

# Balancing Profit and Fairness in Risk-Based Pricing Markets

Jesse Thibodeau<sup>1</sup>, Hadi Nekoei<sup>1,2</sup>, Afaf Taïk<sup>1,2</sup>, Janarthanan Rajendran<sup>3</sup>, Golnoosh Farnadi<sup>1,4</sup>

<sup>1</sup>Mila - Québec AI Institute

<sup>2</sup>Université de Montréal

<sup>3</sup>Dalhousie University

<sup>4</sup>McGill University

{jesse.thibodeau, nekoeihe, afaf.taik, farnadig}@mila.quebec  
janarthanan.rajendran@dal.ca

## Abstract

Dynamic, risk-based pricing can systematically exclude vulnerable consumer groups from essential resources such as health insurance and consumer credit. We show that a regulator can realign private incentives with social objectives through a learned, interpretable tax schedule. First, we provide a formal proposition that bounding each firm’s *local* demographic gap implicitly bounds the *global* opt-out disparity, motivating firm-level penalties. Building on this insight we introduce *MarketSim*—an open-source, scalable simulator of heterogeneous consumers and profit-maximizing firms—and train a reinforcement learning (RL) social planner (SP) that selects a bracketed fairness-tax while remaining close to a simple linear prior via an  $\ell_1$  regularizer. The learned policy is thus both transparent and easily interpretable. In two empirically calibrated markets, i.e., U.S. health-insurance and consumer-credit, our planner simultaneously raises demand-fairness by up to 16% relative to unregulated Free Market while outperforming a fixed linear schedule in terms of social welfare without explicit coordination. These results illustrate how AI-assisted regulation can convert a competitive social dilemma into a win-win equilibrium, providing a principled and practical framework for fairness-aware market oversight.

## 1 Introduction

Firms equipped with modern computational power and extensive consumer data logs may reap financial gains by adopting dynamic (or personalized) pricing, which tailors prices to potential customers or customer segments based on their estimated willingness-to-pay. This approach enables firms to extract the greatest economic value from consumer data. From an efficiency perspective, dynamic pricing has been shown to boost firm profitability and accelerate sales speeds (Schlosser and Boissier 2018; Wang et al. 2023). However, its welfare implications are less consistent. In some markets, including insurance and lending, dynamic pricing can yield undesirable distributional outcomes (Zhu et al. 2023; Betancourt et al. 2022). For instance, while health insurers often rely on dynamic pricing, recent census data indicate that members of the Hispanic population in the U.S. are, on average, roughly twice as unlikely to have healthcare coverage as members of the Afro-descendent population, who in turn are twice as unlikely as members of the Caucasian and Asian populations (Martinez 2022; Keisler-Starkey, Bunch, and Lindstrom 2024). Simi-

larly, data reveal a negative correlation between likelihood of coverage and income, and since income and ethnicity are themselves correlated, there are justifiable concerns that healthcare coverage may be systemically biased.

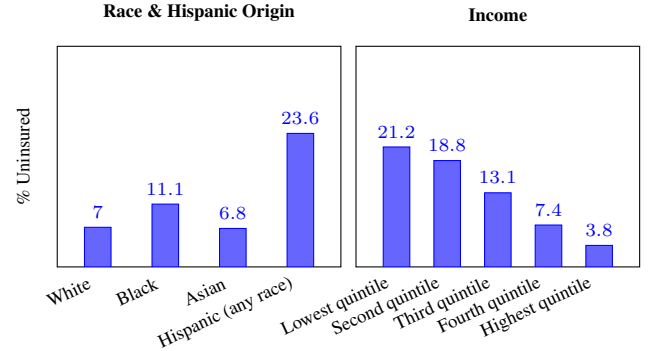


Figure 1: Percentage of working-age adults without health insurance in 2023, by race and income.

Beyond this, substantial price discrepancies between consumers may also give rise to perceptions of unfairness (Lee, Illia, and Lawson-Body 2011), potentially discouraging market participation and perpetuating existing disparities. Moreover, in markets where personal assets can be leveraged to negotiate more favourable terms, goods may become *relatively* more affordable for higher-income consumers. In such scenarios, scarce goods tend to be allocated to privileged groups, leaving fewer units—or lower-quality alternatives—to those with fewer resources. This pattern is well documented in the lending market, where it amplifies wealth gaps (for example, through restricted access to home equity) and drives gentrification. Similar issues arise in sectors such as education services and public transportation, where buyer distributions should ideally reflect those of the underlying population.

Dynamic pricing typically enables firms to set prices for different consumer segments in order to maximize the expected profits derived from each. Thus, a firm adopting dynamic pricing rarely accounts for its resulting buyer distribution, which can be considered unfair if it diverges too sharply from that of the broader population. In this

study, we explore the challenge of dynamic pricing, and specifically that of *regulating its use under demand fairness criteria* in markets where buyer distributions should mirror those of the underlying population. Motivating this, we first demonstrate how profit-maximizing price allocations fail to satisfy demand fairness under purely competitive or collusive dynamics. Broadly, we recognize that it is unrealistic to expect a profit-maximizing firm to voluntarily consider fairness notions in its pricing strategy. Therefore, we consider how a benevolent social planner (SP) might use policy tools such as taxation to encourage market participation among underrepresented consumer groups and penalize unfair firm behavior. To achieve this, we use reinforcement learning to train an SP capable of devising financial incentives that promote demand fairness—namely, by shrinking market opt-out disparities between population subgroups. To conduct our experiments, we introduce `MarketSim`, a simple yet robust and scalable simulation framework wherein arbitrarily many firms engage in price Free Market to capture a market of arbitrarily many consumer profiles with heterogeneous utility. **Our findings indicate that social welfare can be enhanced by taxing firms in ways that incentivize fairer market-specific conduct.** Further, by endowing the learning agent with a domain-specific prior, we observe that interpretability can be maintained at the policy level, confirming the effectiveness of our approach at solving specific markets while retaining certain desirable properties such as tax monotonicity. Contributions of our work include:

- A formal demonstration of how local, firm-level incentives can satisfy a global fairness criterion and thereby raise social welfare in a setting of multiple self-interested stakeholders with incomplete information.
- The introduction of `MarketSim`, a robust and easy-to-use open-source simulator for experimenting with various market dynamics and regulatory policies imposed on arbitrarily many heterogeneous firms and consumers, and evaluating their welfare implications.<sup>1</sup>
- An application of reinforcement learning to generate optimal regulatory policies in two different instances of `MarketSim` (replicating the markets for insurance and consumer credit), showing that we can incentivize welfare-improving firm behaviour in each market, while retaining policy interpretability.

## 2 Related Work

The interdisciplinary nature of this work requires a review of topics from economics, specifically in the subfields of welfare economics and consumer choice theory, as well as a broad overview of applications of artificial intelligence (AI) to welfare economics.

### 2.1 Economics Foundations

In this work, we explore consumer choice and pricing dynamics in competitive markets with heterogeneous agents.

<sup>1</sup>Code will be made publicly available upon acceptance of this work.

On the demand side, consumers exhibit varying sensitivities to price fluctuations, while on the supply side, firms face heterogeneous marginal costs that proxy for technological and scale advantages. Our demand system borrows the random-utility framework of Berry, Levinsohn, and Pakes (1993), yet departs from it in two key respects: first, we replace perfectly rational choice with a stochastic rule that captures bounded rationality and behavioural noise; second, we model multiple competing firms rather than a representative producer, thereby enabling firm-level strategic interaction. These extensions link our analysis to early work on preference heterogeneity by Becker (1962) and its discrete-choice formalisation by McFadden (1972), whose insights remain central to modern consumer-choice theory (Ben-Akiva et al. 2002). Finally, following the process-based welfare view of Fleurbaey (2008), we evaluate outcomes not only by efficiency but also by the fairness of the mechanisms that generate them, an angle largely absent from the original random-utility literature.

### 2.2 Fairness in Dynamic Pricing

The welfare-theoretic study of price design has migrated from economics to operations research (Gallego, Topaloglu et al. 2019) and computer science (Das et al. 2022), giving rise to a rich taxonomy of fairness definitions. Cohen, Elmachoub, and Lei (2022) prove that price, demand, consumer-surplus, and no-purchase fairness cannot be simultaneously satisfied in dynamic settings; we therefore adopt *demand fairness* (Cohen, Elmachoub, and Lei 2022; Kallus and Zhou 2021), which directly measures disparate impact on group participation and is well motivated in education, consumer credit and healthcare domains. Alternative notions such as proportional fairness (Bertsimas, Farias, and Trichakis 2011) highlight welfare trade-offs but do not readily extend to sequential, multi-firm games. RL approaches such as Maestre et al. (2019) impose fairness via Jain’s index under monopoly; in contrast, our regulator shapes *competitive* firms’ incentives so that they voluntarily choose fairer prices, thereby filling the gap between single-seller RL treatments and static constrained-optimisation models.

### 2.3 AI for Economic Policy Generation

The closest application to our work combining economic simulations and sequential modelling is the AI Economist (Zheng et al. 2020), where agents interact in a simulated gather-build-trade society, while a social planner aims to learn an income taxation strategy that improves social welfare, defined as the product of equality and economic productivity. While their work is effective at showcasing emergent behaviours among simulated consumer-workers under incumbent tax regimes, we introduce a new layer to our exploration which instead focuses on how a dynamic regulator can impact societal outcomes by aligning the objectives of self-interested firms with its own. Further, our work focuses on markets involving dynamic pricing, where firms assign prices based on consumer group membership. This added complexity allows for a deeper analysis of firm responses to incumbent policy frameworks. In addition, we

refer to the safe RL literature to impose domain-specific policy constraints in order to maintain a degree of interpretability, which is critical for the widespread adoption of AI-generated public policy. In fact, safe RL methods routinely incorporate domain constraints to ensure an agent’s policy remains feasible or respects regulatory standards (Garcia and Fernández 2015). These constraints can be implicit (e.g., penalizing the agent for violating constraints) (Achiam et al. 2017). By tethering our AI regulator’s reward function to a baseline monotonic schedule, we encourage the learned regulatory policy to align with essential policy norms, namely, that higher fairness should generally not be penalized by higher tax rates, while preserving the safety and interpretability required in real-world economic regulation.

### 3 Preliminaries

We provide an overview of notation and definitions referred to throughout the remainder of this work. Among these, we refer to local and global notions of fairness within this context. Further, we motivate our policy mechanism design by demonstrating how local fairness incentives have global fairness implications. Going forward, let  $A$  be a random variable representing a consumer profile, taking values in the set  $\mathcal{A} = \{1, \dots, m\}$ , and let  $F$  be a random variable denoting the firm selected by a given consumer, taking values in  $\mathcal{F} = \{0, \dots, n\}$ , with  $F = 0$  referring to opting out of the market.

#### 3.1 Definitions and Fairness Metrics

**Definition 1 ( $\epsilon$ -Local Fairness).** For any firm  $j \in \mathcal{F}$ , and for any consumer profile pair  $i, k \in \mathcal{A}$ , we say that firm  $j$  is  $\epsilon$ -locally fair if

$$\max_{i,k} |\Pr(F = j \mid A = i) - \Pr(F = j \mid A = k)| \leq \epsilon.$$

*Remark.* When  $\epsilon = 0$ , it means that every consumer group’s consumption choice is conditionally independent from their group membership.

We quantify market-wide fairness via the opt-out disparity between consumer groups, which we call *global fairness* in order to relate it to its *local* counterpart. This definition draws on *demand fairness*, proposed by Cohen, Miao, and Wang (2021).

**Definition 2 ( $\epsilon$ -Global Fairness).** The entire market is  $\epsilon$ -globally fair if the *opt-out rate* is (approximately) profile-independent:

$$\max_{i,k} |\Pr(F = 0 \mid A = i) - \Pr(F = 0 \mid A = k)| \leq \epsilon.$$

*Remark.* When  $\epsilon = 0$ , every profile opts out with exactly the same probability, i.e.  $\Pr(F = 0)$  is constant in  $A$ . This result is akin to demographic parity (Dwork et al. 2012), where profile  $i \perp$  firm  $j$ .

#### 3.2 Fairness alignment

Local and global fairness capture different levels of discrimination, though both measurements have shortcomings

if considered on their own. On one hand, perfect local fairness fails to capture consumer counts, that is, a firm’s consumers can mirror the population while including very few in total. Similarly, global fairness is satisfied under no market participation, i.e. the case where everyone opts out. Thus, any policy objective involving global fairness should also include some notion of economic productivity. Further, global fairness does not imply local fairness and is not directly addressable. Figure 2 illustrates a market in which *global* fairness holds—both consumer profiles opt out  $\sim 20\%$  of the time—yet the two firms serve very different mixes of profiles ( $F_1$  serving  $A_1$  and  $F_2$  serving  $A_2$ ), violating *local* fairness. How then can an external regulator, unable to

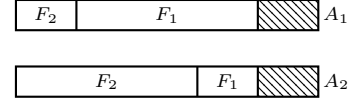


Figure 2: Perfect global fairness does *not* imply local fairness.

address global fairness explicitly, make global fairness improvements by deploying firm-level incentives? This challenge constitutes a mechanism design problem, which we formalize as an optimization problem in the following section.

#### 3.3 From Local to Global Fairness

We first show that *enforcing an  $\epsilon$ -local-fairness constraint on each firm automatically bounds the market-wide opt-out disparity*. The result motivates our policy design: penalties can be assessed at the firm level, yet they control the global metric of interest.

**Proposition 1 (Local  $\Rightarrow$  Global Fairness Bound).** *If every firm  $j \in \{1, \dots, n\}$  satisfies the local-fairness condition*

$$\max_{i,k \in \mathcal{A}} |\Pr(F = j \mid A = i) - \Pr(F = j \mid A = k)| \leq \epsilon,$$

*then the market is  $\epsilon'$ -globally fair, i.e.*

$$|\Pr(F = 0 \mid A = i) - \Pr(F = 0 \mid A = k)| \leq \epsilon' \quad \forall i, k \in \mathcal{A},$$

*with*

$$\epsilon' = \min\{n\epsilon, 1\}.$$

*Proof.* For brevity write  $p_{j|i} := \Pr(F = j \mid A = i)$ . The hypothesis gives  $|p_{j|i} - p_{j|k}| \leq \epsilon$  for every firm  $j \geq 1$  and every pair of profiles  $i, k$ .

**From local to opt-out gap.** Because the opt-out probability is the complement of the in-market mass,

$$p_{0|i} = 1 - \sum_{j=1}^n p_{j|i}, \quad p_{0|k} = 1 - \sum_{j=1}^n p_{j|k}.$$

Hence

$$|p_{0|i} - p_{0|k}| = \left| \sum_{j=1}^n (p_{j|k} - p_{j|i}) \right| \leq \sum_{j=1}^n |p_{j|k} - p_{j|i}| \leq n\epsilon.$$

**Probabilistic range.** Because each  $p_{0|}$  is a probability, their difference cannot exceed 1:  $|p_{0|i} - p_{0|k}| \leq 1$ .

Combining the two bounds,

$$|p_{0|i} - p_{0|k}| \leq \min\{n\epsilon, 1\},$$

which yields the stated  $\epsilon'$ .  $\square$

**Policy insight.** Because global fairness follows directly from firm-level constraints, a regulator can simply penalize firms based on their own  $\epsilon$ -local gap; no market-wide coordination term is needed. This concludes the motivation for our social planner’s policy mechanism. In the following section, we outline the market environments in which policy explorations take place.

## 4 Market Environment

We model an oligopolistic market with heterogeneous consumers and profit-maximizing firms. We then introduce a social planner whose policy consists of a bracketed tax schedule to incentivize fairness. An illustration of this market can be found in Figure 3. We proceed with formal definitions for our simulated agents.

### 4.1 Consumers

Each consumer profile  $i$  obtains utility

$$U_{i,j} = \bar{\alpha} - \beta_i p_{i,j}$$

from consuming firm  $j$ ’s product, where  $\bar{\alpha}_j$  is base product utility,  $\beta_i$  is the price sensitivity of profile  $i$ , and  $p_{i,j}$  is the per-profile price. For the outside option  $F = j = 0$ , let  $U_{i,0} = \bar{\alpha}_0$ . A consumer of profile  $i$  chooses among  $\{0, \dots, n\}$  with probability

$$\mathbf{p}_{j|i} = \frac{\exp(U_{i,j})}{\sum_{j=0}^n \exp(U_{i,j})}.$$

### 4.2 Firms

Each firm  $j \in \{1, \dots, n\}$  may compute its expected profit:

$$\mathbb{E}[\Pi_j] = \sum_{i=1}^m \mathbf{p}_{j|i} (p_{i,j} - mc_{i,j}),$$

where  $mc_{i,j}$  is the average marginal cost for profile  $i$ . Under free-market dynamics, firm  $j$ ’s problem is

$$\max_{\{p_{i,j}\}} \mathbb{E}[\Pi_j] \quad \text{subject to } 0 \leq p_{i,j} \leq p_{\max}.$$

Due to the inherent jump discontinuities in consumer choices under Free Market, firms solve for prices using Powell’s derivative-free method (Powell 1964) in the `SciPy` Python optimization library (Virtanen et al. 2020). While we found this to achieve better stability, we note that alternative optimization methods may be used to solve the firms’ problem.

## 4.3 Social Planner and Welfare Maximization

We now introduce a social planner who aims to maximize overall *social welfare*, a hybrid measure of fairness and firm profits. To maintain policy interpretability, we also penalize large deviations from a simple, *naive* bracketed-tax baseline.

**Bracketed Fairness Tax.** To incentivize firms to adopt fairer outcomes, we partition the fairness range  $[0, 1]$  into  $B$  brackets, each of width  $1/B$ . Suppose firm  $j$  achieves fairness  $f_j \in [0, 1]$  and hence belongs to bracket  $b_j$ , defined by

$$f_j \in \left[ \frac{b_j-1}{B}, \frac{b_j}{B} \right).$$

We index tax brackets by  $b \in \{1, \dots, B\}$  and associate to each bracket a rate  $\tau_b \in [0, 1]$ . In general, we collect these into the vector

$$\boldsymbol{\tau} = [\tau_1, \dots, \tau_B] \in [0, 1]^B.$$

A firm in bracket  $b_j$  thus faces an effective per-profile margin  $(p_{i,j} - mc_{i,j})(1 - \tau_{b_j})$ . Consequently, under regulation, firm  $j$ ’s profit-maximization problem becomes

$$\max_{\{p_{i,j}\}} \mathbb{E}[\Pi_j] (1 - \tau_{b_j}), \quad p_{\min} \leq p_{i,j} \leq p_{\max}.$$

**Social Welfare Objective.** Let  $\mathcal{W}(\boldsymbol{\tau})$  be the total social welfare,

$$\mathcal{W}(\boldsymbol{\tau}) = \left( \frac{1}{n} \sum_{j=1}^n \mathbb{E}[\Pi_j] (1 - \tau_{b_j}) \right) \times \text{fairness}_{\text{global}}(\boldsymbol{\tau}),$$

capturing *both* global fairness (measured via the gap presented in Definition 2) *and* net firm profits (under the chosen  $\boldsymbol{\tau}$ ). We note that this multiplicative welfare expression is one of many possible ways to combine fairness and profit, however the intuition here is that fairness  $\in [0, 1]$  effectively scales profit. A similar formulation of welfare is used in (Zheng et al. 2020). The planner’s goal is to select  $\boldsymbol{\tau}$  to solve  $\max_{\boldsymbol{\tau}} \mathcal{W}(\boldsymbol{\tau})$ . To this end, we use a soft actor-critic algorithm (Haarnoja et al. 2018) to train an RL agent whose reward is  $\mathcal{W}(\boldsymbol{\tau})$ .

In Algorithm 1, we outline the simultaneous Nash Free Market in which firms select prices to optimize for profit given a policy generated by the social planner.

## 5 Market Parameterization and Empirical Results

We construct two market environments, *health insurance* and *consumer lending*, because both combine risk-based pricing with pronounced distributional concerns. Income-group proportions follow Pew Research Center (2024); insurance-coverage rates draw on Keisler-Starkey, Bunch, and Lindstrom (2024); and home-ownership patterns (a proxy for credit demand) follow U.S. Census Bureau (2023). The population is divided into High, Middle, and Low income segments, each assigned a price elasticity ( $\beta$ ) and a firm-specific marginal cost ( $mc$ ). These parameters, which can be found for both markets in table 1, introduce system-wide heterogeneity.

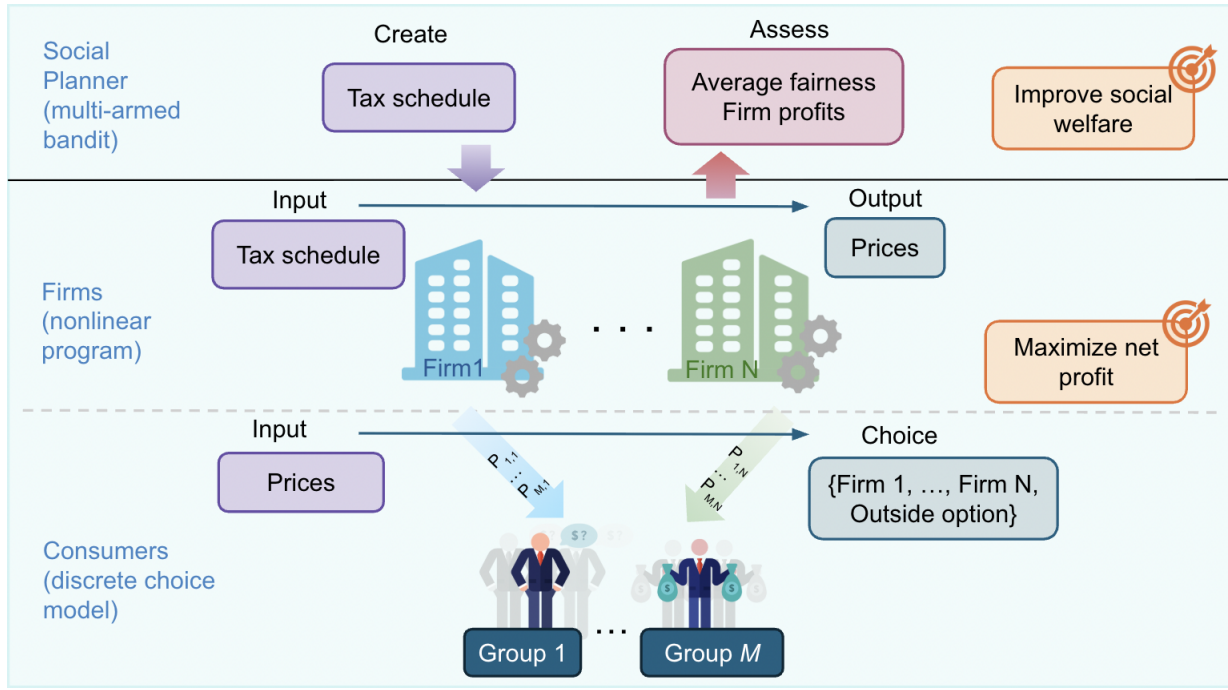


Figure 3: A dynamic-pricing market consisting of 3 agent types, each with their own optimization objective. The social planner generates welfare-maximizing tax schedules applied to firms based on their local fairness gap. Firms then compute their best responses and assign consumer group-level prices. Finally, consumers make their selection from these prices.

**Why heterogeneity matters.** Heterogeneity arises among both agent participant types in the market and shapes every dimension of the policy problem:

- **Consumers.** Differences in disposable income, outside options, and risk exposure create a spread of price elasticities. In insurance, demand is relatively inelastic at higher incomes because coverage quality is valued more than marginal dollars. In credit, by contrast, wealthier households can leverage their assets to negotiate better terms, pay cash or source cheaper capital altogether, making them more price-sensitive. Such a cross-market reversal broadly illustrates how elasticity is a joint outcome of preference intensity and available substitutes.
- **Firms.** Marginal cost is tightly linked to borrowers' or policyholders' income. Low-income consumers are generally riskier to serve as they face higher job volatility, have thinner financial buffers (Bertoletti, Borraz, and Sanroman 2024; Fout et al. 2020), and—in the case of health insurance—experience more occupational hazards and reduced access to preventive care (Nicholson, Bunn, and Costich 2008). These factors translate into (i) higher expected claim costs for insurers and (ii) elevated default probabilities for lenders, raising the average per-unit cost of coverage or credit. By contrast, high-income consumers offer steadier cash flows, better health profiles, and superior collateral, enabling firms to price at lower cost. Even within a single industry, providers differ in their ability to manage this risk—large insurers leverage pooled data and predictive analytics, whereas smaller or niche firms often specialize in higher-risk

pools—amplifying cost dispersion and strategic asymmetry.

From a regulatory standpoint, it is important to consider both sources of heterogeneity in order for policies to achieve realistic social welfare improvements. Our bracketed tax is therefore designed to be *piecewise*—simple enough for transparency yet flexible enough to align marginal incentives across diverse firms. We compare three baselines:

- **Free Market:** Firms compete in a simultaneous price-setting game with the aim of maximizing their individual profits in the absence of policy intervention.
- **Linear regulation:** a monotonic, bracketed linear tax

$$\tau_b^{\text{base}} = 1 - \frac{b}{B}, \quad \text{for } b = 1, \dots, B.$$

intended as a hand-crafted fairness correction. Intuitively, this baseline approximates a simple rule ( $\tau = (1 - \text{fairness})$ ), discretized into  $B$  brackets.

- **Collusion:** Firms jointly maximize aggregate profit. This benchmark reveals the upper bound on total profitability under perfect coordination, and can be seen as an oracle case for profit.

For these benchmarks, we report outcomes after convergence to Nash equilibrium prices.

**Demand elasticity.** Elasticity captures both willingness and ability to substitute. For health insurance, the absence of close substitutes renders wealthier consumers less

---

**Algorithm 1: Multi-Agent Price-Setting Game**


---

```

1: Input:
2:   Number of firms  $n$  (indexed by  $j$ ); consumer profiles
    $i = 1, \dots, m$  with size  $S_i$  and price sensitivity  $\beta_i$ ;
3:   Base utility  $\bar{\alpha}_j$  for each firm  $j$ ; outside option utility
    $\bar{\alpha}_0$ ;
4:   Tax brackets  $\tau = [\tau_1, \dots, \tau_B]$  set by planner;
   marginal costs  $mc_{i,j}$ .
5: Initialize:
6:   Each firm  $j$  has prices  $p_{i,j}$  (possibly random or pre-
   viously set).

7: Step 1: Social Planner Sets Tax Policy
8: for  $j \leftarrow 1$  to  $n$  do
9:   Compute fairness  $f_j$  for firm  $j$ 
10:  Assign bracket  $b_j \leftarrow$  bracket index such that  $f_j \in$ 
    $\left[\frac{b_j-1}{B}, \frac{b_j}{B}\right)$ 
11:  Set effective tax rate  $\tau_{b_j}$  for firm  $j$ 
12: end for
13:   $\triangleright$  Firm  $j$ 's margin becomes  $(p_{i,j} - mc_{i,j})(1 - \tau_{b_j})$ 

14: Step 2: Firms Simultaneously Update Prices
15: for  $j \leftarrow 1$  to  $n$  do
16:   $p_{i,j} \leftarrow \arg \max_{p_{i,j}} \sum_{i=1}^m \mathbf{p}_{j|i} \cdot (p_{i,j} - mc_{i,j})(1 -$ 
    $\tau_{b_j}) \cdot S_i$ 
17:   $\triangleright$  Firm  $j$  chooses prices to maximize expected profit
18: end for

19: Step 3: Consumers Choose Firm or Outside Option
20: for  $i \leftarrow 1$  to  $m$  do
21:   for  $j \leftarrow 1$  to  $n$  do
22:     $U_{i,j} \leftarrow \bar{\alpha}_j - \beta_i p_{i,j}$ 
23:   end for
24:   $U_{i,0} \leftarrow \bar{\alpha}_0$ 
25:   $\mathbf{p}_{j|i} \leftarrow \frac{\exp(U_{i,j})}{\exp(U_{i,0}) + \sum_{k=1}^n \exp(U_{i,k})} \quad \forall j$ 
26: end for

27: Step 4: Outcome and Payoffs
28: for  $j \leftarrow 1$  to  $n$  do
29:  Demand $_{i,j} \leftarrow S_i \cdot \mathbf{p}_{j|i} \quad \forall i$ 
30:   $\Pi_j \leftarrow \sum_{i=1}^m (p_{i,j} - mc_{i,j})(1 - \tau_{b_j}) \cdot \text{Demand}_{i,j}$ 
31: end for

32: Step 5: End of Game
33:  Output final  $\tau$ , prices  $\{p_{i,j}\}$ , demands, and profits
    $\{\Pi_j\}$ .

```

---

price-responsive. In credit, abundant alternatives (e.g. home-equity lines, credit cards, or abstention) make the same group more price-elastic, whereas low-income borrowers confront a near take-it-or-leave-it contract.

**Marginal cost.** Risk-adjusted cost falls with income in both markets but for distinct reasons: fewer costly medical claims in insurance, and lower default probabilities in lending. Firms, heterogeneous in resources and capability, modulate these costs further.

Finally, the social planner learns a piecewise-constant tax via Soft Actor–Critic (Table 2). By explicitly targeting cross-

Group	N	Insurance			Credit					
		$\beta$	$mc_1$	$mc_2$	$\beta$	$mc_1$	$mc_2$	$mc_3$	$mc_4$	$mc_5$
High (H)	200	0.25	2.50	2.25	3.00	0.40	0.65	0.45	0.60	0.44
Middle (M)	520	0.70	3.00	2.75	2.70	1.20	1.45	1.12	1.35	1.29
Low (L)	280	0.825	3.50	3.25	2.25	2.05	2.30	2.25	2.28	2.10

Table 1: Baseline demand elasticities and marginal costs. The insurance market is modelled with two competing insurers, and the credit market with five lenders. Price bounds are  $P_{\min} = 1$ ,  $P_{\max} = 20$ .

Shared social-planner parameters				
Algorithm	Brackets $B$	$\tau_{\min}$	$\tau_{\max}$	$\lambda_{\text{ins}}/\lambda_{\text{cred}}$
SAC	20	0%	100%	100 / 10

Table 2: Social-planner initialization parameters.

segment disparities, the learned policy reflects heterogeneity on *both* the consumer and the firm side, in contrast to the naïve linear schedule.

**$\ell_1$  Penalty on Deviation from Baseline.** To retain policy interpretability, we penalize the *planner* for learning a tax schedule  $\tau$  that deviates excessively from the naïve baseline. Specifically, we add an  $\ell_1$  regularizer

$$\lambda \sum_{b=1}^B \left| \tau_b - \tau_b^{\text{base}} \right|,$$

where  $\lambda \geq 0$  tunes how much the planner is incentivized to remain close to  $\tau^{\text{base}}$ . Thus, the planner tweaks the naïve schedule only to the extent that it increases overall social welfare. This can be interpreted as an “expert planner” capable of taking a known policy mechanism understood by domain experts and adapting it to specific markets. The social planner’s full optimization problem is therefore

$$\max_{\tau} \left[ \mathcal{W}(\tau) - \lambda \sum_{b=1}^B \left| \tau_b - \tau_b^{\text{base}} \right| \right].$$

Hence, while large deviations from the baseline  $\tau^{\text{base}}$  are penalized, the planner retains the flexibility to adjust tax rates to increase overall social welfare (e.g., by nudging firms toward more equitable pricing where the naïve schedule is sub-optimal).

## 5.1 Case Study 1: Health Insurance

**Market Overview and Motivation.** Census data report 7.9% of the U.S. working population uninsured ( $\sim 25\text{M}$ ). Though law-abiding insurance providers do not explicitly use sensitive attributes in the determination of premia, these can often be inferred from occupation and ZIP code (Dwork et al. 2012; Barocas and Selbst 2016), making health insurance an ideal market for dynamic regulation. As mentioned, this instance of our simulation represents a market for a necessity good, and thus we endow low-risk profiles with low demand elasticity, and high-risk profiles with high demand elasticity.

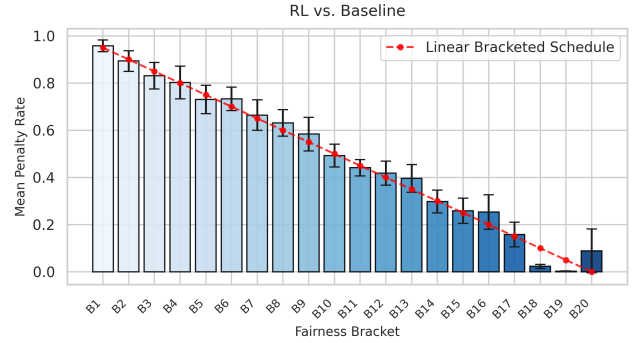


**Policy Generation and Empirical Results.** Use of our method to regulate the health insurance market yields substantial welfare gains over benchmarks. We measure **net profit** (after tax), **fairness**, and **social welfare** at both the per-firm and global level. Table 3 (*left*) shows the outcomes under four market regimes: (**Free Market**) purely competitive, (**Linear-SP**) a simplistic bracket policy with fixed cut-offs, (**RL-SP**) our proposed social planner, and (**Collusion**), a stable cartel-like coordination (used solely for revealing an upper bound on profit from which we provide normalized profit values for the benchmarks of interest). From results, we note that the RL method achieved the highest social welfare, not only by improving fairness, but also in improving total market profitability, indicating that it was able to mitigate the social dilemma inherent to competitive games, instead forcing firms into a kind of policy-induced tacit cooperation that makes the market more stable and friendly to consumers. This is evidence that a fairness-seeking policy can outperform selfish Nash firms in terms of aggregate profits, suggesting that the RL policy might act as an implicit coordination device, getting firms closer to the cooperative outcome without explicit collusion. Broadly, our RL social planner improved social welfare in the market for health insurance compared to the baseline competitive case by approximately 11%, and outperformed the linear policy by 10%. After 2M training steps, the RL social planner converged to the policy found in Figure 4a, with which, as seen in Table 3 (*left*), it was able to funnel firms into welfare-improving fairness brackets. Further, providing the agent with an intuitive prior allowed it to maintain interpretability in brackets with few or no training examples. Specifically, it tweaked the Linear-SP baseline in fairness areas where it was able to achieve substantial welfare gains.

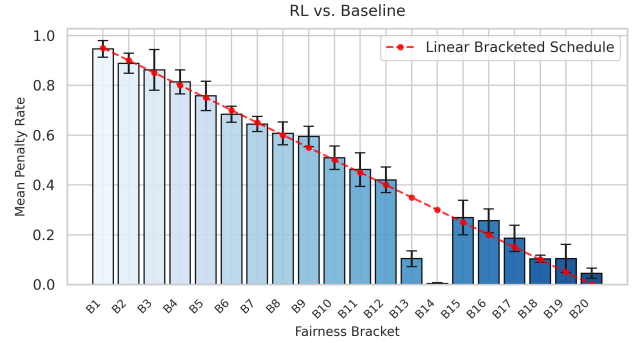
## 5.2 Case Study 2: Consumer Credit

**Market Overview and Motivation.** A second key application of our framework pertains to consumer credit, where financial institutions offer loans or lines of credit to heterogeneous borrowers through competitive interest rates. Unlike health insurance, credit is generally less “essential” yet the stakes remain high for consumers with limited collateral or volatile income. For these groups, restricted credit access can hamper opportunities to secure home equity, reinforcing wealth gaps. As such, credit markets present a core tension between risk-based pricing (necessary for profitable lending) and equitable access (necessary for social welfare and fairness), making them an ideal testing ground for dynamic regulation.

**Policy Generation and Empirical Results.** We evaluate the outcomes of our credit market simulation across the same four market regimes, though in this instance each assessed over five firms. In this market, we observe a qualitatively similar dynamic to the insurance scenario: the RL social planner consistently improves upon linear policy interventions and competitive baselines, even in a setting with greater firm heterogeneity and tighter fairness-profit trade-offs. Evaluated over five competing lenders, Table 3 (*right*) shows that the competitive market yields high profit levels



(a) Taxation policy generated by the RL social planner for the insurance market.



(b) Taxation policy generated by the RL social planner for the credit market.

Figure 4: Comparison of policy generation across two different markets: Insurance (a) and Credit (b).

(0.630) but suffers from low fairness (0.660). The Linear-SP baseline improves fairness to 0.712 but imposes a blunt restriction on risk-based interest rates, thereby driving down profits and yielding modest improvements in participation. By contrast, the RL policy exhibits adaptive behaviour, learning market-specific bracket assignments that improve fairness further (0.767) while also maintaining profitability approaching that of the Free Market baseline. However, we note an increase in the global opt-out rate under the RL policy, a result discussed further in the next section. Nonetheless, overall social welfare increases to 0.477, reversing the decline observed under the Linear-SP baseline (0.392). This represents social welfare gains of 15% and 22% over the competitive and Linear-SP baselines, respectively. Altogether, these results confirm that an RL-based regulator can foster an equilibrium where profitability and fairness minimally conflict.

## 6 Discussion

A notable aspect of our experiments is how tacit cooperation emerges within certain policy frameworks, evidenced by how markets under the RL policy yield comparable or even higher profits than under unregulated Free Market. This is indicative of a social dilemma in unregulated Free Mar-

	Insurance				Credit			
	Free Market	Linear-SP	RL-SP	Collusion	Free Market	Linear-SP	RL-SP	Collusion
<b>Profit</b> ↑	0.697	0.642	$0.707 \pm 0.002$	1.0	0.630	$0.551 \pm 0.003$	$0.622 \pm 0.003$	1.0
<b>Fairness</b> ↑	0.821	0.895	0.895	0.851	0.660	$0.712 \pm 0.006$	$0.767 \pm 0.004$	0.709
<b>Opt Out</b> ↓	0.137	0.120	0.121	0.335	0.173	$0.1593 \pm 0.002$	$0.218 \pm 0.002$	0.395
<b>Welfare</b>	0.572	0.575	<b><math>0.633 \pm 0.002</math></b>	0.851	0.416	$0.392 \pm 0.004$	<b><math>0.477 \pm 0.003</math></b>	0.709

Table 3: Comparison of Profit, Fairness, and Welfare across market scenarios and dynamics. We include market-wide opt-out rates to broadly outline market outcomes. Profit values are normalized with respect to the theoretical maximum determined by the collusive (oracle) market setting. We report standard errors (SE) over 5 seeds, and omit SE values under 0.001.

ket: profit-maximizing individual firm behavior reduces aggregate profits. Meanwhile, by coordinating firm incentives, the regulator resolves this inefficiency, while simultaneously aligning firm and consumer welfare. Crucially, this “cooperation” among firms is driven not by explicit collusion, but instead by an understanding of specific market dynamics acquired by the social planner, allowing it to design incentives which mirror cooperation. From a policy perspective, these insights suggest that partial regulation in the form of adaptive bracket constraints that do not artificially cap or flatten prices can be more attractive than rigid price bounds or unregulated Free Market. This approach preserves profitability for firms while simultaneously broadening access and mitigating discriminatory pricing. In practice, such bracket tuning could be made transparent to both firms and policy-makers, enabling oversight of how fairness categorizations evolve over time.

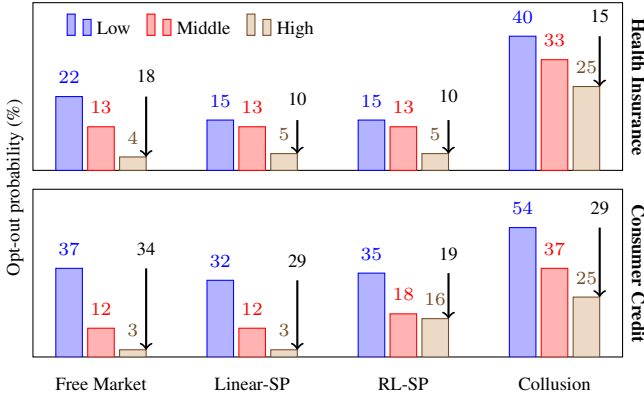


Figure 5: Mean consumer-group opt-out rates under multiple market frameworks. The arrow indicates the maximum difference in opt-out rate means between population groups. By global fairness, lower is better.

Despite obtaining welfare improvements with our RL method, we deem it necessary to highlight a shortcoming regarding our fairness criterion. While demand fairness advocates for equal access to goods (as does *demographic parity*, it is insufficient on its own at ensuring that consumer participation rates improve, as it simply requires that they be agnostic to group membership, with no weight accorded to their actual values. From Table 3, we find that fairness under **collusion**, an illicit market practice in most jurisdictions,

actually outperforms Free Market in both markets, despite yielding the highest opt out rates by a substantial margin. This is not because it truly encourages participation among underrepresented groups, but because it simply makes all consumers more equally *unlikely* to participate. This is evident from Figure 5, where we report group-level opt-out rates. From these, we observe that this share is higher for each subgroup under collusion. We note further that, under the RL policy, there are reductions in opt-out rates among the low-income group in both markets compared to Free Market, albeit at the cost of an opt-out rate increase among the high and middle-income groups. Thus, it is important to examine resulting market distributions in order to qualitatively assess the impact of policy generations. Broadly, while equal access is a desirable goal, it should not come in the form of increased exclusion. A fairness criterion that improves parity in access but leads to a larger number of individuals being priced out or discouraged from participation can ultimately exacerbate structural harms and decrease overall social welfare. To alleviate this concern, one might consider alternate metrics, such as consumer surplus fairness (Cohen, Miao, and Wang 2021), which, when using our probabilistic setting, can be defined via the log-sum rule (Small and Rosen 1981) as

$$|\max_i CS(\tau)_i - \min_k CS(\tau)_k| \leq \epsilon, \forall i, k,$$

where

$$CS(\tau)_i = \frac{1}{\beta_i} \log \left( \sum_j^N \exp(\bar{\alpha}_j - \beta_i p_{i,j}) \right).$$

Accordingly, MarketSim makes it straightforward to swap any welfare metric desired by the user and evaluate outcomes in terms of firm performance and consumer distribution.

### 6.1 Ablation Study: Market Bounds on Fairness

While fairness, according to our local and global definitions, has a theoretical upper bound of 1.0, the empirical upper bound is largely dependent upon market dynamics and parameter initializations. Thus, to uncover the empirical upper bounds on fairness in both markets, we train a new social planner with the sole objective of maximizing global demand fairness with no weight on profit. Results reveal that our instance of the market for insurance has an empirical upper bound on global fairness of 0.895, which was achieved



	Insurance	Credit
Firm A	0.947	0.823
Firm B	0.947	0.885
Firm C	-	0.793
Firm D	-	0.902
Firm E	-	0.763
Global	0.895	0.791

Table 4: Local and global fairness values under a fairness-maximizing SP tax policy.

by both Linear-SP baselines and welfare-maximizing RL policies. Meanwhile, in the credit market, our fairness-maximizing SP reveals an empirical upper bound of 0.791, which was approached but never achieved by our welfare-maximizing SP. This can influence the degree of improvement one can expect when experimenting with regulatory frameworks in *MarketSim*.

## 7 Runtime & Scalability of *MarketSim*

We explore runtimes to demonstrate the approximately-linear scalability of price Free Market in *MarketSim*, reporting wall-clock convergence times on markets of 2 to 100 firms. As the number of participating firms may vary between markets, our simulation should scale well over a broad range of firm counts. For instance, 80% the U.S.

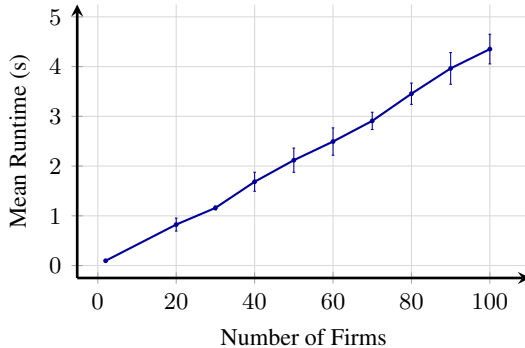


Figure 6: Mean wall-clock runtime per 10 rounds as a function of the number of firms (on CPU). Error bars represent standard deviation over 5 seeds.

health insurance market is dominated by 10 firms, with 57% concentrated in the top 3 (American Medical Association 2023). Meanwhile, the credit market is more fragmented, with many non-bank lenders making up 65% of the mortgage lending market (Chopra 2024). Conversely, 70% of the credit card market is dominated by 5 issuers (McCann 2025), indicating a broader range for participation within lending markets. *MarketSim* nonetheless accommodates arbitrarily many firms and consumers, with runtimes scaling linearly.

**Limitations and Future Work.** While these findings offer promising evidence, we highlight several directions for further inquiry:

1. **Multi-step Taxation Dynamics:** While this work focused on a single-step taxation policy, an exciting direction for future research is to frame taxation as a sequential decision-making task. Such settings could further emphasize the advantages of RL approaches.
2. **Data Quality and Bias:** Real-world transactional datasets often contain inaccuracies, missing fields, or historical biases, which could skew bracket assignments. Investigating robustness under such conditions remains an essential next step.
3. **Bracket Design:** We implement a straightforward bracket structure here. Future work might explore more sophisticated, flexible brackets or alternative reward functions that accommodate multiple fairness definitions or risk preferences.
4. **Legal and Ethical Context:** As bracketed interventions shape competitive behavior at scale, deeper analysis is warranted to confirm compatibility with antitrust laws and to guard against new forms of collusion or bias.

In summary, our case studies demonstrate that dynamic bracket regulation can notably enhance fairness and welfare in a competitive market scenario featuring dynamic pricing with heterogeneous firms and consumer profiles. By aligning private incentives with public interest goals, RL-driven brackets exemplify a viable pathway toward more inclusive market systems for essential products and services.

## 8 Conclusion

In this work, we explore social welfare outcomes under various policy mechanisms in markets featuring risk-based dynamic pricing. We mathematically demonstrate the relationship between local and global notions of fairness, thereby motivating a firm-level policy approach. We then introduce *MarketSim*, a simulator for risk-based dynamic pricing, which we render open-source. Finally, we reproduce two distinct real world markets where global opt-out gaps have been empirically recorded (health insurance and consumer credit) and demonstrate how RL can be successfully leveraged to generate interpretable policy mechanisms aimed at improving social welfare.

## 9 Ethical Statement

If rigorously validated and overseen, AI-assisted regulation could help align private incentives with societal fairness goals in essential-goods markets (e.g. health insurance, consumer credit). Conversely, careless application risks legitimising opaque pricing schemes and deepening existing disparities. We urge future work to prioritise transparency, multidisciplinary oversight, and the voices of affected communities.

## Acknowledgements

Funding support for project activities has been partially provided by Canada CIFAR AI Chair, Google award, and FRQNT scholarships. We also express our gratitude to Compute Canada and Mila clusters for their support in providing facilities for our evaluations.

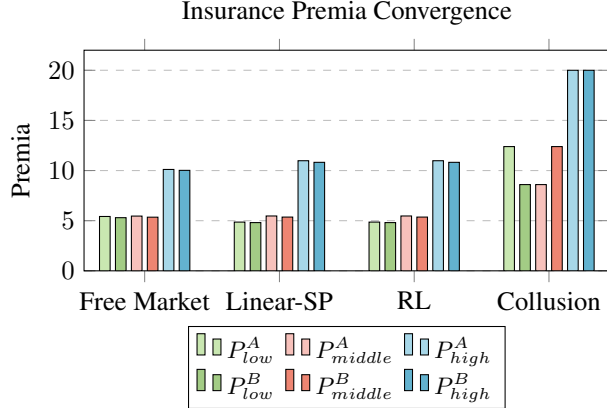
## References

- Achiam, J.; Held, D.; Tamar, A.; and Abbeel, P. 2017. Constrained policy optimization. In *Proceedings of the 34th International Conference on Machine Learning, ICML '17*, 22–31. JMLR.org.
- American Medical Association. 2023. AMA identifies market leaders in health insurance. Accessed: 2025-04-19.
- Barocas, S.; and Selbst, A. D. 2016. Big data’s disparate impact. *California Law Review*, 104(3): 671–732.
- Becker, G. S. 1962. Irrational behavior and economic theory. *Journal of political economy*, 70(1): 1–13.
- Ben-Akiva, M.; McFadden, D.; Train, K.; Walker, J.; Bhat, C.; Bierlaire, M.; Bolduc, D.; Boersch-Supan, A.; Brownstone, D.; Bunch, D. S.; et al. 2002. Hybrid choice models: Progress and challenges. *Marketing Letters*, 13: 163–175.
- Berry, S. T.; Levinsohn, J. A.; and Pakes, A. 1993. Automobile prices in market equilibrium: Part I and II.
- Bertoletti, L.; Borraz, F.; and Sanroman, G. 2024. Consumer Debt and Poverty: the Default Risk Gap. Technical report, GLO Discussion Paper.
- Bertsimas, D.; Farias, V. F.; and Trichakis, N. 2011. The price of fairness. *Operations research*, 59(1): 17–31.
- Betancourt, J. M.; Hortaçsu, A.; Oery, A.; and Williams, K. R. 2022. Dynamic price competition: Theory and evidence from airline markets. Technical report, National Bureau of Economic Research.
- Chopra, R. 2024. Statement by CFPB Director Rohit Chopra on the Financial Stability Oversight Council’s Nonbank Mortgage Company Report. Accessed: 2025-04-21.
- Cohen, M. C.; Elmachtoub, A. N.; and Lei, X. 2022. Price discrimination with fairness constraints. *Management Science*, 68(12): 8536–8552.
- Cohen, M. C.; Miao, S.; and Wang, Y. 2021. Dynamic pricing with fairness constraints. Available at SSRN 3930622.
- Das, S.; Dhamal, S.; Ghalme, G.; Jain, S.; and Gujar, S. 2022. Individual fairness in feature-based pricing for monopoly markets. In *Uncertainty in Artificial Intelligence*, 486–495. PMLR.
- Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, 214–226.
- Fleurbaey, M. 2008. *Fairness, responsibility, and welfare*. OUP Oxford.
- Fout, H.; Li, G.; Palim, M.; and Pan, Y. 2020. Credit risk of low income mortgages. *Regional Science and Urban Economics*, 80: 103390.
- Gallego, G.; Topaloglu, H.; et al. 2019. *Revenue management and pricing analytics*, volume 209. Springer.
- Garcia, J.; and Fernández, F. 2015. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1): 1437–1480.
- Haarnoja, T.; Zhou, A.; Hartikainen, K.; Tucker, G.; Ha, S.; Tan, J.; Kumar, V.; Zhu, H.; Gupta, A.; Abbeel, P.; and Levine, S. 2018. Soft Actor-Critic Algorithms and Applications.
- Kallus, N.; and Zhou, A. 2021. Fairness, welfare, and equity in personalized pricing. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 296–314.
- Keisler-Starkey, K.; Bunch, L. N.; and Lindstrom, R. 2024. Health Insurance Coverage in the United States: 2023.
- Lee, S.; Illia, A.; and Lawson-Body, A. 2011. Perceived price fairness of dynamic pricing. *Industrial Management & Data Systems*.
- Maestre, R.; Duque, J.; Rubio, A.; and Arévalo, J. 2019. Reinforcement learning for fair dynamic pricing. In *Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 1*, 120–135. Springer.
- Martinez, M. E. 2022. QuickStats: Percentage of Uninsured Adults Aged 18–64 Years, by Race and Selected Hispanic Origin Subgroup — National Health Interview Survey, United States, 2020. *MMWR Morb Mortal Wkly Rep* 2022, 71(834).
- McCann, A. 2025. Credit Card Market Share by Issuer. Accessed: 2025-04-19.
- McFadden, D. 1972. Conditional logit analysis of qualitative choice behavior.
- Nicholson, V. J.; Bunn, T. L.; and Costich, J. F. 2008. Disparities in work-related injuries associated with worker compensation coverage status. *American journal of industrial medicine*, 51(6): 393–398.
- Pew Research Center. 2024. Are you in the American middle class? [Accessed: 2025-02-28].
- Powell, M. J. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2): 155–162.
- Schlosser, R.; and Boissier, M. 2018. Dynamic pricing under competition on online marketplaces: A data-driven approach. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 705–714.
- Small, K. A.; and Rosen, H. S. 1981. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, 105–130.
- U.S. Census Bureau. 2023. Homeownership by Race and Ethnicity of Householder. <https://www.census.gov/library/visualizations/interactive/homeownership-by-race-and-ethnicity-of-householder.html>. Interactive visualization, released September 28, 2023. Accessed April 29, 2025.
- Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3): 261–272.
- Wang, Q.; Huang, Y.; Singh, P. V.; and Srinivasan, K. 2023. Algorithms, artificial intelligence and simple rule based pricing. Available at SSRN 4144905.
- Zheng, S.; Trott, A.; Srinivasa, S.; Naik, N.; Gruesbeck, M.; Parkes, D. C.; and Socher, R. 2020. The ai economist: Improving equality and productivity with ai-driven tax policies. *arXiv preprint arXiv:2004.13332*.

Zhu, Y.; Yuan, Y.; Gu, J.; and Fu, Q. 2023. Neoliberalization and inequality: disparities in access to affordable housing in urban Canada 1981–2016. *Housing Studies*, 38(10): 1860–1887.

## A Price Convergence in Insurance Simulation

Firms converge to different price allocations dependent upon the market dynamics. Under fairness incentives (linear and RL policies), firms are nudged to adopt lower prices for the low income group than under Free Market. When firms collude, they set prices much higher to maximize total market profits.



We note that the RL policy leads firms to converge to prices similar to those resulting from the linear policy. This is likely due to these fairness brackets being optimal, as uncovered in Section 6.1. However, the SP is able to improve welfare further by lowering taxes in those brackets, allowing firms to keep more of their profit for achieving optimal fairness.

## B A Probabilistic Bound on Global Fairness

With more assumptions, we derive a tighter, albeit probabilistic bound on opt-out gaps as a result of bounds on local gaps.

**Proposition 2** (Local fairness  $\Rightarrow$  global fairness). *We wish to bound, with high probability, the deviation*

$$\begin{aligned} & \left| \Pr(F = 0 \mid A = i) - \Pr(F = 0 \mid A = k) \right| \\ &= \left| \sum_{j=1}^N \left( \Pr(F = j \mid A = k) - \Pr(F = j \mid A = i) \right) \right| \\ &= \left| \sum_{j=1}^N X_j \right|, \end{aligned}$$

where we set the random variables

$$X_j := \Pr(F = j \mid A = k) - \Pr(F = j \mid A = i), \quad j = 1, \dots, N.$$

**Assumptions.**

1. The variables  $\{X_j\}_{j=1}^N$  are independent.
2. Each  $X_j$  is bounded:  $|X_j| \leq \epsilon$ .
3. Optionally,  $\text{Var}[X_j] \leq \sigma^2$  (needed only for Bernstein).

**a) Hoeffding’s Inequality (bounded differences)** *Because the  $X_j$ ’s are bounded in  $[-\epsilon, \epsilon]$ , Hoeffding’s inequality gives, for a candidate bound  $t$ ,*

$$\Pr\left(\left|\sum_{j=1}^N X_j\right| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{N(2\epsilon)^2}\right).$$

Setting  $t = 2\epsilon \sqrt{\frac{N}{2} \log\left(\frac{2}{\delta}\right)}$  yields, with probability at least  $1 - \delta$ ,

$$\begin{aligned} & \left| \sum_{j=1}^N \left( \Pr(F = j \mid A = k) - \Pr(F = j \mid A = i) \right) \right| \\ & \leq 2\epsilon \sqrt{\frac{N}{2} \log\left(\frac{2}{\delta}\right)}. \end{aligned} \quad (1)$$

**b) Bernstein’s Inequality (variance information available)** *If in addition  $\text{Var}[X_j] \leq \sigma^2$ , Bernstein’s inequality states*

$$\Pr\left(\left|\sum_{j=1}^N X_j\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2N\sigma^2 + \frac{2}{3}\epsilon t}\right),$$

which can be tighter when most differences are very small (i.e.  $\sigma^2 \ll \epsilon^2$ ).

**High-Probability Bound** From (1), we conclude that, with probability at least  $1 - \delta$ ,

$$\begin{aligned} & \left| \Pr(F = 0 \mid A = i) - \Pr(F = 0 \mid A = k) \right| \leq \\ & 2\epsilon \sqrt{\frac{N}{2} \log\left(\frac{2}{\delta}\right)}. \end{aligned}$$

This Hoeffding bound, while not a convergence guarantee, is sufficient to demonstrate the tight relationship between local and global fairness.

## C SAC Hyperparameters

We outline the hyperparameters used with SAC to train our Social Planner agent. They are the same as the default hyperparameters encouraged by Haarnoja et al. (2018).

### SAC Hyperparameters

Hyperparameters	
Hidden-layer sizes (actor/critic)	[256, 256]
Activation function	ReLU
Learning rate $\eta$	0.0003
Batch size	256
Replay-buffer size	1M
Discount factor $\gamma$	0.00 (RL)
Target-update coef. $\tau_{\text{soft}}$	0.005
Entropy coef. $\alpha_{\text{ent}}$	auto
Updates per env step	1
Warm-up steps	100
Total training steps	2M
Random seed	[0, 1, 2, 3, 4]