to nonetheless lead to disparate impacts on individuals [8]. The setting of personalized pricing is particularly interesting because it fundamentally involves considerations of both resource allocation in response to price; as well as predictive models for such a price response. Auditing challenges arise precisely because valuations are in general not known or observed; rather only binary-feedback demand response is observed.

Studying the case of personalized pricing is *conceptually* challenging because prices are a shared tool in drastically different domains: we consider lending/insurance, consumer goods, and public provision. A crucial distinction is between *value-based* pricing that offers different prices to customers based on their estimated willingness to pay, and *risk-based* pricing which offers different prices to customers based on their estimated costs, as in lending and insurance [34]. While discrimination law is strongest in insurance and lending, in lending, discrimination concerns often arise from individual agents providing offers from an actuarially-fair securitized rate sheet [9]. In particular, distributional concerns regarding price optimization reflect overall concern for differentially adept/prepared/educated negotiating customers in insurance and lending, but slight optimism in value-based pricing since low-income individuals may be more price-sensitive [9]. Hence, the majority of our analysis will focus on value-based pricing, which lends itself more readily to price optimization.

In the case of value-based pricing, incidents where price targeting leads to disparities are often subject to media coverage and public outcry. We recollect just a few of these incidents: Staples changed prices based on available brick-and-mortar locations of competitors leading to higher prices for rural areas [57] and Asians faced higher prices as a result of the Princeton Review's zip-code based price-targeting [45, 56]. While these *covariate-based* pricing schemes were based on non-group contextual information, nonetheless they induced disparities in prices along group lines. Although there are not clear anti-discrimination principles that govern the setting of value-based pricing, understanding the tradeoffs introduced by considering constraints or fairness penalties on a myopic price optimization problem can shed light on tradeoffs between various intuitive notions to inform algorithm design.

In this paper, we study the interplay of fairness and welfare considerations as they arise in the setting of personalized pricing. Our first task is conceptual: we square these real-world problem settings alongside previously expressed concerns regarding price optimization. For example, acknowledging the empirical reality of economic inequality informs the expected distributional impacts of covariate-conditional pricing schemes along racial and economic segments. We then turn to analytical modeling to identify the multiple objectives which pricing decisions affect with fairness and/or welfare implications. Prices impose a burden on customers (perhaps only on those who purchase), result in allocations of the good itself, may be optimized based on noisy predictors, and the good itself may have downstream impacts of interest. In Section 4, we taxonomize these considerations into price parity (marginal and conditional), model error fairness, preferences for access, allocative efficiency, and actuarial fairness; we offer these operationalizations

(all of which may in themselves not be novel) alongside contextual discussion. We propose operationalizations of these normative considerations which avoid classical assumptions of known valuations, focusing on identifiability based on available information. In Section 5 we provide further characterization of implications for fairness and welfare considerations for considering modifications to price optimization in order to improve on some of these notions, in particular price parity and market share. Where possible, we provide analytical insights on tradeoffs. In Section 6 we build empirical case studies on datasets related to the "public-interest goods" setting based on a study of willingness-to-pay for an elective vaccine and interest rates for microcredit loans.

2 RELATED WORK

The study of algorithmic pricing and revenue management is very extensive and spans economics, operations research [30], and computer science. Price discrimination has also been studied for ethical and normative considerations, especially in relation to privacy [50, 51]. We now highlight methodological and empirical work of particular relevance to algorithmic considerations. In Appendix A we discuss domain-level considerations and more broad related work in greater detail.

We first briefly overview the classical economic taxonomy of price discrimination [59]. First-degree price discrimination offers individual prices to customers exactly at their willingness to pay, which is assumed to be known. Second-degree price discrimination depends on the *quantity purchased* but does not differ across consumers, such as bulk discounts. Third-degree price discrimination charges different prices to different *groups* of consumers, such as offering senior or student discounts. We focus on analyzing covariate-conditional prices, which are a form of third-degree pricing but draw nearer to first-degree pricing (up to the noise of random valuations).

The economic literature studies welfare (consumer and producer surpluses) implications of idealized first-degree personalized pricing, assuming valuation distributions are known. This classical notion of welfare is hence pegged to the valuation distribution (e.g., consumer welfare is valuation minus price). [58] studies third-degree price discrimination; using first-order conditions of valuation distributions, they show that consumer welfare increases with additional price discrimination as long as total output increases. [2] provide analogous conditions for similar analysis. [10] shows that a seller can choose various segmentations that can achieve any combination of increase/decrease in consumer surplus; [20] study the theoretical computational efficiency of finding such segmentations.

This classical theory suggests that ideal personalized pricing may improve welfare relative to uniform monopoly pricing, but that different segmentations lead to possibly indeterminate outcomes for consumer welfare. The empirical literature indicates this indeterminacy in important settings. [29] empirically study implications of machine learning predictors of default probability on disparities not only for predictive performance, but on using these predictions to set interest rates for loans via risk-based pricing. While richer machine learning predictors expand access to credit; they also result in greater price dispersion for the minority borrower on the margin. Hence, greater access comes at the cost of a greater price burden. [24] study personalized pricing in a Bayesian setting with posterior uncertainty quantification in a business-to-business marketing setting; they find that finer-grained personalized pricing overall increases consumer welfare; though this is not monotonic in segmentation granularity.

Pricing in the context of mechanism design follows another approach and assumes elicitation of ontologically valid valuations from strategic participants due to narrowly bracketed contexts such as auctions, kidney exchanges, and matching markets. [27] show that for a two-sided market, market-clearing competitive equilibrium pricing is not necessarily optimal if a market designer has distributional preferences. Noting that classical theory on quasi-linear utility "implicitly embeds the assumption that each agent values money equally," they study implications of dispersion in marginal values for money of market participants for optimal market mechanisms.

[18] study fairness considerations when each protected group is assigned one price and the valuation distribution is known. In contrast, we focus on the prediction setting with rich covariates, and propose metrics that are completely independent of the valuation distribution. Our analytical insights focus on covariate-personalized prices and implications of joint structure of covariate distributions and group variability on fairness considerations.

Finally, we mention work that studies tensions in fair machine learning, specifically the role of algorithms in allocating decisions or conferring utility, to highlight questions of interest that have analogies in the pricing setting. [33, 35, 42, 48] broadly study tensions between fairness and welfare when machine learning enacts allocations, e.g., via classification. An interest of this work is to study analogous considerations for price optimization, under corresponding notions of fairness and welfare. Longer-term considerations of fairness constraints have also been studied [19, 21], typically with the formalism of dynamical systems. In machine learning, regulatory considerations barring "disparate treatment," e.g., using the protected attribute information in the predictor or in algorithmic interventions, may be in fundamental tension with achieving proposed fairness notions. [46] studies tensions that arise from interventions that do not use attribute information. Again, these broad concerns may additionally be of interest in the setting of price optimization.

3 PROBLEM SETUP

We let X denote customer covariates (features) and A the protected attribute. To simplify discussion we focus on binary comparisons between two protected groups, $A = a$ and $A = b$. A personalized pricing policy is a function mapping covariates and possibly protected attribute information to a real-valued price, $p(X, A) \in \mathbb{R}^+$ (or, $p(X)$ for attribute-blind pricing rules). Each sale instance is associated with a hypothetical and unknown demand curve representing the demand for each possible price, $D(p) \in \mathbb{R}^+$. Often, when each sale instance is with an individual customer, $D(p) \in \{0, 1\}$ is binary where $D(p) = 1$ denotes the individual would purchase or take-up at price p. In section 4.4, we further restrict $D(p) = \mathbb{I}[V \leq p]$ to be given by a customer valuation. Otherwise, we do not make this restriction, allowing arbitrary, possibly non-rational behavior.

We are primarily interested in studying the question of *covariate-based pricing*. One approach is to estimate personalized demand at a price, given covariates, via a parametric or semi/non-parametric model, which we will denote as $D(p \mid x, a) = \mathbb{E}[D(p) \mid X = x]$.

A personalized pricing function satisfies:

$$p^*(x, a) \in \underset{p(x,a)}{\operatorname{argmax}} \mathbb{E}[D(p(X, A))p(X, A)]$$

Often, policy or domain-level restrictions may prohibit prices that directly use the attribute A:

$$p^*(x) \in \underset{p(x)}{\operatorname{argmax}} \mathbb{E}[D(p(X))p(X)]$$

We define $P = p(X, A)$ as the *per-individual* personalized price; it may sometimes also refer to $P = p(X)$. We additionally introduce notation for the revenue, $R(P) = PD(P)$, and its covariate-conditional counterpart.

We will also consider cases where the good has an effect on some outcome in itself, such as repeat purchasing behavior, health benefits, downstream welfare, so that $D(p)$ is also itself a *treatment*. We denote the corresponding *potential outcomes* as $Y(D(p))$. These represent the causal outcomes of take-up/non-take-up, e.g., the effect on contracting malaria of purchasing/not purchasing a bed net.

4 DEFINITIONS AND METRICS

We now introduce definitions of different aspects of fairness or welfare in personalized pricing, addressing normative considerations motivated by different contexts, and offer formalisms and operationalizations of these. We consider a generic personalized price function P which may reflect either first or third degree price discrimination. In Section 5 we discuss these operationalizations in more depth and analyze potential trade-offs.

Observational metrics. We introduce the following notions of allocative fairness based on what we may *observe*: prices p, demand outcome $D(p)$ (e.g., purchase/no purchase), and potential downstream outcomes due to the good $Y(D(p))$. In Appendix A we provide further discussion on why, if we are concerned about fairness in the first place, we might be skeptical about defining fairness relative to valuations or other *un*revealed latent preferences/valuations.

4.1 Price parity

Context. A customer always benefits from a lower price. The difference between distributions of prices faced by different groups measures potential unfairness in price burdens. An extreme example is the so-called "pink tax": [22], commissioned by the Mayor's office in New York City, studies gender-based pricing differentials and finds an average of 7% higher prices paid by women; for example, women's pink razors are more expensive, for the exact same product[1]. This highlights the capacity of price optimization to extract consumer value from behavioral failures of economic regarding valuations. In particular, using "pinkness" to segment products corresponds to extracting valuation from social constructions of gender, all other functionality being the same for a product with no signaling value. While the "pink razor" is an extreme example, considerations regarding price parity are a common intuitive objection to personalized pricing.

Operationalization. We introduce a definition based on distributional equivalence of prices for each group.

Definition 1 (Price parity)**.** $P \perp\!\!\!\perp A$

We may also consider parity in moments of the price distribution, which also enables a simple way to give a scalar metric to disparity.

Definition 2 (Marginal price parity)**.**

$$\mathbb{E}[P \mid A = a] = \mathbb{E}[P \mid A = b]$$

Correspondingly, the marginal price *dis*parity is

$$\mathbb{E}[P \mid A = a] - \mathbb{E}[P \mid A = b].$$

The notion of "disparate treatment" in fair machine learning suggests that the following notion of covariate-conditional price parity is intuitively appealing.

Definition 3 (x-conditional price parity)**.** $p(x, a) = p(x, b)$

Notice, however, that satisfying definition 3 generally does not ensure satisfying definitions 1 and 2 due to differing distributions of covariates between groups. Indeed, in general contexts, it is well-understood that equal treatment need not lead to equal impact; this remains true in personalized pricing.

Finally, it is often helpful to consider price parity conditional on take-up, or more generally on demand. Conditioning on take-up reflects that price only affects consumer utility *if* the customer purchases.

Definition 4 (Take-up-conditional parity)**.**

$$P \perp\!\!\!\perp A \mid D(P)$$

More generally, we may wish to condition on the *effect* of pricing. Consider the case where there is a nominal price p_0 and $P \leq p_0$ represents a personalized potential discount. The event $D(P) > D(p_0)$ is the event that demand increases as an *effect* of the discount. In the binary demand case: that an individual purchases if and only if given the discount (rather than purchasing irrespective of discount or not purchasing irrespective), which we term a *responder* to the discount. Conditioning on responsiveness accounts for the possibility that different groups have different valuations or willingness to pay, and to the extent that one deems it acceptable to personalize to leverage such differences (and often it is not) parity conditional on response requires we do not price-discriminate more than is justified by response to discount.

4.2 Model error fairness

Context. Given that fairness in machine learning studies how predictive models may exhibit differential model performance (predictive accuracy, error distributions) by group, a natural question is how such disparities in predictive model performance might affect the suboptimality of different prices; and whether different groups might experience different price suboptimality burdens due to error patterns of the predictive model.

[1]Notably, while legislation has been proposed to try to address the pink tax, this has not been established as gender-based discrimination. See [37] for more discussion.

Operationalization. The rest of this paper studies pricing based on a true conditional demand model, $D(p \mid x, a)$. In practice however, only an estimate $\hat{D}(p \mid x, a)$ is available from observed data. For example, for the pricing problem:

$$\hat{p}^*(x) \in \operatorname{argmin} \mathbb{E}[\hat{D}(p(X))p(X))]$$

Price suboptimality fairness is concerned with the decision suboptimality of a price based on a risk model vs. the price derived from the actuarially fair "true risk", $\hat{p}^*(x) - p^*(x)$. For example, $\hat{p}^*(x)$ may differ from $p^*(x)$ when we learn the prediction $\hat{D}(p \mid x, a)$ from finite samples and differential accuracy thereof could lead to fairness concerns. In Section 5 we will focus on the revenue objective.

4.3 Access and equal access

Context. We are often concerned about *access* to the good being sold, especially when the good has benefits that are deemed crucial such as vaccines (see also our empirical study in section 6.1), loans, or broadband internet. In terms of welfare, it is important to consider the total access that personalized schemes lead to, namely the total demand. In terms of fairness, we may be concerned with allocative parity in the form of parity in market shares or take-up probabilities by group. Personalized pricing schemes may in fact enhance both of these measures by allowing revenue extraction from high-valuation groups to enable offering lower price offers to low-valuation individuals, hence "pricing people into the market." High-valuation groups are usually those with financial means that allows them to have a higher willingness to pay, and low-valuation groups are usually those with less financial means.

Operationalization. Total access is simply the marginal demand. When demand is binary, this is the fraction of individuals who take-up the good.

Definition 5 (Total access)**.**

$$\mathbb{E}[D(P)].$$

The idea of access parity suggests requiring equal allocation of access / market share / take-up.

Definition 6 (Access parity)**.**

$$\mathbb{E}[D(P) \mid A = a] = \mathbb{E}[D(P) \mid A = b].$$

Correspondingly, access *dis*parity is $\mathbb{E}[D(P) \mid A = a] - \mathbb{E}[D(P) \mid A = b]$.

4.4 Allocative efficiency and fairness

Context. In settings where interpreting revealed preferences as rational choices due to latent valuations, efficiency considerations are concerned with how prices *sort* individuals by their valuations and ensure that a good may be *targeted* towards those who value it the most. For example, the literature on pricing health interventions in development is particularly interested in the difference between free and low-price provisions based on whether prices better target households who are more likely to use a product [16]. An important further concern is whether errors in sorting individuals disproportionately affect one group more than another so that, on average, certain groups are more often incorrectly given priority over others.

Operationalization. We focus on providing observational metrics which assess sorting/targeting *without* assuming access to the full valuation distribution. For this section, we focus on the binary feedback setting where $D(p) \in \{0, 1\}$.

Assumption 1 (Monotonicity)**.** For any p, p' :

$$p > p' \implies D(p) \geq D(p')$$

Assuming monotonicity of binary demand with respect to price as in Asn. 1 is equivalent to assuming a random latent threshold model, i.e., $D(p) = \mathbb{I}[V \leq p]$. This perspective recognizes that observations of the binary event $D(p)$, under Asn. 1, are censored observations of the underlying valuation.

The question is whether a given pricing scheme (for which we have observed the binary outcomes) appropriately ranks valuations of individuals. A marginal measure of such efficiency may be *concordance*:

Definition 7 (Concordance)**.** Given two individuals drawn independently at random, *concordance is*

$$\mathbb{P}(V_1 > V_2 \mid P_1 > P_2).$$

While concordance captures how efficiently prices sort individuals by valuation, such efficiency may have disparate effects. To capture this disparity, we propose the *class-crossed concordance disparity* metric.

Definition 8 (Class-crossed concordance disparity)**.** Given two individuals drawn independently at random from groups $A = a$ and $A = b$, respectively, *class-crossed concordance disparity* is

$$\mathbb{P}(V_b > V_a \mid P_a < P_b) - \mathbb{P}(V_a > V_b \mid P_b < P_a)$$

Class-crossed concordance has the following probabilistic interpretation. The term $\mathbb{P}(V_b > V_a \mid P_a < P_b)$ can be interpreted as: of those whose valuations can be ordered under Asn. 1, what is the probability that valuations drawn from one group are *stochastically greater* than valuations from another group? Class-crossed concordance measures a groupwise disparity in the difference in these probabilities.

4.5 Targeting long-run dynamics

Context. A key domain-level consideration that justifies *preferring take-up* is that take-up of the good is itself a treatment with a downstream outcome, such as future purchases/customer loyalty in e-commerce, net present value of continued borrowing [40], or usage and downstream health outcomes of a preventive health intervention in development economics [16, 26, 38]. Therefore, price impacts an allocation which itself may have heterogeneous effects on longer-term outcomes for the customer and/or decision-maker: we identify this as targeting long-run dynamics. This, for example, justifies an overall preference for expanding market shares.

The possibility of a "triple bottom line". Price personalization may be beneficial due to its increasing take-up of a good which is beneficial for individuals and the decision-maker, targeted price subsidies enable budget-balanced public provision (in contrast to a complete subsidy), and the access expansion might particularly those who would benefit the most. Of course, whether these benefits compound (or whether certain contributors are irrelevant) need to be assessed in the data of any particular setting, as in [18, 25].

	Insurance	Lending	Consumer goods	Public-interest goods
Moral hazard/Adverse selection	✓	✓	✗	✗
Revenue-driven price optimization	✗	✗	✓	✓
Risk-driven price optimization	✓	✓	✗	✗
Prefer marginal price parity	✗	✗	✓	✗
Prefer conditional price parity	✓	✓	✗	✗
Prefer access	✓	✓	✓	✓
Actuarial fairness	✓	✓	✗	✗
Allocative efficiency/sorting	✗	✗	✓	✓
Targeting long-run dynamics	✓	✓	✓	✓

Table 1: Different problem settings and *what* fairness/welfare notions are relevant *when*.

Operationalization. We recognize the firm's objective function as population welfare downstream of a price allocation:

$$\mathbb{E}[Y(D(P))].$$

4.6 Summary of problem settings and relevant notions

Abstractly, we might summarize some of the above considerations by considering a $\lambda-$scalarized multi-objective optimization problem, which represents the expansion of considerations beyond myopic revenue maximization:

$$\begin{aligned} \max_{P(\cdot)} \ & \mathbb{E}[PD(P)] + \lambda_1(\textstyle\sum_{a\in\mathcal{A}} \mathbb{E}[D(P) \mid A=a]) + \lambda_2\mathbb{E}[Y(P)] \\ \text{s.t. } & \mathbb{E}[P \mid A=a] - \mathbb{E}[P \mid A=b] \le \Gamma \\ & \mathbb{E}[D(P)(P-c)] \ge 0 \end{aligned} \tag{1}$$

Sections 4.1 and 4.3 and **??** (price parity, access, long-run welfare) might conceivably be included in a conceptual "multi-objective" version of the firm's problem, while Sections 4.2 and 4.4 (price suboptimality, class-crossed concordance) are idealized measures.

In table 1 we apply our conceptual taxonomy of problem domains to these different fairness/welfare notions in pricing. The first category of criteria summarizes how different problem settings differ in the presence of moral hazard/adverse selection that justifies risk-based pricing), and the capacity for price optimization. The second category identifies which notions of fairness or equity are more or less relevant in different settings.

We caution that the above optimization problem is a conceptual device to illustrate how these notions might justify revenue sub-optimal allocations. Table 1 suggests that in any particular application setting, these notions may not be simultaneously relevant.

5 ANALYSIS OF DEFINITIONS AND METRICS

In this section, we expand further on each definition. Where possible, e.g. by making additional assumptions, we provide analytical insight on implications of these fairness notions for price optimization or corresponding specializations of eq. (1).

5.1 Price-parity

We study the price optimization problem with additional price parity constraints, focusing on highlighting tradeoffs with implications for algorithm design. To simplify the analysis, we make the following assumptions. We assume a partially linear demand model with a link function of the non-price, covariate-driven demand $\overline{D}(x,a)$ and a linear component for price elasticity.

Assumption 2 (Partially linear demand model)**.**

$$D(p \mid x, a) = \beta_a p + \overline{D}(x,a).$$

Assumption 3 (Downward sloping linear demand with respect to price.)**.** $\beta_a < 0, \forall a \in \mathcal{A}$

For example, the linear model corresponds to $\overline{D}(x,a) = \alpha + \gamma^\top x$. Linear demand is a common assumption for contextual pricing [6, 11, 52]. We also assume that price elasticities are negative.

Without loss of generality, assume group a has higher average price at the unconstrained solution). The marginal parity[2] constrained revenue maximization problem, is a specialization of eq. (1). Let $p^*(x,a;\Gamma)$ denote the corresponding $\Gamma-$parametrized solution.

$$\begin{aligned} p^*(x,a;\Gamma) \in \operatorname*{argmax}_{p(\cdot)} \ & \mathbb{E}[p(X,A)D(p(X,A))] \\ \text{s.t. } & \mathbb{E}[p(X,A) \mid A=a] - \mathbb{E}[p(X,A) \mid A=b] \le \Gamma \end{aligned} \tag{2}$$

The *attribute-blind* personalized price is $p^*(x;\Gamma)$, which restricts the above optimization to prices which only personalize on x. We derive the parity-constrained optimal price.

Theorem 1. *Let*

$$\xi(A) = (\mathbb{P}[A=a]^{-1}\mathbb{I}[A=a] - \mathbb{P}[A=b]^{-1}\mathbb{I}[A=b]).$$

The optimal attribute-based personalized price under marginal parity solving eq. (2) *is:*

$$p^*(x,a;\Gamma) = \frac{-\overline{D}(x,a) + \xi(a)\lambda^*_{xa}}{2\beta_a},$$
$$\textit{where } \lambda^*_{xa} = \frac{\mathbb{E}[\overline{D}(X,A)\xi(A) + 2\beta_A\Gamma]}{\mathbb{E}[\xi(A)^2]}.$$

[2]We discuss constraining first moments of the price distributions (marginal parity) to provide analytical insights. Constraining higher order moments via e.g. a set constrained by Kolmogorov-Smirnov statistic [49], or a moment-based hierarchy as in [4], is computationally possible.

The optimal attribute-blind personalized price is

$$p^*(x;\Gamma) = \frac{-\overline{D}(x) + \lambda_x^* \mathbb{E}[\xi(A) \mid X = x]}{2\mathbb{E}[\beta_A \mid X = x]},$$
$$\text{where } \lambda_x^* = \frac{\mathbb{E}[\overline{D}(X)\xi(A) + 2\beta_A\Gamma]}{\mathbb{E}[\mathbb{E}[\xi(A) \mid X]^2]}.$$

□

The proof is included in Appendix B; the key idea is to study the Lagrangian dual of the knapsack-constrained quadratic program and solve by swapping the order of minimum and maximum.

In the following analysis, we focus on an equality constraint in eq. (2). Interpreting the solution, $p^*(x,a)$ differs from the unconstrained personalized price by a penalty whose size depends on the discrepancy of the price-independent covariate-based demand within groups.

We highlight some tradeoffs induced by marginal price parity against other fairness considerations. We first consider a very special setting where the demand function is invariant across groups; it is only the group-conditional covariate distribution which induces price disparity.

Proposition 1 (Attribute-based vs. attribute-blind pricing under marginal parity)**.** Suppose Asns. 2,3, and further that:

(1) $\overline{D}(x,a) = \overline{D}(x,b) = \overline{D}(x)$,
(2) $\overline{D}(x)$ is linear in x,
(3) and $\beta_a = \beta_b$.

Then we have

$$p^*(x,a;0) - p^*(x;0) < 0 \iff$$
$$\frac{\lambda_{xa}^*}{\lambda_x^*} < \mathbb{P}(A = a \mid X = x) - \frac{\mathbb{P}[A=a]}{\mathbb{P}[A=b]}\mathbb{P}(A = b \mid X = x),$$
$$\text{and } p^*(x,b;0) - p^*(x;0) < 0.$$

In this very special case, we find that group b would uniformly prefer attribute-based marginal parity pricing. The relationship is not necessarily uniform for group a. A sufficient condition, for example, is if for some x, $\mathbb{P}(A = a \mid X = x) >> \mathbb{P}(A = b \mid X = x)$ and $\mathbb{P}[A = a] = \mathbb{P}[A = b]$: then $p^*(x,a;0) - p^*(x;0) < 0$. Hence, for $x-$outliers within group a satisfying this condition, attribute-based pricing is Pareto optimal relative to attribute-blind pricing for *both* groups. Since however price disparity in this special case is exactly driven by variability in $\mathbb{P}(A = a \mid X = x)$, we might also expect that $p^*(x,a;0) - p^*(x;0) > 0$ for some other x.

Building on the results of Theorem 1 and Proposition 1, we also provide bounds on the revenue loss due to attribute-blind pricing, now in the setting where the non-price-based demand may differ across groups.

Corollary 1 (Revenue loss of $p^*(x)$). *Suppose Asn. 3, $\beta_a = \beta_b$ and $\overline{D}(x,a) \neq \overline{D}(x,b)$. Then,*

$$\mathbb{E}[R(p^*(X,A)) - R(p^*(X))] \geq \frac{1}{4\beta}\mathbb{E}[\overline{D}^2(X) - \overline{D}^2(X,A)] \geq 0.$$

We now use these characterizations from Theorem 1, Proposition 1, and Corollary 1 to study tradeoffs between price parity and other desiderata, in particular x-conditional parity, and summarize some implications for algorithm design.

(1) A firm may generically prefer attribute-blind pricing to attribute-based pricing schemes due to regulatory considerations or x-conditional price parity (definition 3).
(2) If achieving marginal parity is of interest in view of price disparities, attribute-blind marginal parity achieves lower firm revenue than attribute-based marginal parity. We provide a quantitative bound on the gap in a simple case in Corollary 1.
(3) Under the special case of Proposition 1, for some x, attribute-based marginal parity is strictly preferable to attribute-blind marginal parity for *both* groups.

These considerations might outweigh an intuitive preference for attribute-blind pricing in the case of marginal parity.

5.2 Price suboptimality fairness

We provide a decomposition that gives structural conditions on when the sign of the prediction error is informative of the sign of the mispricing or decision error. In particular, this highlights a distinction between analyzing fairness in data-driven optimal prices vs. fair prediction in machine learning.

Proposition 2 (Price suboptimality error decomposition)**.** Assume that $\nabla\hat{D}(\hat{p}^*(x) \mid x) \neq \nabla\hat{D}(p^*(x) \mid x)$. Up to first order terms,

$$\hat{p}^*(x) - p^*(x) = \frac{\hat{D}(p^*(x) \mid x) - D(p^*(x) \mid x)}{\nabla\hat{D}(\hat{p}^*(x) \mid x) - \nabla\hat{D}(p^*(x) \mid x)} + p^*\left(1 + \frac{\nabla\hat{D}(p^*(x) \mid x) - \nabla D(p^*(x) \mid x)}{\nabla\hat{D}(\hat{p}^*(x) \mid x) - \nabla\hat{D}(p^*(x) \mid x)}\right)$$

The decomposition is not computable from observed data (since some quantities depend on the unknown $p^*(x)$). One implication is that the sign of $(\hat{p}^*(x) - p^*(x))$ ultimately depends on the sign of a few quantities:

1	$\hat{D}(p^*(x) \mid x) - D(p^*(x) \mid x)$	estimation error
2	$\nabla\hat{D}(p^*(x) \mid x) - \nabla D(p^*(x) \mid x)$	gradient est. error
3	$\nabla\hat{D}(\hat{p}^*(x) \mid x) - \nabla\hat{D}(p^*(x) \mid x)$	price elast. subopt.

There are some cases where we may be able to conclude the sign of decision error $\hat{p}^*(x) - p^*(x)$: we can conclude $\hat{p}^*(x) - p^*(x) > 0$ if $1, 2, 3$ are all positive.

However, in general, the main implication of the above proposition is that the direction of *decision disparity* is not immediate from *prediction error* of $D(p(x) \mid x)$ alone: it also depends on estimation error of the gradient, and the difference in gradients due to suboptimality. It is more difficult to conclude implications of pricing decisions (and more broadly, optimization decisions) based on uncertain nuisance predictions.

Applying the above result to $p^*(x,a) - p^*(x)$ allows us to make a similar conclusion for the discrepancy of pricing with respect to attribute-based $D(p(x) \mid x,a)$ vs. the attribute-blind $D(p(x) \mid x)$ setting.

$$p^*(x,a) - p^*(x) = \frac{D(p^*(x) \mid x,a) - D(p^*(x) \mid x)}{\nabla D(p^*(x,a) \mid x,a) - \nabla D(p^*(x) \mid x,a)} + p^*\left(1 + \frac{\nabla D(p^*(x) \mid x,a) - \nabla D(p^*(x) \mid x)}{\nabla D(p^*(x,a) \mid x,a) - \nabla D(p^*(x) \mid x,a)}\right)$$

5.3 Market share

We study a multi-objective version of eq. (1) with additional weights on group-conditional market share objectives. We consider demand that arises from an underlying valuation distribution. (This is true without loss of generality under Asn. 1 of monotonicity.) The most general assumption that admits a concave price optimization program is *log-concavity* of valuation distributions.

Assumption 4 (Log-concave valuation distribution)**.** Suppose that

$$V = g(x, a) + \epsilon,$$

where ϵ has a log-concave probability density function.

We assume log-concavity so that the cdf and cumulative cdf of ϵ (effectively $D(p \mid x, a)$) are also log-concave. Log-concavity is quite general; log-concave pdfs include the normal, exponential, logistic, extreme value, Laplace, gamma, Weibull, etc.

Then the population-level market share personalized-price specialization of eq. (1) is a concave program under Asn. 4, since maximization is equivalent under the monotonic increasing log transformation:

$$p^*(x, a; \lambda) \in \underset{p(\cdot)}{\operatorname{argmax}}\, \mathbb{E}[(p(X, A) + \lambda) D(p(X, A))] \tag{3}$$

We may consider the *attribute-blind* restriction of the above:

$$p^*(x; \lambda) \in \underset{p(\cdot)}{\operatorname{argmax}}\, \mathbb{E}[(p(X) + \lambda) D(p(X))] \tag{4}$$

The above problems consider a *population-level* market share penalty; we also consider group-conditional market share penalties. These may arise from the penalty formulation of the market-share constrained problem, where *some* λ is the optimal dual Lagrange multiplier for the constraints $\mathbb{E}[D(p) \mid A = a] \geq \Gamma$.

$$\max\{\mathbb{E}[p(X, A) D(p(X, A))] : \mathbb{E}[D(p(X, A)) \mid A = a] \geq \Gamma_a\} \tag{5}$$

Correspondingly, the *attribute-based* group-conditional market share price $p^*(x, a; \lambda_a)$ solves, pointwise over x, a:

$$p^*(x, a; \lambda_a) \in \underset{p(\cdot)}{\operatorname{argmax}}\, (p + \lambda_a / \rho_a) D(p \mid x, a), \quad \forall x, a \tag{6}$$

We study the sensitivity of the unconstrained-optimal attribute-blind and attribute-based personalized price to *local increases* in the penalty parameter, λ, relative from the unconstrained optimal price, i.e. $\nabla_\lambda p^*(x, a; 0)$. This describes how much the price changes in response to implementing distributional preferences for market share. Quantifying these sensitivities sheds light on the dependence on R'', the second derivative of the (conditional) revenue function.

Lemma 1 (Optimality conditions for different penalties). *Suppose Asn. 4.*

(1) Population-level market share, attribute-based sensitivity is

$$\nabla_\lambda p^*(x, a; 0) = R''(p^*(x, a; 0) \mid x, a)^{-1} p^*(x, a; 0)^{-2}.$$

(2) Population-level market share, attribute-blind sensitivity is

$$\nabla_\lambda p^*(x; 0) = R''(p^*(x; 0) \mid x)^{-1} p^*(x; 0)^{-2}.$$

(3) Group-level market share, attribute-based sensitivity is

$$\nabla_\lambda p^*(x, a; 0) = \tfrac{1}{\rho_a} R''(p^*(x, a; 0) \mid x, a)^{-1} p^*(x, a; 0)^{-2}.$$

Observe that these sensitivities are negative, under Asn. 4. The proof, included in the appendix, identifies the (pointwise) optimality conditions of the constrained optimizations, eq. (4), eq. (6), and applies the implicit function theorem.

We highlight some implications of Lemma 1 for algorithm design.

(1) For larger $|R''|$ (greater curvature), the less price decrease is required to increase market share, and conversely for smaller $|R''|$, the larger price decrease is required.
(2) Quantifying these sensitivities in terms of R'' highlights the *revenue* implications of these price fairness changes. Considering a second-order expansion of the revenue, smaller $|R''|$ suggests that the larger price decrease may not have extreme *revenue decrease*.
(3) Curvature also quantifies the rate of convergence of the optimal price, e.g. if optimizing over a parametrized pricing policy via M-estimation [44]. Hence, finite-sample variability of the optimal price (which may be assessed empirically by bootstrapping) may suggest low curvature. This suggests a robust approach which ensures out-of-sample market share may incur small revenue tradeoff in the low-curvature regime.

5.4 Allocative efficiency: concordance

Assumption 1, of almost sure monotonicity, suggests that the combination of continuous treatment and binary outcome can be viewed as a censored observation of the valuation. Again, we do not assume access to the underlying realizations of valuation distribution, but study what may be concluded about valuations given that we only observe the censored realizations $D(P) = \mathbb{I}[V > P]$.

Consider ranking the prices and valuations of two generic price-valuation-demand triples, $(p_1, v_1, D(p_1)), (p_2, v_2, D(p_2))$. The only joint outcome of demands and prices that admits concluding an ordering *on the underlying valuations* v_1, v_2 is that

$$\{p_1 < p_2, D(p_1) = 0, D(p_2) = 1\} \iff \{v_1 < p_1 < p_2 < v_2\}$$

Using this observation (which is highly dependent on almost sure monotonicity), we can identify a lower bound on concordance from observational data.

Theorem 2. *Assume Asn. 1 and* $D(p) \in \{0, 1\}$.

$$\mathbb{P}(D_b(P_b) > D_a(P_a) \mid P_a < P_b) \leq \mathbb{P}(V_a > V_b \mid P_a < P_b)$$

Note that $\mathbb{P}(D_b(P_b) > D_a(P_a) \mid P_a < P_b)$ is related to the concordance index of sensitivity analysis, in particular the perspective studied by [54] who suggest a ranking-based approach to survival analysis.

In survival analysis, right-censoring occurs when there is a finite horizon end to data collection for the survival time of patients, so that the observed survival times are the minimum of the censoring time and the actual survival time. The concordance score is a generalization of Wilcoxon-Mann-Whitney statistics and the AUC that applies to continuous output variables, and accounts for censoring of the data. It is the fraction of all pairs of subjects whose predicted survival times are correctly ordered among all subjects that can be ordered. [54] observe that two subjects' survival times can be ordered not only if both of them are uncensored; but also if the uncensored time of one is smaller than the censored time of another.

Relative to the concordance index of right-censored survival analysis, the setting of allocative efficiency is more difficult: we are required to further restrict attention to pairs $p_1 < p_2$, and we can at best order $v_1 < p_1 < p_2 < v_2$.

5.5 Targeting for long-run dynamics: optimal encouragement designs

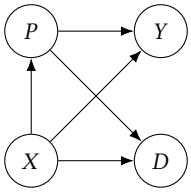
An important justification for cross-subsidy (preferring take-up) is in recognizing that take-up of the good, $D(p)$ may itself be a treatment for downstream outcomes. Such outcomes of interest might include long-term customer value, social learning, repeated purchase behavior, "compliance", attrition, etc. We highlight that one might view p as either a continuous treatment or instrument.

For example, this is a major focus of [40] which considers the amortized net present value of customers over a long time-horizon after they take-up a microcredit loan. Their analysis suggests that while price discrimination to expand take-up may result in losses in short-term profits, this can be outweighed by clients "aging into" a loan portfolio and becoming more profitable. This is also of concern for health interventions in development: take-up of the good is not the final outcome of interest, but rather health outcomes are.

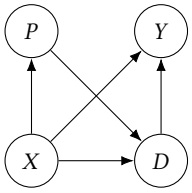
In particular, the possibility of such a "triple bottom line" highlights a situation where *non-ideal theory* that propagates the effects of known inequality to the expected failures of classical economic theory may highlights possible opportunities for personalized pricing to achieve practical benefit. For example, a plausible narrative recognizes that poorer households "undervalue" preventive healthcare not because of some fundamental "underlying preference" for poor health, but due to behavioral considerations and cognitive burdens which prevent endogenizing the full health benefits of a product [7], or for far more practical reasons since they may simply have lower incomes. As a result, they may be more price sensitive. And, they may receive outsized additional health benefits from using the preventive health intervention if they are indeed liquidity constrained due to lack of other ancillary health interventions. Personalized price offers allow making lower price offers, increasing take-up, and if indeed these households "priced in on the margin" receive greater heterogeneous benefits, larger welfare improvements.

There are two perspectives. In Figure 1b, price is a continuous instrument for treatment, and hence, outcome[3]. That is, recognizing that one cannot directly assign the actual treatment of interest – one can neither force loans upon individuals nor ethically randomly reject individuals who apply for loans, exogenous price variation may be the only tool from observational data for assessing the causal impacts of loans on welfare. Crucially, price satisfies the main assumptions for instrumental variables; *relevance*, that it predicts treatment (loan take-up), and more importantly the *exclusion restriction*: that $Y(D(p)) = Y(D), \forall p$, the only effect of the price on outcome is via its impact on treatment [36]. This may be plausible when the outcome of interest is a quantity such as health impacts of a bednet on malaria incidence [13], impacts of sanitation on health outcomes, [38], or social learning for sustained use of the

[3]Estimating a covariate-conditional local instrumental variable curve, or optimal policy when the price instrument is the control variable, remains an open problem. See [43] for doubly robust estimation of the continuous instrument or [55] for policy optimization with heterogeneous effects and discrete instruments.



(a) Price as treatment



(b) Price as instrument

intervention via subsidized first use [26]. In the lending setting, this may be plausible if it is believed interest rates do not affect default event, or the amount borrowed.

Another perspective views price as continuous treatment, e.g. Figure 1a. There are some posited behavioral economics effects which may lead to a failure of the exclusion restriction such that price affects outcome, such as anchoring to reference prices (which attenuates future take-up) or sunk-cost fallacies (when high prices encourage usage/non-wastage); this is explored in [26]. Alternatively, interest rates might affect default probability if individuals are liquidity-constrained. Evidence is mixed in lending: [29] assumes this, [3] finds no effect, and [41] finds some effect of rates on default. Alternatively, interest rates might have an effect on the extensive demand margin (amount borrowed). In this setting, we might instead consider price as a continuous treatment with a composite outcome of take-up and observed outcome, conditional on take-up (such as amount borrowed or default outcome).

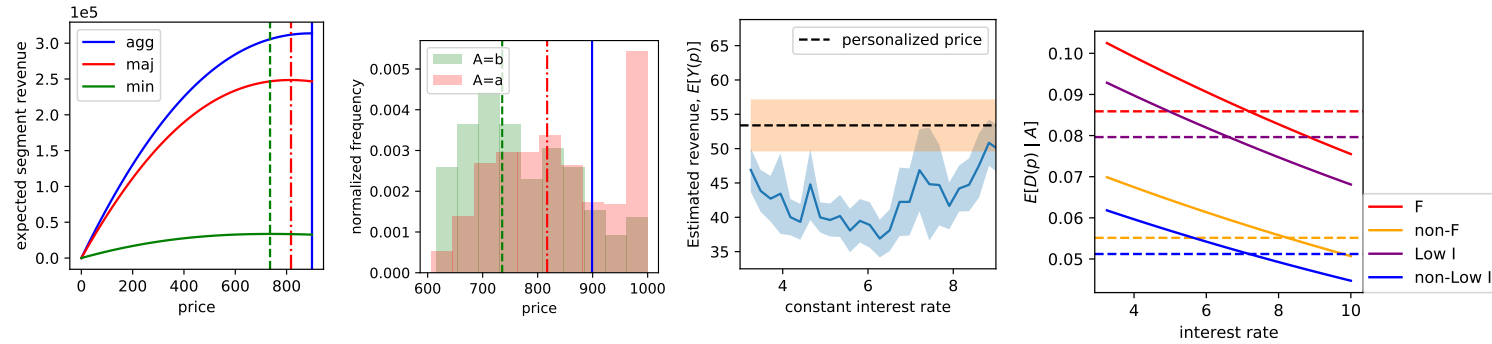
From the perspective of optimization, we generally view price as a treatment and optimize for corresponding downstream outcomes, e.g. conduct an "intention to treat" analysis.

6 CASE STUDIES

6.1 Willingness to pay for elective vaccine

We build a case study from [53], a willingness-to-pay study for vaccination against tick-borne encephalitis in Sweden. The vaccine for tick-borne encephalitis (TBE) is elective and the study is interested in assessing determinants of willingness to pay to inform health policy. Demand is associated with price and income; as well as individual contextual factors such as age, geographic risk factors, trust, perceptions and knowledge about tick-borne disease. The health policy recommendation uses the learned demand model to estimate the vaccination rate under a free, completely subsidized vaccine. This setting corresponds to the setting of public provision, where a decision-maker has a preference for higher take-up due to dynamic externalities of vaccines (which are nonetheless difficult to precisely estimate or target). The study was a contingent valuation study which asked individuals about take-up at a random price of 100, 250, 500, 750, or 1000 SEK. The study finds that "The current market price of the TBE vaccine deters a substantial share of at-risk people with low incomes from getting vaccinated."

In Figures 2a and 2b, we compare distributional considerations of segmented vs personalized pricing. Let $A = b$ indicate low-income. We follow [53] and learn a logistic regression model of binary demand by simply appending the price covariate with the other covariates, so that $D(x, p) = \sigma(\gamma^\top x + \beta p)$. A natural approach in the setting where a free subsidy is not feasible due to budget constraints, is *third-degree* price discrimination: segment based on



(a) Revenue curves for $A = \{a, b\}$, a, b segments

(b) Distribution of $p^*(X) \mid A = a, p^*(X) \mid A = b$

(c) Estimated revenue comparison

(d) Comparing take-up

Figure 2: Willingness to pay for elective vaccine (Figures 2a and 2b) and Microcredit (Figures 2c and 2d).

income and offer a price to low-income and high-income groups separately. We consider such a group-segmented approach in Figure 2a. The blue curve ("agg") represents the revenue curve for a uniform price. The red ("maj") and green ("min") curves are revenue curves from the majority (high-income) and minority (low-income) groups, respectively. We indicate the resulting optimal prices for each of these curves with vertical dashed lines. Notably, the $A = a$ revenue curve has greater second-degree curvature than the $A = b$ revenue curve. Because of the flat revenue curve, the market share of the minority group can be substantially increased without much extra cost.

We next consider a covariate-driven personalization approach. In Figure 2b, we plot histograms of the group-conditional distributions of these optimal prices, $p^*(X) \mid A = a$ and $p^*(X) \mid A = b$. The optimal group-based prices are indicated by the vertical lines for reference. We solve

$$p^*(x) \in \underset{p(\cdot)}{\operatorname{argmax}} \ \mathbb{E}[D(p(X))p].$$

Notably, the optimal prices for the low-income group are overall lower than those for the high-income group. Expected take-up in this segment increases to 27.5% from 24.2% under the uniform monopoly price or 26.7% under $p^*(b)$, the optimal group-based price of Figure 2a. Compared to the uniform monopoly price which obtains expected optimal revenue of $313.6 \cdot 10^3$, the group-based segment scheme $p^*(A)$ obtains expected revenue of $282.2 \cdot 10^3$, and the personalized pricing scheme $p^*(X)$ obtains expected revenue of $318.8 \cdot 10^3$. In this setting, covariate-based personalized pricing is strictly beneficial in terms of (mild) *revenue benefits* that are also able to achieve *greater market share* for the minority group. While the group-based segmentation results in a lower price for the minority group, it is overall not incentive-compatible for a decision-maker to use this segmentation because it attains less revenue than even uniform monopoly pricing.

6.2 Credit Elasticities

[41] randomize prices for a microfinance lender for repeat borrowers. An extensive literature on microfinance sought to assess whether microcredit was able to provide longer term benefits in improving outcomes for household. The rise of the sector led to partially subsidized lenders as well as interest from the private sector. The question of the study was to leverage price randomization in the microcredit setting and assess the effects of lower, or higher, interest rates on revenue for the lender. Overalal, the findings suggest lower rates could decrease profits by a small amount. But the paper considers that at the domain level, since microfinance initiatives may have targeting preferences, e.g. for financial inclusion for women or lower-income individuals, such potential mild profit losses could be offset by expanded inclusion of these target groups due to heterogeneity in take-up.

This setting could present an opportunity for personalized pricing to differentially lower interest rates and expand revenue to targeted groups. Adopting an "intention to treat" analysis, we use the method of [39] to consider off-policy evaluation and optimization of a continuous linear personalized pricing policy from the randomized controlled trial data. The policy parametrization is linear in the covariates, which include income, demographics, location, and loan history information. The method of [39] considers a kernel-based estimator of the counterfactual value of a pricing policy. We use the Epanechnikov kernel and a bandwidth of 0.3; the optimization is non-convex. Because of the fundamental problem of causal inference, we lack the ability to directly assess outcomes.

Nonetheless, we provide some comparison of the estimated revenue and market shares under the personalized policy. We consider a 50/25/25 training/nuisance estimation/validation split, training a random forest on the nuisance estimation split, and learning an optimal policy on the training data with a doubly robust estimator. We use the random forest to estimate the revenue of the personalized policy in comparison to constant interest rates on the validation set.[4] Finally, to indicate the sampling variation in our comparison induced by training the benchmark model, we repeat draws of the nuisance/validation sets, and report the sampling variation in revenue estimates via confidence bands of one standard error.

We include the results in Figure 2c and Figure 2d. Figure 2c plots the random-forest imputed revenue of the personalized allocation

[4]This is in general a biased "direct method", and we take care to avoid extrapolation from interest rates. However, for the sake of comparing against constant treatment assignment, using the direct method reduces variance.

(indicated in black dashed, plotting one standard error), comparing against the imputed revenue via the random forest model of constant interest rates (on the x-axis), in blue. The personalized allocation rule increases estimated revenue (as expected). We assess some of the distributional characteristics of the resulting allocation. In Figure 2d, we compare the access properties of the personalized decision rules for subgroups of interest, namely female and non-female borrowers and low-income and non-low-income borrowers. The estimates of access in Figure 2d are based on a logistic regression of demand, learned on the nuisance estimation dataset.

In horizontal dashed lines, we plot the access estimates for these subgroups under the personalized allocation rule. Note the achievable subgroup access levels correspond to intersections of vertical lines with the demand curves. In comparison, the personalized pricing allocation rule is able to increase access for female and low-income borrowers, relative to the optimal constant interest rate (around 7.2). The unconstrained optimal personalized price however achieves lower takeup for non-low income borrowers (i.e. it increased interest rates for them). Stronger distributional guarantees may be possible by further constraining the price optimization problem. Overall, the goal is to highlight that personalized pricing can improve firm revenue, as well as increase access and improving targeting abilities.

REFERENCES

[1] Case c-236/09 association belge des consommateurs test-achats asbl and others v conseil des ministres, 2011. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62009CJ0236:EN:HTML.

[2] I. Aguirre, S. Cowan, and J. Vickers. Monopoly price discrimination and demand curvature. *American Economic Review*, 100(4):1601–15, 2010.

[3] S. Alan and G. Loranth. Subprime consumer credit demand: Evidence from a lender's pricing experiment. *The Review of Financial Studies*, 26(9):2353–2374, 2013.

[4] A. Aswani and M. Olfat. Optimization hierarchy for fair statistical decision problems. *arXiv preprint arXiv:1910.08520*, 2019.

[5] R. Avraham, K. D. Logue, and D. Schwarcz. Understanding insurance antidiscrimination law. *S. Cal. L. Rev.*, 87:195, 2013.

[6] G.-Y. Ban and N. B. Keskin. Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. *Available at SSRN 2972985*, 2020.

[7] O. Bar-Gill. Algorithmic price discrimination: When demand is a function of both preferences and (mis) perceptions. *The Harvard John M. Olin Discussion Paper Series*, (05):18–32, 2018.

[8] S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2018. http://www.fairmlbook.org.

[9] R. Bartlett, A. Morse, R. Stanton, and N. Wallace. Consumer-lending discrimination in the fintech era. Technical report, National Bureau of Economic Research, 2019.

[10] D. Bergemann, B. Brooks, and S. Morris. The limits of price discrimination. *American Economic Review*, 105(3):921–57, 2015.

[11] O. Besbes and A. Zeevi. On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. *Management Science*, 61(4):723–739, 2015.

[12] J. Beshears, J. J. Choi, D. Laibson, and B. C. Madrian. Behavioral household finance. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 1, pages 177–276. Elsevier, 2018.

[13] D. Bhattacharya and P. Dupas. Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics*, 167(1):168–196, 2012.

[14] S. Block-Lieb and E. J. Janger. The myth of the rational borrower: Rationality, behavioralism, and the misguided reform of bankruptcy law. *Tex. L. Rev.*, 84:1481, 2005.

[15] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[16] J. Cohen, P. Dupas, et al. Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *Quarterly journal of Economics*, 125 (1):1, 2010.

[17] J. Cohen, P. Dupas, and S. Schaner. Price subsidies, diagnostic tests, and targeting of malaria treatment: evidence from a randomized controlled trial. *American Economic Review*, 105(2):609–45, 2015.

[18] M. Cohen, A. N. Elmachtoub, and X. Lei. Pricing with fairness. *Available at SSRN 3459289*, 2019.

[19] E. Creager, D. Madras, T. Pitassi, and R. Zemel. Causal modeling for fairness in dynamical systems. *arXiv preprint arXiv:1909.09141*, 2019.

[20] R. Cummings, N. R. Devanur, Z. Huang, and X. Wang. Algorithmic price discrimination. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2432–2451. SIAM, 2020.

[21] A. D'Amour, H. Srinivasan, J. Atwood, P. Baljekar, D. Sculley, and Y. Halpern. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 525–534, 2020.

[22] B. De Blasio and J. Menin. From cradle to cane: the cost of being a female consumer. a study of gender pricing in new york city. *The New York Department of Consumer Affairs*, 2015.

[23] A. L. Dontchev and R. T. Rockafellar. *Implicit functions and solution mappings*, volume 543. Springer, 2009.

[24] J.-P. Dubé and S. Misra. Personalized pricing and customer welfare. *Available at SSRN 2992257*, 2019.

[25] P. Dupas. What matters (and what does not) in households' decision to invest in malaria prevention? *American Economic Review*, 99(2):224–30, 2009.

[26] P. Dupas. Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica*, 82(1):197–228, 2014.

[27] P. Dworczak, S. D. Kominers, and M. Akbarpour. Redistribution through markets. *Becker Friedman Institute for Research in Economics Working Paper*, (2018-16), 2019.

[28] D. Fernandes, J. G. Lynch Jr, and R. G. Netemeyer. Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8): 1861–1883, 2014.

[29] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. Predictably unequal? the effects of machine learning on credit markets. *The Effects of Machine Learning on Credit Markets (November 6, 2018)*, 2018.

[30] G. Gallego, H. Topaloglu, et al. *Revenue management and pricing analytics*, volume 209. Springer, 2019.

[31] P. Gonzaga. Personalised pricing in the digital era: Background note by the secretariat. Technical report, Directorate for Financial and Enterprise Affairs Competition Committee, 2018.

[32] K. L. Haws and W. O. Bearden. Dynamic pricing and consumer fairness perceptions. *Journal of Consumer Research*, 33(3):304–311, 2006.

[33] H. Heidari, C. Ferrari, K. Gummadi, and A. Krause. Fairness behind a veil of ignorance: A welfare analysis for automated decision making. In *Advances in Neural Information Processing Systems*, pages 1265–1276, 2018.

[34] W. House. Big data and differential pricing. *White House Council of Economic Advisers (https://obamawhitehouse. archives. gov/sites/default/files/whitehouse_files/docs/Big_Data_Re port_Nonembargo_v2. pdf*, 2015.

[35] L. Hu and Y. Chen. Fair classification and social welfare. *arXiv preprint arXiv:1905.00147*, 2019.

[36] G. W. Imbens and D. B. Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[37] K. A. Jacobsen. Rolling back the pink tax: Dim prospects for eliminating gender-based price discrimination in the sale of consumer goods and services. *Cal. WL Rev.*, 54:241, 2017.

[38] T. R. Johnson and M. Lipscomb. Pricing people into the market: Targeting through mechanism design. Dec. 2019.

[39] N. Kallus and A. Zhou. Policy evaluation and optimization with continuous treatments. *arXiv preprint arXiv:1802.06037*, 2018.

[40] D. Karlan and J. Zinman. Long-run price elasticities of demand for credit: evidence from a countrywide field experiment in mexico. *The Review of Economic Studies*, 86(4):1704–1746, 2019.

[41] D. S. Karlan and J. Zinman. Credit elasticities in less-developed economies: Implications for microfinance. *American Economic Review*, 98(3):1040–68, 2008.

[42] M. Kasy and R. Abebe. Fairness, equality, and power in algorithmic decision making. Technical report, Working paper, 2020.

[43] E. H. Kennedy, S. Lorch, and D. S. Small. Robust causal inference with continuous instruments using the local instrumental variable curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(1):121–143, 2019.

[44] M. R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media, 2007.

[45] J. Larson, S. Mattu, and J. Angwin. Unintended consequences of geographic targeting. *Technology Science*, 2015.

[46] Z. Lipton, J. McAuley, and A. Chouldechova. Does mitigating ml's impact disparity require treatment disparity? In *Advances in Neural Information Processing Systems*, pages 8125–8135, 2018.

[47] Z. Liscow. Is efficiency biased? *University of Chicago Law Review*, 2018.

[48] L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt. Delayed impact of fair machine learning. *Proceedings of the 35th International Conference on Machine Learning (ICML), Stockholm, Sweden*, 2018.

[49] F. Luo and S. Mehrotra. Distributionally robust optimization with decision dependent ambiguity sets. *Optimization Letters*, pages 1–30, 2020.
[50] A. A. Miller. What do we worry about when we worry about price discrimination-the law and ethics of using personal information for pricing. *J. Tech. L. & Pol'y*, 19:41, 2014.
[51] A. Odlyzko. Privacy, economics, and price discrimination on the internet. In *Economics of information security*, pages 187–211. Springer, 2004.
[52] S. Qiang and M. Bayati. Dynamic pricing with demand covariates. *Available at SSRN 2765257*, 2016.
[53] D. Slunge. The willingness to pay for vaccination against tick-borne encephalitis and implications for public health policy: evidence from sweden. *PloS one*, 10 (12):e0143875, 2015.
[54] H. Steck, B. Krishnapuram, C. Dehing-Oberije, P. Lambin, and V. C. Raykar. On ranking in survival analysis: Bounds on the concordance index. In *Advances in neural information processing systems*, pages 1209–1216, 2008.
[55] V. Syrgkanis, V. Lei, M. Oprescu, M. Hei, K. Battocchi, and G. Lewis. Machine learning estimation of heterogeneous treatment effects with instruments. In *Advances in Neural Information Processing Systems*, pages 15193–15202, 2019.
[56] K. Vafa, C. Haigh, A. Leung, and N. Yonack. Price discrimination in the princeton review's online sat tutoring service. *Technology Science*, 2015090101, 2015.
[57] J. Valentino-Devries, J. Singer-Vine, and A. Soltani. Websites vary prices, deals based on users' information.
[58] H. R. Varian. Price discrimination and social welfare. *The American Economic Review*, 75(4):870–875, 1985.
[59] H. R. Varian. Price discrimination. *Handbook of industrial organization*, 1:597–654, 1989.
[60] J. D. Wright. Behavioral law and economics, paternalism, and consumer contracts: An empirical perspective. *NYUJL & Liberty*, 2:470, 2006.

Overview of the appendix.

- Appendix A provides a literature review of how considerations regarding price personalization have arisen in different problem settings.
- Appendix B provides proofs of the analysis in Section 5.
- Appendix C provides further details on the empirics.

A LITERATURE REVIEW

A.1 Different problem settings (summary)

Lending. Discrimination in lending is an important problem subject to regulation by the CFPB. Not only do banks decide whether or not to offer loans or extend credit (previously considered as classification problems), they also decide on the terms, i.e. interest rates, of the loan or credit, using risk-based pricing. These rates have significant welfare implications for individuals if discrimination leads to differences in rate terms. Standard discretionary pricing practices, where individual loan agents have some discretion to set rates above/below securitized rates based on local operating costs and competitive factors, may lead to disparate impacts due to, e.g. differential consumer bargaining leverage or discrimination.

There is extensive literature studying household finance, including credit constraints of subprime borrowers with heterogeneity. For example, firms may find via price optimization that they are able to raise prices on subprime borrowers without affecting demand for credit-constrained consumers.

Price optimization in the lending setting is primarily based on discretionary markups/discounts, e.g. negotiated on an individual basis with consumers. There are subsets of loans that are securitized by the government, which are presumed to be actuarially fair prices. Price dispersion beyond this can be explained a combination of individual discretionary pricing (negotiation) or market concentration (price optimization). A common strategy in papers that study potential discriminatory pricing in lending (e.g. [9]) is to regress deviations from the Fannie Mae-securitized rate schedule on controls (e.g. for location-varying costs of lending); the coefficient on race in such a regression corresponds to unexplained disparate impact in pricing.

Insurance. Insurance is the one of the originating domains of risk-assessment. However; it has the least available data. Our analysis correspondingly does not focus on or shed light on important questions in this area.

Consumer products. There are growing concerns regarding the use of price optimization for consumer goods. Many concerns in consumer product pricing are complicated due to the difficulty of disambiguating consumer valuation based on "legitimate" aspects such as individual preferences (which may be culturally conditioned; hence associated with categories) vs. aspects that may seem "illegitimate" or repugnant to extract rent from consumers on the basis of. A different source of concern arises when universal provision of a good is expected.

However, studies show that consumers react strongly negatively to perceptions of price unfairness (perception is an important mediator, because different formats for the same posted price tend to have different fairness perceptions). This consumer backlash may be so strong it is posited as a reason that retailers do not wish to conduct extremely fine first-degree price discrimination. Therefore, we might be interested in general notions of price equity.

Goods with public externalities. In this setting which arises in development, public, or health economics, a centralized decision-maker (DM) has some utility or preference for individuals receiving the good in addition to individuals' idiosyncratic distributions of willingness to pay. Price optimization is beneficial because it can subsidize participation in the market and take-up of the good by pricing at willingness-to-pay for individuals with high valuations to subsidize lower price offers to lower willingness-to-pay [16, 25, 38]. Cross-subsidy is particularly beneficial when understanding willingness-to-pay is in part related to ability to pay, which is commonly discussed in development economics with regards to credit-constrained consumers.

A.2 Implications of price optimization in lending

Lending and implications of regulation. In this setting, such as risk-based pricing in insurance and discretionary pricing in mortgage lending, discretionary pricing is highly regulated and the leeway for discretionary pricing or "price optimization" (rent extraction from consumers) is severely limited. Limited discretion is given to lending agents, for example to match competing offers or possibly (under interpretations of fair lending law) to cover operating costs (for example in certain geographic regions), e.g. "justified business necessity". [9] study mortgage lending for loans bought by Fannie Mae, which are subject to price schedules based on creditworthiness, and find that while financial tech companies that engage in algorithmic pricing reduce discrimination on LatinX and African-American borrowers by 40%, nonetheless bias still persists (as measured by residual coefficients on race in a regression of interest rates on controls).

Nonetheless, concerns about actuarial risk may still arise when actuarial risk is estimated from covariates. [9] summarizes their interpretation of discrimination law and its implications for the legitimacy of personalized pricing:

> (a) Scoring or pricing loans explicitly on credit-risk macro-fundamental variables is legitimate; (b) Scoring or pricing on a proxy variable that only correlates with race or ethnicity through hidden fundamental variables is legitimate; (c) Scoring or pricing on a proxy variable that has significant residual correlation with race or ethnicity after orthogonalizing with respect to hidden fundamental credit-risk variables is illegitimate.

Distributional concerns of interest in insurance/lending and risk-based pricing. In a recent case "Association belge des Consommateurs Test-Achats ASBL and Others v. Conseil des ministres" ([1]), the EU banned the use of sex for pricing in insurance policies. [5] points out some aspects in which the redistributive transfers which result may be preferred to transfers implemented through the typical public economics perspective of tax-and-transfer because the magnitude is pegged to the difference in marginal risk by group; and the transfer does not distort labor incentives as do changes in tax schedules.

[9] points out that regulators may have distributional concerns about price discretion in lending that differ from other settings (such as revenue management and pricing). In revenue management and pricing, and more generally settings with high fixed costs and lower marginal costs, price inelastic customers such as business-class travelers may subsidize access to a service via fare or price discrimination. In contrast, price-sensitive consumers in insurance and lending may be more fiscally savvy and/or "shop around" more for better terms; there is some concern that underprivileged borrowers may also be less equipped to leverage competition as such. This points to the importance of using *non-ideal* theory and domain-specific considerations in motivating considerations of fairness.

Underlying normative considerations barring risk classification based on protected attribute. [5] points out that normative grounds for arguments against risk classification based on protected attribute include luck egalitarian arguments (not holding individuals responsible for factors out of their control, which is difficult with risk to operationalize), as well as the ontological invalidity of social classifications for absolute risk except insofar as social categories designate the effects of inequities in history. Additionally, the consequences of inequity in accuracy of risk classification, arguments regarding causality (which however are more effectively levied along the lines of differential accuracy due to the presence of unobserved confounders) and privacy.

Concerns about price optimization: rent extraction from behavioral bias. A key concern in household finance is in explaining empirical puzzles that contradict predictions from conventional economic theory. To this end, behavioral household finance considers implications of behavioral economics for explaining some of these puzzles.

However, the expansion of the consumer credit market in the 1970s and 1980s led to concerns regarding rising debt, and many puzzles in household finance where empirical phenomena contradicted the assumptions of classical economic theory. [14] studies the "myth of the rational borrower" in the context of the discussion about bankruptcy law. The central controversy was about whether or not the law provides incentives to declare bankruptcy, or if increases in the number of bankruptcies are instead driven by circumstances typically out of a borrower's control such as income shocks. They argue that a key counter-argument is "the myth of the rational borrower"; that behavioral biases from economic theory lead suboptimally rational borrowers to overborrow relative to their actual marginal returns to credit. Whether or not consumers are behaviorally biased, or suboptimal in ways predicted by behavioral economics, is a research topic: [60] counter-argues that empirical evidence is not so strong that deviations from predictions of economic theory are explained by behavioral economics. Behavioral household finance remains an active area of study; although a meta-analysis on the evidence for the effectiveness of financial education interventions suggests that they are overall ineffective [28]. We refer to [12] for a fuller discussion.

We argue that the main takeaways to guide our analysis are therefore:

- Conventional pricing theory and welfare analysis that assumes revealed purchasing behavior reflects valuation may be inappropriate under strong evidence of behaviorally suboptimal consumers, and in particular concerns about credit- and liquidity-constrained customers, especially poorer households.
- Price optimization based on behavioral considerations may be profit-maximizing, but value-based pricing in this setting may be suspect (or may introduce distributional concerns opposite the slight favor in revenue management for price optimization). Competitive considerations may support the utility of price optimization, but we will abstract away from most competitive considerations and focus on implications for an idealized actuarially-fair pricing problem.

A.3 Perceptions of unfairness from consumers due to different prices (equal treatment)

It is unclear whether there is sufficient regulation to merit equal treatment as a normative rule for pricing, or to understand this squarely within the sphere of discrimination law. Some consumer research such as [32] studies consumer perceptions of fairness. Since in fact consumer reaction may differ to personalized pricing based on *salience* – e.g. showing prices vs. showing discounts – the inconsistency of these reactions themselves suggests that intuitive perceptions of fairness in pricing do not reflect rational economic behavior; but rather that behavioral considerations interact with the predictions of standard expected utility frameworks.

There are many high-profile instances where differential pricing appears exploitative but we would argue that what is doing the normative work is more broadly related to concerns regarding perpetuating historical injustices or overt extraction of customer rent for reasons beyond their agency and correlated with historical inequity. In that light, the normative force is really in the "leakage" of structural considerations and the correlation between price optimization and these specific aspects. Conversely, it is difficult to reason abstractly about these settings without further grounded context.

We posit some examples. Consider the razor tax. The razor tax is objectionable because it extracts customer welfare on the grounds of gender roles for functionally the same product. At some level, our objection is far more difficult to levy on the level of *homophily or taste* associated with "pinkness", than it would be to levy on the objection that societal expectations regarding grooming have led to endogenizing higher willingness to pay. Another example is higher prices for shipping/delivery to certain zip codes (or not providing service at all: Here, there is some expectation of equitable service provision, or not reifying historical disadvantages due to location; even though antitrust

provisions would allow charging different prices based on differential costs of doing business. Another important point contrasts price optimization based on first party vs. third-party data. Here, there is a violation of privacy norms. Consumers may object to certain kinds of data being used for price optimization vs. other kinds of data.

Disentangling aspects of valuations driven by the social constructions of race and gender, e.g. via mediation analysis or other counterfactual notions, is likely a very difficult task to operationalize.

A.4 Personalized pricing and public provision

In this specialized setting, price optimization allows significant benefit for cross-subsidizing allocation of a good. Cross-subsidization is particularly beneficial when a central decision-maker has a utility preference for more individuals obtaining the good vs. not; for example in situations with positive externalities.

These pricing schemes are of particular relevance for provision of health interventions in development, where targeting can ensure take-up of households of distributional interest [38]. A natural question is, why not provide the good for free? The answer is due to a combination of scarce resources and efficacy that requires take-up/compliance. In these scarce-resource settings in health and development, if free provision to all were possible and feasible, that would be optimal. However, when resource constraints are binding, and the effects of health interventions are also realized only by "compliance" of the consumer in using it (preventive healthcare, bednets, contraceptives), the DM prefers subsidies to increase access to those who would use the intervention if offered. So there are additional targeting concerns that prefer allocation to individuals who would take-up and use the health intervention; that is, the DM is interested in balancing over-provision and over-exclusion [17]. To this end, price discrimination is considered a tool to screen-in individuals who would use the good, rather than simply allocating for free.

This setting is common in development and public economics. [38] conceive a pricing mechanism to price individuals into a market for mechanical desludging in Kenya; a health intervention that has positive externalities but which may be out of reach for the poorest households. Estimating a demand and reservation price model, they learn personalized prices, estimate the personalized price solution via a multinomial logistic model, and run a RCT to evaluate out-of-sample the effects of personalized pricing on take-up as well as overall health outcomes.

A.5 Research questioning the ontological stability of valuations

Ontological stability of valuations. While a typical economic rejoinder to these concerns highlights that first-degree price discrimination achieves a Pareto-efficient allocation of goods to individuals at their valuation and universal access, we would note the actual distributional implications of personalized pricing depend on how consumer behavior does or does not reflect an implicitly assumed ontologically stable and valid idiosyncratic "valuation".

While there is strong mechanism design theory for soliciting individual valuations in, for example, auctions, we argue that many of the domains where consumer-facing personalized pricing has faced pushback are different from the restricted domains where auctions are deployed in practice. We propose a taxonomy of types of failures of the "ideal theory" of willingness to pay which raise concerns for fairness. These settings are not merely pathological or point failures of economic theory, but rather exactly how structural inequities manifest in price optimization settings. A first concern is *behavioral biases* (and price optimization which extracts rent from them). For example, credit card contracts with promotional offers but high terms afterwards. This motivates the use of mandatory disclosure notices in that setting. A second concern is that of *differential endowments* which surface via different incomes or credit constraints, budget constraints, and liquidity constraints. This is a concern when individuals' reflected purchase behavior is not a realization of their ideal valuation or willingness to pay, of the item, but rather of their *ability* to pay. [47] discusses a similar argument against Kaldor-Hicks efficiency arguments for welfare analysis in empirical legal studies. In settings with credit constraints, individuals are credit or liquidity-constrained, and therefore borrow at high (perhaps even irrationally high) interest rates because of lack of other options. [3]

B PROOFS

B.1 Proofs for price parity

Proof of theorem 1 . **Optimizing over attribute-based personalized prices $p^*(x)$:**

We consider the Lagrangian dual of eq. (2). We then applying Sion's minimax theorem to swap the order of min and max operations, which is valid under compactness [15]:

$$\min_{p(x,a)} \max_{\lambda} \mathbb{E}\left[-pD(p) + \lambda(p\xi(A) - \Gamma)\right] = \max_{\lambda} \min_{p(x)} \mathbb{E}\left[-pD(p) + \lambda(p\xi(A) - \Gamma)\right] \tag{7}$$

For a linear demand model, observing that when $p^*(x,a)$ is unrestricted, we let $p^*(x,a;\lambda)$ be the optimal solution parameterized by λ:

$$p^*(x,a;\lambda) \in \operatorname{argmax} \mathbb{E}[-pD(p) + \lambda(p\xi(A) - \Gamma) \mid X = x, A = a]. \tag{8}$$

The optimal price with the λ penalty is computable in closed form:

$$p^*(x,a;\lambda) = \frac{-\overline{D}(x,a) + \lambda\xi(a)}{2\beta}.$$

Plugging in this solution:

$$\max_{\lambda} L(\lambda, p^*(X,A;\lambda)) = \max_{\lambda} \mathbb{E}\left[-p^*(X,A;\lambda)D(p^*(X,A;\lambda)) + \lambda(p^*(X,A;\lambda)\xi - \Gamma)\right]$$

We maximize the above over λ. Taking derivatives with respect to λ, we obtain the first order necessary condition for optimality, letting $p^*(\lambda) = p^*(X,A;\lambda)$, $\overline{D} = \overline{D}(X,A)$ for brevity is:

$$\begin{aligned}
0 &= \mathbb{E}\left[-p^*(\lambda)\frac{d}{d\lambda}D(p^*(\lambda)) + -(\frac{d}{d\lambda}p^*(\lambda))D(p^*(\lambda)) + (p^*(\lambda)\xi - \Gamma) + \lambda\frac{\xi^2}{2\beta}\right] \\
&= \mathbb{E}\left[-\left(\frac{-\overline{D} + \lambda\xi}{2\beta}\right)\frac{\xi}{2} - \frac{\xi}{2\beta}(\overline{D} + \frac{1}{2}(-\overline{D} + \lambda\xi)) + (p^*(\lambda)\xi - \Gamma) + \lambda\frac{\xi^2}{2\beta}\right] \\
&= \mathbb{E}\left[\frac{\overline{D}\xi}{4\beta} - \frac{\overline{D}\xi}{4\beta} - \lambda\frac{\xi^2}{2\beta} + (p^*(\lambda)\xi - \Gamma) + \lambda\frac{\xi^2}{2\beta}\right] \qquad (9) \\
&= \mathbb{E}\left[\left(\frac{-\overline{D} + \lambda\xi}{2\beta}\xi - \Gamma\right)\right] \qquad (10)
\end{aligned}$$

From eq. (11) we conclude

$$\lambda^* = \frac{\mathbb{E}[\overline{D}(X,A)\xi(A) + 2\beta\Gamma]}{\mathbb{E}[\xi^2(A)]}.$$

Note that further taking the derivative of eq. 11 with respect to λ verifies $L(\lambda, p^*(\lambda))$ is concave in λ, $\frac{d^2L(\lambda,p^*(\lambda))}{d\lambda^2} = \mathbb{E}[\frac{\xi^2}{2\beta}] < 0$ under Asn. 3. Therefore the first-order necessary condition is also sufficient.

Optimizing over attribute-blind personalized prices $p^*(x)$:

The counterpart of eq. (8) is, scaling by a constant $f(x)$, the covariate density of x, to simplify

$$\Delta f(x \mid a) := \mathbb{P}(X = x \mid A = a) - \mathbb{P}(X = x \mid A = b) = \mathbb{E}[\xi(A) \mid X = x]f(x) = (\rho_a^{-1}\mathbb{P}(A = a \mid X = x) - \rho_b^{-1}\mathbb{P}(A = b \mid X = x))g(x)$$

$$\begin{aligned}
& p^*(x;\lambda) \in \operatorname{argmin} -pD(p \mid x) + \lambda(p\mathbb{E}[\xi(A) \mid X = x] - \Gamma) \\
& p^*(x;\lambda) \in \operatorname{argmin} -pD(p \mid x)f(x) + \lambda(p\mathbb{E}[\xi(A) \mid X = x] - \Gamma)f(x) \\
\iff\ & p^*(x;\lambda) \in \operatorname{argmin} -pD(p \mid x)f(x) + \lambda p\Delta f(x \mid a)
\end{aligned}$$

Note $D(p \mid x) = D - \alpha + \mathbb{E}[\beta_A \mid X = x] + \mathbb{E}[g(x,A) \mid x]$.

The optimal $\lambda-$ parametrized price is

$$p^*(x;\lambda) = \frac{-\overline{D}(x) + \lambda\mathbb{E}[\xi(A) \mid X = x]}{2\mathbb{E}[\beta_A \mid X = x]} = \frac{-\overline{D}(x) + \lambda\frac{\Delta f(x|a)}{f(x)}}{2\mathbb{E}[\beta_A \mid X = x]}.$$

We correspondingly solve (analogous to the previous):

$$\max_{\lambda} L(\lambda, p^*(X;\lambda)) = \max_{\lambda} \mathbb{E}\left[-p^*(X;\lambda)D(p^*(X;\lambda)) + \lambda(p^*(X;\lambda)\xi - \Gamma)\right]$$

with the corresponding first-order conditions

$$0 = \mathbb{E}\left[-p^*(\lambda)\frac{d}{d\lambda}D(p^*(\lambda)) + -(\frac{d}{d\lambda}p^*(\lambda))D(p^*(\lambda)) + (p^*(\lambda)\xi - \Gamma) + \lambda\frac{\xi^2}{2\mathbb{E}[\beta_A \mid X = x]}\right]$$
$$= \mathbb{E}\left[\left(\frac{-\overline{D} + \lambda\mathbb{E}[\xi(A) \mid X = x]}{2\mathbb{E}[\beta_A \mid X = x]}\mathbb{E}[\xi(A) \mid X = x] - \Gamma\right)\right] \tag{11}$$

Analogously we conclude

$$\lambda^* = \frac{\mathbb{E}[\overline{D}(X)\mathbb{E}[\xi(A) \mid X] + 2\mathbb{E}[\beta_A \mid X]\Gamma]}{\mathbb{E}[\mathbb{E}[\xi(A) \mid X]^2]}.$$

□

Proof of Proposition 1. We would like to derive conditions that might inform of the sign of $p^*(x) - p^*(x,a)$.
There are a few extreme cases which might be informative (of one regime or another).

(1) $\beta_a = \beta_b, \overline{D}(x,a) = \overline{D}(x,b)$
Disparities are solely due to covariate distributions across groups.
(2) $\beta_a > \beta_b, \overline{D}(x,a) = \overline{D}(x,b) = \overline{D}(x)$
Disparities are solely due to group-level differences in price elasticity or differences in covariate distribution across groups.

We only included conclusions for case 1 in the main text. In this appendix we also provide a sufficient condition for case 2 (though this is less interpretable; hence less useful).

Under **case 1**,

$$p^*(x,a) - p^*(x) = \frac{-\overline{D}(x) + \lambda^*(x,a)\xi(a)}{\beta} - \frac{-\overline{D}(x) + \lambda^*(a)\mathbb{E}[\xi(A) \mid X = x]}{\beta} = \frac{\xi(a)\lambda^*(x,a) - \mathbb{E}[\xi(A) \mid X = x]\lambda^*(x))}{\beta}$$
$$= \frac{\xi(a)\lambda^*(x,a) - \xi(a)\lambda^*(x) + \xi(a)\lambda^*(x) - \mathbb{E}[\xi(A) \mid X = x]\lambda^*(x))}{\beta}$$
$$= \frac{1}{\beta}\left(\xi(a)(\lambda^*(x,a) - \lambda^*(x)) + \lambda^*(x)(\xi(a) - \mathbb{E}[\xi(A) \mid X = x])\right) \tag{12}$$

We conclude the signs of the following terms:

(1) $(\lambda^*(x,a) - \lambda^*(x)) \leq 0$,
by assumption of linearity of demand, Jensen's inequality, since $f(x) = x^2$ is a convex function, and by iterated expectation:

$$\mathbb{E}[\mathbb{E}[\xi(A) \mid X]^2] \leq \mathbb{E}[\mathbb{E}[\xi(A)^2 \mid X]] = \mathbb{E}[\xi(A)^2] \tag{13}$$

Note that $\lambda^*(x,a) \leq \lambda^*(x) \iff \mathbb{E}[\mathbb{E}[\xi(A) \mid X]^2] \leq \mathbb{E}[\xi(A)^2]$.
(2) $\lambda^*(x) < 0$, under assumption for theorem 1 that group a faces the higher unrestricted personalized prices.
(3) $(\xi(a) - \mathbb{E}[\xi(A) \mid X = x]) > 0$ and $(\xi(b) - \mathbb{E}[\xi(A) \mid X = x]) < 0$.
This may be verified by observing

$$\frac{1}{\rho_a} > \frac{\mathbb{P}(A=a|X=x)}{\rho_a} - \frac{\mathbb{P}(A=b|X=x)}{\rho_b} \iff 1 - \mathbb{P}(A = a \mid X = x) > -\mathbb{P}(A = b \mid X = x)\frac{\rho_a}{\rho_b} \iff \rho_b > -\rho_a,$$

and concluding based on nonnegativity of ρ_a, ρ_b.

Therefore, identifying signs of terms in eq. (12):

$$p^*(x,a) - p^*(x) = \underbrace{\frac{1}{\beta}}_{<0}\left(\underbrace{\xi(a)(\lambda^*(x,a) - \lambda^*(x))}_{\leq 0} + \underbrace{\lambda^*(x)}_{\leq 0}\underbrace{(\xi(a) - \mathbb{E}[\xi(A) \mid X = x])}_{\geq 0}\right)$$
$$p^*(x,b) - p^*(x) = \underbrace{\frac{1}{\beta}}_{<0}\left(\underbrace{\xi(b)}_{<0}\underbrace{(\lambda^*(x,a) - \lambda^*(x))}_{\leq 0} + \underbrace{\lambda^*(x)}_{\leq 0}\underbrace{(\xi(b) - \mathbb{E}[\xi(A) \mid X = x])}_{\leq 0}\right) \tag{14}$$

The claim for $p^*(x,a) - p^*(x) < 0$ follows by simplifying and factoring out ρ_a and $\lambda^*(x)/\beta_a > 0$:

$$\begin{aligned} p^*(x,a) - p^*(x) < 0 &\iff \frac{1}{\beta}\left(\rho_a^{-1}(\lambda^*(x,a) - \lambda^*(x)) + \lambda^*(x)(\rho_a^{-1}(1 - (\mathbb{P}(A=a \mid X=x) - \rho_a/\rho_b\mathbb{P}(A=b \mid X=x))\right) < 0 \\ &\iff \left(\frac{\mathbb{E}[\mathbb{E}[\xi(A)|X=x]^2]}{\mathbb{E}[\xi(A)^2]} - 1\right) + ((1 - (\mathbb{P}(A=a \mid X=x) - \rho_a/\rho_b\mathbb{P}(A=b \mid X=x)) < 0 \\ &\iff \frac{\mathbb{E}[\mathbb{E}[\xi(A)|X=x]^2]}{\mathbb{E}[\xi(A)^2]} < (\mathbb{P}(A=a \mid X=x) - \rho_a/\rho_b\mathbb{P}(A=b \mid X=x)) \end{aligned}$$

The claim follows for $p^*(x,b) - p^*(x)$ directly from eq. (14).

To interpret this condition, denote $\Delta f(x;a,b)$ as the *covariate-driven group divergence*:

$$\Delta f(x;a,b) := \mathbb{P}(A=a \mid X=x) - \frac{\rho_a}{\rho_b}\mathbb{P}(A=b \mid X=x) = \frac{\mathbb{E}[\xi(A) \mid X=x]}{\rho_a^{-1}}.$$

Observe that

$$\Delta f(x;a,b) < 0 \iff \frac{\mathbb{P}(A=a \mid X=x)}{\mathbb{P}(A=b \mid X=x)} < \frac{\rho_a}{\rho_b} \iff \frac{\mathbb{P}(X=x \mid A=a)}{\mathbb{P}(X=x \mid A=b} < 1,$$

i.e. the sign of $\Delta f(x;a,b)$ depends on the covariate likelihood under the two classes (and the magnitude depends on the magnitude of this covariate-based divergence relative to the covariate-uninformed ratio, ρ_a/ρ_b).

Under **case 2**:

Let $\Delta\beta_a(x) = \frac{\beta_a}{\mathbb{E}[\beta_A|X=x]}$ and recall that $\frac{\lambda^*_{xa}}{\lambda^*_x} = \frac{\mathbb{E}[\mathbb{E}[\xi(A)|X=x]^2]}{\mathbb{E}[\xi(A)^2]} \le 0$. Adding and subtracting $\frac{-\overline{D}(x)+\xi(a)\lambda^*(x)}{\beta_a}$,

$$\begin{aligned} & p^*(x,a) - p^*(x) > 0 \\ \iff & \left(\frac{-\overline{D}(x) + \lambda^*_{xa}\xi(a)}{\beta_a} - \frac{-\overline{D}(x) + \xi(a)\lambda^*_x}{\beta_a}\right) + \left(\frac{-\overline{D}(x) + \xi(a)\lambda^*_x}{\beta_a} - \frac{-\overline{D}(x) + \mathbb{E}[\xi(A) \mid X=x]\lambda^*_x}{\mathbb{E}[\beta_A \mid X=x]}\right) > 0 \\ \iff & \frac{\xi(a)}{\beta_a}(\lambda^*(x,a) - \lambda^*(x)) + -\overline{D}(x)\left(\frac{1}{\beta_a} - \frac{1}{\mathbb{E}[\beta_A \mid X=x]}\right) + \left(\frac{\xi(a)\lambda^*_{xa}}{\beta_a} - \frac{\mathbb{E}[\xi(A) \mid X=x]\lambda^*_x}{\mathbb{E}[\beta_A \mid X=x]}\right) > 0 \\ \iff & \xi(a)\underbrace{\left(\frac{\mathbb{E}[\mathbb{E}[\xi(A) \mid X=x]^2]}{\mathbb{E}[\xi(A)^2]} - 1\right)}_{<0 \text{ by } eq.\ (13)} + -\frac{\overline{D}(x)}{\lambda^*_x}(1 - \Delta\beta_a(x)) + \left(\xi(a)\frac{\mathbb{E}[\mathbb{E}[\xi(A) \mid X=x]^2]}{\mathbb{E}[\xi(A)^2]} - \mathbb{E}[\xi(A) \mid X=x]\Delta\beta_a(x)\right) > 0 \end{aligned}$$

where in the last line, we factor out $\lambda^*(x)/\beta_a > 0$. Correspondingly for $A = a, b$ respectively:

$$\begin{aligned} p^*(x,a) - p^*(x) > 0 &\iff \underbrace{\rho_a^{-1}\left(\frac{\lambda^*_{xa}}{\lambda^*_x} - 1\right)}_{<0} + -\frac{\overline{D}(x)}{\lambda^*_x}(1 - \Delta\beta_a(x)) + \rho_a^{-1}\left(\frac{\lambda^*_{xa}}{\lambda^*_x} - \Delta f(x;a,b)\cdot\Delta\beta_a(x)\right) > 0 \\ p^*(x,b) - p^*(x) > 0 &\iff \underbrace{-\rho_b^{-1}\left(\frac{\lambda^*_{xa}}{\lambda^*_x} - 1\right)}_{>0} + -\frac{\overline{D}(x)}{\lambda^*_x}(1 - \Delta\beta_b(x)) + \rho_b^{-1}\left(-1\cdot\frac{\lambda^*_{xa}}{\lambda^*_x} - \rho_b\mathbb{E}[\xi(A) \mid X=x]\Delta\beta_b(x)\right) > 0 \end{aligned}$$

Unlike the previous case, this case does not admit determinate conclusions on signs.

We may simplify the condition to obtain that if $-\Delta f(x;a,b)\lambda^*_x + \overline{D}(x)\rho_a > 0$,

$$p^*(x,a) - p^*(x) > 0 \iff \Delta\beta_a(x) < \frac{\lambda^*_x - 2\lambda^*_{xa} + \overline{D}(x)\rho_a}{-\Delta f(x;a,b)\lambda^*_x + \overline{D}(x)\rho_a},$$

with the inequality on $\Delta\beta_a(x)$ holding in the opposite direction if instead $-\Delta f(x;a,b)\lambda^*_x + \overline{D}(x)\rho_a < 0$. □

B.2 Model error fairness

Proof of proposition 2 . We omit dependence on fixed x and denote $D(p) = D(p \mid x)$. Gradient ∇ is with respect to p. We assume price elasticity of demand is nonpositive, $\nabla_p\eta < 0$. We specialize to a revenue setting by observing that $\nabla h = \eta(p) + \nabla\eta \cdot p$, so that $\hat{p}^*, p^*$ satisfy the first order optimality conditions:

$$\hat{p}^* = -\frac{\hat{D}(\hat{p}^*)}{\nabla\hat{D}(\hat{p}^*)}, \qquad p^* = -\frac{\eta(p^*)}{\nabla\eta(p^*)}.$$

Taylor expanding $\hat{D}(\hat{p}^*)$ around p^*:

$$\hat{p}^* - p^* = -\frac{\hat{D}(\hat{p}^*)}{\nabla\hat{D}(\hat{p}^*)} + \frac{D(p^*)}{\nabla D(p^*)} = \frac{\hat{D}(p^*) + \nabla\hat{D}(p^*)(\hat{p}^* - p^*)}{\nabla\hat{D}(\hat{p}^*)} + \frac{D(p^*)}{\nabla D(p^*)}$$

so that

$$\begin{aligned}(\hat{p}^* - p^*)\left(1 - \frac{\nabla\hat{D}(p^*)}{\nabla\hat{D}(\hat{p}^*)}\right) &= -\frac{\hat{D}(p^*)}{\nabla\hat{D}(\hat{p}^*)} + \frac{D(p^*)}{\nabla\hat{D}(\hat{p}^*)}\frac{\nabla\hat{D}(\hat{p}^*)}{\nabla D(p^*)} + o((\hat{p}^* - p^*)^2)\\ &= -\frac{\hat{D}(p^*)}{\nabla\hat{D}(\hat{p}^*)} + \frac{D(p^*)}{\nabla\hat{D}(\hat{p}^*)}\left(1 + \frac{\nabla\hat{D}(\hat{p}^*) - \nabla D(p^*)}{\nabla D(p^*)}\right) + o((\hat{p}^* - p^*)^2)\\ &= \frac{\hat{D}(p^*) - D(p^*)}{\nabla\hat{D}(\hat{p}^*)} + \frac{D(p^*)}{\nabla\hat{D}(\hat{p}^*)}\left(\frac{\nabla\hat{D}(\hat{p}^*) - \nabla D(p^*)}{\nabla D(p^*)}\right) + o((\hat{p}^* - p^*)^2)\end{aligned}$$

Therefore,

$$\begin{aligned}(\hat{p}^* - p^*) &= \left(\frac{\nabla\hat{D}(\hat{p}^*)}{\nabla\hat{D}(\hat{p}^*) - \nabla\hat{D}(p^*)}\right)\left(\frac{\hat{D}(p^*) - D(p^*)}{\nabla\hat{D}(\hat{p}^*)} + \frac{D(p^*)}{\nabla\hat{D}(\hat{p}^*)}\left(\frac{\nabla\hat{D}(\hat{p}^*) - \nabla D(p^*)}{\nabla D(p^*)}\right)\right) + o((\hat{p}^* - p^*)^2)\\ &= (\nabla\hat{D}(\hat{p}^*) - \nabla\hat{D}(p^*))^{-1}\left(\hat{D}(p^*) - D(p^*) + \frac{D(p^*)}{\nabla D(p^*)}\left(\nabla\hat{D}(\hat{p}^*) - \nabla D(p^*)\right)\right) + o((\hat{p}^* - p^*)^2)\\ &\overset{1}{=} (\nabla\hat{D}(\hat{p}^*) - \nabla\hat{D}(p^*))^{-1}\left(\hat{D}(p^*) - D(p^*) + \frac{D(p^*)}{\nabla D(p^*)}\left(\nabla\hat{D}(\hat{p}^*) - \nabla\hat{D}(p^*) + \nabla\hat{D}(p^*) - \nabla D(p^*)\right)\right)..\\ &= \left(\frac{\hat{D}(p^*) - D(p^*)}{\nabla\hat{D}(\hat{p}^*) - \nabla\hat{D}(p^*)} + p^*\left(1 + \frac{\nabla\hat{D}(p^*) - \nabla D(p^*)}{\nabla\hat{D}(\hat{p}^*) - \nabla\hat{D}(p^*)}\right)\right) + o((\hat{p}^* - p^*)^2)\end{aligned}$$

where in 1 we expand $\left(\nabla\hat{D}(\hat{p}^*) - \nabla D(p^*)\right)$; then simplify. □

B.3 Market share

The main tool for analyzing local sensitivities of the optimal prices is an implicit function theorem to differentiate the optimal solution with respect to the parameter. We state it for completeness.

Theorem 3 (Dini Classical Implicit Function Theorem (Thm. 1B.1 of [23]). *Consider a function $f : \mathbb{R}^d \times \mathbb{R}^n \mapsto \mathbb{R}^n$ with values $f(\lambda, p)$ with λ the parameter and p the variable to solve for. The equation $f(\lambda, p)$ is associated with the solution mapping*

$$S\colon \lambda \mapsto \{p \in \mathbb{R}^n \mid f(\lambda, p) = 0\},\ \textit{for } \lambda \in \mathbb{R}^d$$

Let f be continuously differentiable in a neighborhood of $(\overline{\lambda}, \overline{p})$ such that $f(\overline{\lambda}, \overline{p}) = 0$, and let the partial Jacobian of f with respect to p at $(\overline{\lambda}, \overline{p})$, namely $\nabla_x f(\overline{\lambda}, \overline{p})$.

Then the solution mapping S has a single valued localization s around $\overline{\lambda}$ for $\overline{p}$ which is continuously differentiable in a neighborhood Q of $\overline{p}$ with Jacobian satisfying

$$\nabla s(\lambda) = -\nabla_p f(\lambda, s(\lambda))^{-1}\nabla_\lambda f(\lambda, s(\lambda)) \textit{ for every } \lambda \in Q$$

Using the implicit function theorem, we can characterize the sensitivities of solutions under attribute-blind vs. attribute-based, and group market share vs. population market share penalties. We restate an expanded versio of Lemma 1.

Lemma 1[Optimality conditions for different penalties]

The sensitivities of price with respect to λ, $\frac{\partial p^*(x)}{\partial\lambda}$.

(1) $p^*(x)$ with population market share penalty satisfies $\frac{1}{p^*(x)+\lambda} + \frac{D'(p^*(x)|x)}{D(p^*(x)|x)} = 0,\ \ \forall x$ so that

$$\nabla_\lambda p^*(x; 0) = \frac{R''(p^*(x) \mid x))^{-1}}{p^*(x)^2}.$$

(2) $p^*(x, a)$ with population market share satisfies $\frac{1}{p^*(x,a)+\lambda} + \frac{D'(p^*(x,a)|x,a)}{D(p^*(x,a)|x,a)} = 0,\ \ \forall x, a$ so that

$$\nabla_\lambda p^*(x, a; 0) = \frac{R''(p^*(x, a) \mid x, a))^{-1}}{p^*(x, a)^2}.$$

(3) $p^*(x, a)$ with group-level market share $\frac{1}{p^*(x,a)+\lambda_a/\rho_a} + \frac{D'(p^*(x,a)|x,a)}{D(p^*(x,a)|x,a)} = 0$ so that

$$\nabla_\lambda p^*(x, a; 0) = \frac{1}{\rho_a}\frac{R''(p^*(x, a) \mid x, a))^{-1}}{p^*(x, a)^2}.$$

Proof of Lemma 1. (1) $p^*(x,a)$ with population market share

$$\begin{aligned} & p^*(x,a) \in \operatorname{argmax} \mathbb{E}[pD(p)] + \lambda \mathbb{E}[D(p)] \\ \iff & p^*(x,a) \in \operatorname{argmax} \log((p+\lambda)D(p \mid x,a)) \end{aligned}$$

Therefore $p^*(x,a)$ satisfies the following:

$$\frac{1}{p^*(x,a)+\lambda} + \frac{D'(p^*(x,a) \mid x,a)}{D(p^*(x,a) \mid x,a)} = 0, \quad \forall x,a$$

The expression for $\nabla_\lambda p^*(x)(0)$ follows by applying the Implicit function theorem on the optimality condition.

(2) $p^*(x)$ with population market share

$$\begin{aligned} & p^*(x) \in \operatorname{argmax} \mathbb{E}[pD(p)] + \lambda \mathbb{E}[D(p)] \\ \iff & p^*(x) \in \operatorname{argmax} \mathbb{E}[\mathbb{E}[(p+\lambda)D(p) \mid X]], \forall x \\ \iff & p^*(x) \in \operatorname{argmax} (p+\lambda)D(p \mid x), \forall x \\ \iff & p^*(x) \in \operatorname{argmax} \log((p+\lambda)D(p \mid x)), \forall x \end{aligned}$$

Therefore $p^*(x)$ satisfies the following:

$$\frac{1}{p^*(x)+\lambda} + \frac{D'(p^*(x) \mid x)}{D(p^*(x) \mid x)} = 0, \quad \forall x$$

(3) $p^*(x,a)$ with group-level market share

$$\begin{aligned} & \iff p^*(x,a) \in \operatorname{argmax} \mathbb{E}[(p+\lambda_a/\rho_a)D(p) \mid X=x, A=a] \\ & \iff p^*(x,a) \in \operatorname{argmax} (p+\lambda_a/\rho_a)D(p \mid x,a) \\ & \iff p^*(x,a) \in \operatorname{argmax} \log((p+\lambda_a/\rho_a)D(p \mid x,a)) \end{aligned}$$

Therefore $p^*(x,a;\lambda)$ is such that

$$\frac{1}{p+\lambda_a/\rho_a} + \frac{D'(p \mid x,a)}{D(p \mid x,a)} = 0$$

□

B.4 Allocative efficiency: Concordance

Proof of Theorem 2. Let $\mathcal{I}\{A=a\}$ denote index sets for data points within group a, etc.

$$\begin{aligned} \mathbb{P}(D(p_i) < D(p_j) \mid P_a < P_b) &= \frac{1}{|\{(i,j) : p_i < p_j\}|} \frac{1}{n^2} \sum_{i \in \mathcal{I}\{A=a\}} \sum_{j \in \mathcal{I}\{A=b\}} \mathbb{I}[D(p_i)=0, D(p_j)=1, p_i < p_j] \\ &= \frac{1}{|\{(i,j) : p_i < p_j\}|} \frac{1}{n^2} \sum_{i \in \mathcal{I}\{A=a\}} \sum_{j \in \mathcal{I}\{A=b\}} \mathbb{I}[v_i < p_i < p_j < v_j] \\ &\le \frac{1}{|\{(i,j) : p_i < p_j\}|} \frac{1}{n^2} \sum_{i \in \mathcal{I}\{A=a\}} \sum_{j \in \mathcal{I}\{A=b\}} \mathbb{I}[\{v_i < v_j\} \cap \{p_i < p_j\}] \\ &= \mathbb{P}(V_a > V_b \mid P_a < P_b) \end{aligned}$$

where the first equality holds because under Asn. 1 (a.s. monotonicity), the following events are a.s. equivalent:

$$\{p_i < p_j, D(p_i)=0, D(p_j)=1\} \iff v_j > v_i.$$

The second inequality holds because $\{v_i < p_i < p_j < v_j\} \subset \{\{v_i < v_j\} \cap \{p_i < p_j\}\}$.

□

C DATASET DETAILS

Details about the study [53]. We omit concerns about non-response. The survey was distributed online in 2013. $N = 1116$. There are 28 data columns with information including categorical age values, gender, geographic factors and risk factors, and information about knowledge and trust about vaccines.