Inequality-Aware Regulation*

Mikael Mäkimattila†

October 31, 2025

Click here for the latest version

Abstract

I study regulation of a screening monopolist when the regulator cares about the distribution of surplus across heterogeneous consumers and the firm. Consumers have private information about their valuations for quality and the firm about its market demand. The main result shows that optimal regulation combines two policies commonly observed in practice: *inequality-aware pricing* and *cost-plus pricing*. In inequality-aware pricing, the regulator mandates an affordable basic option while granting full flexibility in pricing premium qualities—this targets redistribution within the consumer side. In cost-plus pricing, the firm is required to sell each quality at production cost plus a fixed fee—this limits information rents accruing to the firm. I also compare firm-side regulation to consumer-side subsidies, showing that the latter are weaker along two dimensions: they cannot implement cross-subsidization where the firm would serve some consumers at a loss, and they cannot properly screen the firm's private information.

*I am especially grateful to Matti Liski, Pauli Murto, and Juuso Välimäki for guidance. I am grateful to Georgy Egorov, Alexander Frug, Daniel Hauser, Toomas Hinnosaar, Johannes Hörner, Andreas Kleiner, Johan Lagerlöf, Tuomas Markkula, Ludmila Matysková, Léopold Monjoie, Juan-Pablo Montero, David Myatt, Frikk Nesje, Robin Ng, Jukka Pirttilä, Markus Reisinger, Julia Salmi, François Salanié, Anna Sanktjohanser, Ryo Shirakawa, Peter Norman Sørensen, Egor Starkov, Nick Vikander, Julia Wirtz, and audiences at the University of Copenhagen, at Inequality-aware Market Design Days (Warsaw), at the Nordic Economic Theory workshop 2025 (Copenhagen), and at the Toulouse School of Economics for their insightful comments.

†Aalto University and Helsinki Graduate School of Economics. Email: mikael.makimattila@aalto.fi.

1 Introduction

Market power and distributional considerations are central motivations for regulatory intervention in private markets. These two rationales for intervention often interact: a firm exercising market power typically designs its offering in a way that generates distributional effects both among consumers and across market sides. For example, concentration in the private health care industry—a major policy concern in many countries—has implications for whether the poor receive high-quality care and for the mark-ups the rich pay for their care.

Motivated by these considerations, I study optimal regulation of a monopolist when the regulator cares about distribution of surplus in the market. This is a non-trivial mechanism design problem because both the firm and consumers hold private information relevant for their market behavior and the regulator's objective. I characterize trade-offs arising from the interaction of screening problems on both sides of the market. The distributional objective shapes the optimal allocation of information rents across market participants.

I analyze the following three-stage game. First, the regulator—the *upstream principal*—commits to a regulatory mechanism that specifies transfers to the firm as a function of what the firm offers in the market. Second, the firm—the *downstream principal*—designs a menu of qualities and associated prices. Third, consumers select from that menu. Consumers have privately known willingness to pay (WTP) for quality. Reflecting its superior information about its customer base and demand conditions relative to the regulator, the firm privately observes a demand state that parameterizes the distribution of consumer types. The regulator maximizes a weighted utilitarian objective, placing on average a higher weight on low-WTP consumers than high-WTP consumers, and a higher weight on consumers than the firm.

My main result characterizes optimal inequality-aware regulation, which combines two policies commonly observed in real-world regulation: *inequality-aware pricing* and *cost-plus pricing*. In inequality-aware pricing, the regulator requires the firm to offer a basic-quality option at a below-cost price and grants the firm full flexibility in pricing premium qualities. In cost-plus pricing, the firm is required to offer each type of the good at cost plus a fixed fee. The mechanism assigns the former schedule if the firm's demand state is above a cutoff (so that there are many high-WTP consumers) and the latter if the demand state is below the cutoff.

Inequality-aware pricing targets redistribution within the consumer side. It is justified by consumers' private information: consumption behavior is used to screen the value of redistributing to different consumers. "Poor" consumers with high welfare weights are more likely to select the subsidized basic-quality option. The firm's participation is ensured through markups paid by "rich" consumers on high-quality products. The basic quality consumed by the poor is distorted downward from the efficient level: reducing the quality relaxes the incentive compatibility constraint of the rich, allowing larger monetary redistribution among the consumers. Yet, the basic quality is higher than what the poor would consume in an unregulated monopoly market.

Inequality-aware pricing is widely used in real-world regulation and procurement mechanisms.[1] In health care, it corresponds to requiring private clinics or health care insurers to offer basic-level care or insurance at a low price, cross-subsidized by markups on premium care. Quality in health care could refer to, e.g., scopes of service, waiting times, visit lengths, and staffing levels. A low level of basic-quality mandate is justified by the fact that it enables larger gap between the prices of the basic option and premium options. In some applications, quality maps to quantity: inequality-aware pricing may rationalize charging a higher per-unit price for higher levels of water or electricity consumption. For example in California, regulators have long required privately owned water and electricity utilities to use this kind of pricing scheme.[2]

Since the firm has private information about its demand relative to the regulator, inequality-aware pricing disproportionately benefits a firm facing many consumers with high WTP for quality (e.g., a health care clinic serving a rich clientele). This creates a trade-off between redistributing within the consumer side and across the two sides of the market. The firm's private information thus justifies that optimal regulation combines inequality-aware pricing with cost-plus pricing, where the regulator controls where the firm is required to offer each type of the good at cost plus a fixed fee. Cost-plus regulation is one of the most commonly used types of regulation (Armstrong and Sappington, 2007).

In low-demand states, the regulator's motive to limit redistribution to the

[1] While I refer to the intervention in the paper as "regulation," it can also be interpreted as a single-bidder procurement setting.

[2] Such pricing is often called "increasing block tariffs". For an institutional description and an empirical study on non-linear electricity pricing in California, see Borenstein (2012). For an institutional description and an empirical study on non-linear water pricing in North America, see Olmstead et al. (2007).

firm dominates the motive to redistribute within the consumer side. Implementing inequality-aware pricing in such markets would require giving excessive rents to high-demand firms to satisfy the firm's incentive constraints. The mechanism therefore assigns cost-plus pricing. By contrast, in high-demand states, redistribution among consumers dominates. The cut-off equates the marginal social gain from redistributing within the consumer side with the firm's information-rent cost of doing so. The mechanism is incentive-compatible for the firm because, while choosing cost-plus pricing gives no profit, a low-demand firm lacks sufficient high-WTP consumers opting for premium qualities to recoup losses from selling the basic quality below cost.

The optimal regulation mechanism also screens the average marginal value of transferring one unit of money from the regulator's funds to consumers in the market. When there are many poor consumers, this marginal value is high; the mechanism sets a low overall price level and compensates the firm from public funds. When there are few poor consumers, the mechanism permits a higher price level and reduces transfers to the firm. This screening margin appears when there is a smoothly increasing marginal cost of public funds and disappears if the regulator needs to satisfy budget balance.

The optimal mechanism characterized above requires sophisticated regulation. In many markets, public policy operates through simpler consumer-side instruments: subsidies or taxes tied to a consumer's purchase, such as reimbursements in health care. Such tools may be administratively lighter and sometimes legally more feasible. Motivated by these considerations, I also analyze a variation of the model where the regulator directly contracts with consumers, designing a subsidy schedule that depends only on what a consumer purchases from the firm. The timing mirrors the baseline model: as an upstream principal, the regulator first commits to the subsidy scheme, after which the firm, as the downstream principal, best-responds by offering a menu to consumers. The regulator should anticipate the firm's strategic response when designing the subsidy schedule.

Optimal subsidy design may give either higher or lower subsidies to consumers purchasing expensive high-quality goods relative to those purchasing cheaper low-quality goods. Subsidizing high-quality purchases strengthens the firm's incentives to offer higher qualities, but it conflicts with redistributive objectives. If reducing inequality is sufficiently important to the regulator, the optimal policy assigns a higher subsidy to inexpensive basic-quality options, more often selected by the poor.

While consumer-side subsidies may be a more realistic policy instrument relative to flexible firm-side regulation, contracting with the firm is more powerful along two dimensions. First, the regulator can require the firm to serve poor consumers at a loss and finance that loss with markups on premium purchases. Pure consumer-side subsidies leave the firm a non-negative surplus on each transaction, and the firm typically earns positive profit even absent firm-side private information. Second, regulation of the firm can properly screen the firm's private information by offering state-contingent options. The advantage of consumer-side subsidies is that they can allow redistribution across consumers without rewarding the firm excessively for serving many rich consumers relative to serving many poor consumers.

At a more abstract level, the paper is a novel attempt to model a situation where an upstream principal contracts on a downstream contract, screening both the downstream principal's and downstream agents' private information. The mechanism balances the distribution of information rents between the downstream principal and downstream agents. The comparison of firm-side regulation and consumer-side subsidies highlights how contracting with the downstream principal differs from direct contracting with downstream agents.

Related literature. The regulatory economics literature analyzes settings in which the regulator is less informed than the firm about costs or consumer demand (Baron and Myerson, 1982; Lewis and Sappington, 1988). I focus on the latter information problem but model the consumer side more richly: consumers have heterogeneous, privately known types that are screened through a price-quality menu, and the regulator is concerned with redistribution among consumers.

A small strand of the literature examines regulation of a screening monopolist when the firm has no private information but the regulator's instruments are limited (and the regulator is not concerned with consumer-side inequality). Besanko et al. (1987) consider two policies: a price cap and a quality floor. Krishna (1990) studies the impact of linear taxes. Schlom (2024) examines regulation of a monopolist's price distribution.

Of course, there is also a large literature on redistribution in markets. Public finance has long studied nonlinear taxation, especially in labor markets (e.g., Mirrlees, 1971; Diamond, 1998; Saez, 2001). A recent strand on "inequality-aware market design" analyzes general instruments (e.g., nonlinear taxation and rationing) for addressing inequality in commodity markets (e.g., Akbarpour et al., 2024; Pai and Strack, 2022). However, the literature on redistribution typically abstracts from the supply side by assuming competitive supply or a government-controlled producer; by contrast, the firm's incentive compatibility is central in my model. In a similar vein, da Costa and Maestri (2019) study optimal income taxation in monopsonistic labor markets. Stantcheva (2014) studies optimal income taxation with adverse selection in labor markets, where firms screen workers via compensation contracts. Martimort et al. (2020) and Kang and Watt (2024) study inequality-aware menu design for a government-controlled firm in the presence of an unregulated competitive fringe.

Finally, the paper connects to a scattered literature on dynamic games in which multiple principals contract sequentially. Pavan and Calzolari (2009) provide general characterization results for settings in which multiple principals contract sequentially with the same agent, who is the only party with private information. A recent, innovative application of related ideas is Dworczak and Muir (2024), who analyze the allocation of property rights. A "designer" (upstream principal) chooses an agent's menu of outside options for a subsequent interaction with a "principal" (downstream principal). To connect their analysis with my setting, the designer of their model could be interpreted as a regulator specifying a price cap for each quality. Their framework does not have a participation constraint for the principal that would restrict the designer's policy; also, they study a linear environment and assume the designer weakly prefers transfers from the agent to the principal—opposite to what I assume.

Organization of the article. The paper proceeds as follows. Section 2 presents the environment and the regulator's problem. Sections 3.1 and 3.2 analyze optimal regulation when consumers have a binary type: first in a benchmark with a publicly known demand state, then when the firm privately observes demand. Section 3.3 extends the analysis to a continuous distribution of consumers' WTP. Section 4 studies consumer-side subsidies as an alternative instrument. Proofs are collected in the Appendix.

2 Set-up

Market environment. I study a canonical monopoly screening framework (see, e.g., Mussa and Rosen, 1978; Maskin and Riley, 1984) with a single firm serving a unit mass of consumers. The profit-maximizing firm can produce a good of quality $q \in Q = \mathbb{R}_+$ at cost $c(q)$, where $c : Q \to \mathbb{R}_+$ is continuously differentiable, strictly increasing, and convex with $c(0) = 0$. The firm offers a menu of quality–price pairs to consumers. A consumer with willingness-to-pay (WTP) type θ who consumes quality q at price p enjoys utility

$$\theta v(q) - p,$$

where $v : Q \to \mathbb{R}_+$ is continuously differentiable, strictly increasing, and strictly concave with $v(0) = 0$, $\lim_{q\to\infty} v'(q) = 0$, and $\lim_{q\to 0} v'(q) = \infty$. In some applications, it is natural to interpret q as quantity rather than quality.

Consumers' types. Each consumer is characterized by a privately known WTP parameter $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$ and a binary unobservable label $i \in \{R, P\}$ indicating whether the consumer is "rich" or "poor". Let F_R and F_P denote the distributions of θ among rich and poor consumers, respectively, with F_R strictly first-order stochastically dominating F_P: $F_R(\theta) < F_P(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta})$. Label i does not matter directly for consumers' preferences over quality and consequently, its observability to consumers is not important.

Demand states. The firm privately observes a demand state $\gamma \in \Gamma = [0, 1]$ which is the share of rich consumers in the market. The state is drawn from distribution G with differentiable density g. Conditional on label i, WTP is independent of the state, so that when the realized share is γ, the market WTP distribution is

$$F(\theta \mid \gamma) = \gamma\, F_R(\theta) + (1 - \gamma)\, F_P(\theta).$$

Because F_R strictly first-order stochastically dominates F_P, for $\gamma' > \gamma$, $F(\cdot \mid \gamma')$ strictly first-order stochastically dominates $F(\cdot \mid \gamma)$. Higher γ therefore corresponds to stronger demand: more rich consumers with a high willingness to pay for quality.

Regulatory game. The timeline of the regulatory game is shown in Figure 1. First, the firm privately observes γ, and consumers privately observe their WTP types θ. Next, the regulator publicly commits to a transfer rule $\hat{t} : \mathcal{M} \to \mathbb{R}$ that maps any menu $M \in \mathcal{M}$ of quality–price options (q, p) to a transfer from the regulator to the firm.[3] The firm—the downstream principal—then either

[3] A posted menu is any subset $M \subseteq Q \times \mathbb{R}$ of quality–price pairs. Although the transfer rule $\hat{t}$ is, in principle, defined on the large domain of all menus, $\mathcal{M} = \{M : M \subseteq Q \times \mathbb{R}\}$, it would also suffice for the regulator to define it on a simpler "sufficient statistic" of the menu such as the set of options on the menu that are undominated for consumers. Section 3 shows that even

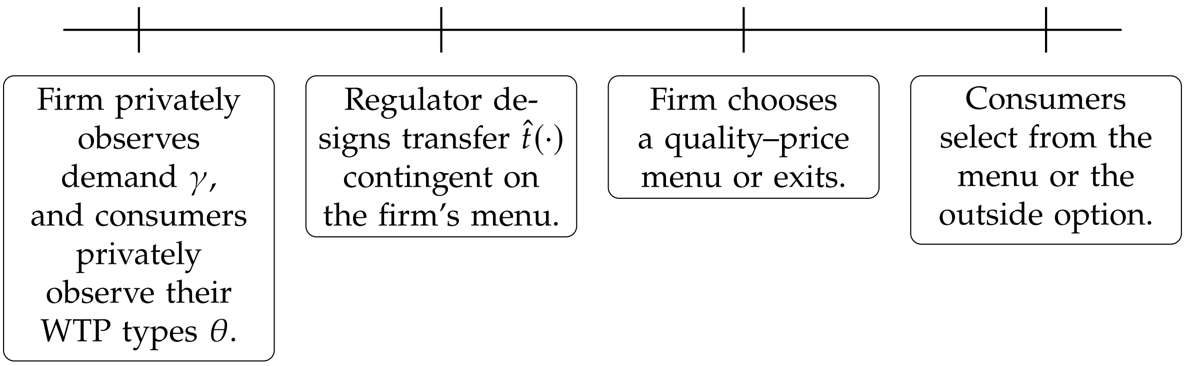

Figure 1: Model timeline.

chooses its menu or exits (to obtain earnings zero). Finally, consumers choose from the posted menu or take the outside option with payoff zero. I focus on regulator-optimal perfect Bayesian equilibria of this game.

As is standard in the models of regulation with private demand information (see Armstrong and Sappington, 2007), the regulator cannot condition transfers on realized sales (here, the measure of units sold at each option). Policies that avoid monitoring realized demand are easier to implement and not susceptible to demand manipulation. If realized demand were contractible, the problem would reduce to the case with no firm-side informational asymmetry (see Section 3.1).

By a taxation–principle argument, the framework is equivalent to a direct–mechanism formulation in which the regulator commits ex ante to a mapping from the firm's message to a posted menu and a regulatory transfer, and the firm then reports a message rather than explicitly choosing a menu. This leaves the set of implementable allocations and the incentive constraints unchanged.

Consumers' incentives. Let $q : \Theta \times \Gamma \to \mathbb{R}_+$ and $p : \Theta \times \Gamma \to \mathbb{R}$ denote the equilibrium qualities and prices assigned to consumers. In state γ, the equilibrium payoff of a consumer with WTP θ is then

$$u(\theta, \gamma; q(\theta, \gamma), p(\theta, \gamma)) = \theta\, v(q(\theta, \gamma)) - p(\theta, \gamma).$$

Individual rationality requires that for all (θ, γ),

$$u(\theta, \gamma; q(\theta, \gamma), p(\theta, \gamma)) \geq 0. \tag{IR-B}$$

Incentive compatibility requires that, given the firm's offering in any fixed demand state $\gamma \in \Gamma$, every consumer weakly prefers her equilibrium choice of

simpler transfer rules sometimes suffice.

quality–price pair to any choice made by some other consumer. So for all $\theta \in \Theta$ and $\gamma \in \Gamma$,

$$\theta \in \arg\max_{\theta' \in \Theta} u(\theta, \gamma; q(\theta', \gamma), p(\theta', \gamma)). \tag{IC-B}$$

Firm's incentives. Let $t : \Gamma \to \mathbb{R}$ denote the equilibrium transfer from the regulator to the firm. Given demand state γ, the firm's equilibrium profit is

$$\Pi\big(\gamma; q(\cdot, \gamma), p(\cdot, \gamma), t(\gamma)\big) = \int_{\Theta} \big(p(\theta, \gamma) - c(q(\theta, \gamma))\big)\, dF(\theta \mid \gamma) + t(\gamma).$$

The firm's individual rationality requires that it makes non-negative profit in every demand state $\gamma \in \Gamma$,

$$\Pi\big(\gamma; q(\cdot, \gamma), p(\cdot, \gamma), t(\gamma)\big) \geq 0. \tag{IR-F}$$

Furthermore, the firm's incentive compatibility requires

$$\gamma \in \arg\max_{\gamma' \in \Gamma} \Pi(\gamma; q(\cdot, \gamma'), p(\cdot, \gamma'), t(\gamma')), \tag{IC-F}$$

so that the firm with demand γ cannot profitably deviate to the menu designed for demand state γ'; the full details can be found in the proof of Lemma 1.

Lemma 1. *The regulator can implement the triple $q : \Theta \times \Gamma \to \mathbb{R}_+$, $p : \Theta \times \Gamma \to \mathbb{R}$, $t : \Gamma \to \mathbb{R}$ with some regulation $\hat{t}$ if and only if the constraints* (IR-B), (IC-B), (IR-F) *and* (IC-F) *are satisfied.*

Regulator's objective. The regulator maximizes a weighted utilitarian objective in which welfare weights measure the regulator's value of giving a unit of money to each market participant. A consumer's weight depends on whether she is rich or poor, $\omega_i \in \{\omega_P, \omega_R\}$ with $\omega_P > \omega_R$.[4] The regulator also places weight $\omega_F \in \mathbb{R}_{++}$ on firm profits, with $\omega_F \leq \omega_R$. A natural special case is $\omega_F = \omega_R$: the firm is owned by the rich and the mechanism trades off limiting rents accruing to the rich via ownership or consumption against losses in total surplus.

[4]Part of the literature on mechanism design and redistribution explicitly models heterogeneous values for money and assumes a designer maximizes a utilitarian objective where each consumer's weight is 1, so that the consumer's value-for-money parameter effectively becomes her welfare weight. See, e.g., Dworczak et al. (2021); the approach dates back at least to Weitzman (1977). A similar approach could be used here as well. Another approach to modeling redistributive concerns, often taken in public finance literature, is to apply a concave transformation to individual utilities to obtain the individual's social contribution; the weights can be interpreted to arise from a local approximation of such a welfare function.

The cost of public funds is $k(t)$, where k is continuously differentiable, strictly increasing and convex with $\lim_{t\to\infty} k'(t) = \infty$, $k'(0) \leq \omega_R$. A hard budget constraint—i.e., a fixed cap on the regulator's transfer—arises as a limiting case of admissible k (let the cost of public funds become arbitrarily large at the cap); in this sense, k generalizes a hard budget constraint.

The regulator's expected payoff is

$$\int_\Gamma \int_\Theta \Big(\mathbb{E}[\omega_i \mid \theta, \gamma] u(\theta, \gamma) + \omega_F \, \Pi(\gamma) - k\big(t(\gamma)\big)\Big) \, dF(\theta \mid \gamma) dG(\gamma) \tag{1}$$

$$= \int_\Gamma \left[\gamma \omega_R \int_\Theta u(\theta, \gamma) dF_R(\theta) + (1-\gamma)\omega_P \int_\Theta u(\theta, \gamma) dF_P(\theta) + \omega_F \Pi(\gamma) - k\big(t(\gamma)\big)\right] dG(\gamma).$$

By Lemma 1, an optimal regulation mechanism (i.e., an equilibrium of the game) is then a triple (q, p, t) maximizing (1) subject to constraints (IR-B), (IC-B), (IR-F) and (IC-F). Any such optimum is constrained Pareto-efficient: there is no other regulation that would make all consumers and the firm weakly better off and someone strictly better off without increasing the regulator's expected cost.

Discussion on the assumptions. This paper considers a partial equilibrium framework for tractability and transparency: a single market is modeled, and welfare weights are exogenous. The model therefore abstracts from other redistributive instruments, such as income taxation. One rationale is political feasibility: policymakers regulating a specific market may lack authority or political capital in other policy domains. Another rationale is that governments may have market-specific redistributive tastes (Tobin, 1970, calls this "specific egalitarianism"). Recent work incorporating labor and consumption choices in competitive settings with flexible instruments (e.g., Doligalski et al., 2025; Ahlvik et al., 2024) shows that even when optimal income taxation is available, it is still typically optimal to distort commodity markets because of redistributive concerns. Nevertheless, those instruments are optimally designed jointly across markets, so focusing on a single market entails a loss along that dimension.

While the framework assumes a monopolist (typical of industries such as electricity, water, and postal services), it can also apply in markets where product differentiation or search frictions create effectively captive demand (e.g., isolated health care clinics). For screening under oligopolistic competition, see Rochet and Stole (2002).

3 Optimal inequality-aware regulation

3.1 Binary WTP type and known demand state

Preliminary analysis. I begin with a benchmark in which the demand state $\gamma \in (0,1)$ is publicly known.[5] A consumer's willingness to pay is deterministic conditional on her label $i \in \{R, P\}$: rich consumers ($i = R$) have $\theta = \theta_R$ and poor consumers ($i = P$) have $\theta = \theta_P$ with $\theta_R > \theta_P$. Thus there is a perfect negative correlation between θ and the welfare weight ω_i. Throughout this subsection, I write

$$q_i := q(\theta_i, \gamma), \quad p_i := p(\theta_i, \gamma) \quad \text{for } i \in \{P, R\}.$$

With known γ, the firm has no private information, so the firm's incentive compatibility (IC-F) does not have to be considered. The regulator must only ensure that the firm makes non-negative profit. Given that the regulator values redistribution from the firm to consumers, as $\mathbb{E}[\omega_i \mid \gamma] = \gamma\omega_R + (1-\gamma)\omega_P > \omega_R \geq \omega_F$, the firm makes zero profit at the optimum,

$$\gamma\big(p_R - c(q_R)\big) + (1-\gamma)\big(p_P - c(q_P)\big) + t \;=\; 0. \tag{2}$$

If the firm earned strictly positive profit, the regulator would benefit from requiring the firm to reduce both prices slightly, which would redistribute from the firm to consumers without violating any constraints.

Furthermore, given that the regulator values redistribution from the rich to the poor ($\omega_P > \omega_R$), she wants to increase price gap $p_R - p_P$ subject to incentive compatibility. The rich consumers' IC constraint therefore binds at the optimum,

$$\theta_R v(q_R) - p_R \;=\; \theta_R v(q_P) - p_P; \tag{3}$$

otherwise the regulator would benefit from requiring the firm to slightly increase p_R and decrease p_P, keeping feasibility.

Let $S_i(q) := \theta_i v(q) - c(q)$ denote money-metric surplus at quality q for type $i \in \{P, R\}$. Using the two binding constraints (2) and (3) to eliminate prices (p_P, p_R) from the planner's objective, the regulator's problem is equivalent to choosing (q_P, q_R, t) to maximize

$$\mathbb{E}[\omega_i \mid \gamma]\Big(\overbrace{\gamma S_R(q_R) + (1-\gamma)S_P(q_P)}^{\text{unweighted surplus}} + t\Big) \tag{4}$$

[5] The analysis is also very similar if the state γ is unknown to both the regulator and the firm with a common prior.

$$- \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big) \underbrace{\gamma\big(S_R(q_P) - S_P(q_P)\big)}_{\text{rich's info rents}} - k(t).$$

The first line aggregates total gains from trade and the regulator's transfer, all weighted by the average consumer weight $\mathbb{E}[\omega_i \mid \gamma]$ as if these could be rebated to all consumers as lump-sum payments. However, rich consumers receive information rents that are subtracted on the second line, weighted by the difference between the average consumer weight and the rich consumers' weight. The second line also accounts for the cost of public funds $k(t)$.

Optimal regulation with known γ.

Proposition 1. *In the binary-WTP environment with known $\gamma \in (0,1)$, there is an optimal regulation mechanism with the following properties:*

(i) Transfer to the firm satisfies $k'(t) = \mathbb{E}[\omega_i \mid \gamma]$.

(ii) The rich consume the efficient quality, uniquely defined by $q_R = \arg\max_q S_R(q)$.

(iii) The poor face a downward distortion from the efficient quality. If $\gamma \frac{\omega_P - \omega_R}{\omega_P} < \frac{\theta_P}{\theta_R}$ then $q_P > 0$ is the unique solution to

$$\mathbb{E}[\omega_i \mid \gamma](1-\gamma)S_P'(q_P) = (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(S_R'(q_P) - S_P'(q_P)),$$

and otherwise $q_P = 0$.

(iv) The prices are pinned down by two binding constraints: the firm's IR constraint (2) and the rich consumers' IC constraint (3).

Parts (ii)-(iv) of the proposition formalize inequality-aware pricing. The rich consume the efficient quality, while the poor face a downward distortion that relaxes the rich consumers' IC constraint (3) and enables lower prices for the poor. The higher the value of redistributing from the rich to the poor, the lower the quality allocated to the poor.[6] When redistributive taste is sufficiently strong, the poor may optimally be allocated the minimal quality $q_P = 0$.

The regulated prices are pinned down by the binding firm's IR constraint (2) and the binding rich consumers' IC constraint (3):

$$p_P = c(q_P) - t - \gamma\big[S_R(q_R) - S_R(q_P)\big], \tag{5}$$

$$p_R = c(q_R) - t + (1-\gamma)\big[S_R(q_R) - S_R(q_P)\big]. \tag{6}$$

[6]The comparative statics are discussed in greater detail in Section 3.2.

Comparison with efficient outcome. That inequality-awareness results in a downward distortion of the poor's quality from the efficient level may seem paradoxical. However, the optimal regulation utilizes the property that rich consumers value marginal changes in quality more, so lowering the poor's quality leads beneficially softens the rich's incentive to choose that option. The distortion stems from the unobservability of consumer types; if consumers' types were not private, the regulator could redistribute without respecting the rich consumers' IC constraint, and there would be no reason to distort any quality in the market.

The downward distortion depends on the utilitarian social welfare specification. The regulator's objective results in policies that are *inequality-aware* but not *inequality-averse*, at least in terms of dispersion in consumed quality. The utilitarian specification is not paternalistic: it respects the preferences of the poor, who prefer to take a small downward distortion in quality given that it allows a substantial decrease in the price they must pay. In particular, the optimal regulation mechanism is always constrained Pareto-efficient, as noted in Section 2. By contrast, an alternative regulatory objective, such as an attempt to allocate quality equally regardless of the willingness to pay, might well lead to Pareto-inefficient policies that do not similarly respect the poor's preferences.

Comparison with unregulated market. There is an interesting similarity between the inequality-aware regulation and unregulated profit-maximizing monopoly screening. The unregulated monopoly would also set the rich consumers' IC constraint to bind. Furthermore, it would make the poor's IR constraint bind, and hence its profit could be written as a function of the allocation as

$$\overbrace{\gamma S_R(q_R) + (1-\gamma) S_P(q_P)}^{\text{unweighted surplus}} \quad - \overbrace{\gamma \big(S_R(q_P) - S_P(q_P)\big)}^{\text{rich's info rents}}. \tag{7}$$

Comparing (4) with (7), both the inequality-aware regulator and the unregulated monopolist dislike rich consumers' information rents. However, the regulator discounts them relatively less because $\omega_R > 0$: unlike the firm, the regulator has a positive value for giving information rents to the rich. Hence, both use downward distortion of the low option to relax the rich consumers' IC, but the regulator distorts less:

Remark 1. *In the optimal regulation mechanism of Proposition 1, the poor receive a weakly higher quality than under unregulated profit-maximizing monopoly pricing.*

Two differences relative to laissez-faire therefore emerge: (a) in terms of quality, optimal regulation increases the poor's quality, and (b) in terms of prices, optimal regulation lowers the price level so that the firm breaks even.

Simple implementation. The result of Remark 1 is also important for showing that the regulator does not have to monitor the firm's entire menu:

Remark 2. *The optimum in Proposition 1 admits a simple implementation: it suffices to mandate that the firm includes one option* $(\underline{q}, \overline{p})$ *in its menu (and receives the transfer* t*), where* $(\underline{q}, \overline{p}, t) = (q_P, p_P, t)$ *from the optimal mechanism. No further restrictions on other qualities or prices are required.*

The regulator can therefore implement the optimal mechanism by simply making transfer t to the firm contingent on the firm including a single option $(\underline{q}, \overline{p})$ in its menu. Given the mandated option $(\underline{q}, \overline{p})$, the firm optimizes the allocation subject to the constraint that both the poor and the rich obtain at least the payoff they receive from the mandated option, solving the following problem:

$$\max_{q_P, p_P, q_R, p_R} \quad \gamma\big(p_R - c(q_R)\big) + (1-\gamma)\big(p_P - c(q_P)\big)$$

$$\text{s.t.} \quad \theta_i v(q_i) - p_i \;\geq\; \theta_i v(\underline{q}) - \overline{p}, \text{ for } i \in \{P, R\}, \qquad \text{(IR vs mandate)}$$

$$\theta_i v(q_i) - p_i \;\geq\; \theta_i v(q_j) - p_j, \text{ for } i, j \in \{P, R\}, \qquad \text{(IC)}$$

Given that the regulator sets $(\underline{q}, \overline{p})$ to be the poor's assigned option in Proposition 1, the firm solves the maximization problem by allocating the efficient quality to the rich and quality $\underline{q}$ to the poor, choosing prices so that the poor pay price $\overline{p}$ and the IC constraint of the rich binds. The allocation of the optimal regulation mechanism is therefore reproduced. Offering the poor (i) lower quality at a lower price or (ii) higher quality at a higher price is not profit-maximizing: (i) moves the poor further from their efficient quality without reducing the rich's information rents so long as $(\underline{q}, \overline{p})$ remains available; (ii) is unprofitable because $\underline{q}$ already exceeds the laissez-faire poor quality by Remark 1.

3.2 Binary WTP type and unknown demand state

Preliminary analysis. I now let the firm privately know the share of rich consumers γ in its market demand. The state γ is continuously distributed with full support on $\Gamma = [0, 1]$ according to a cdf G with density $g > 0$. All other primitives are as in Section 3.1: consumers are rich or poor, and willingness to pay is fixed within each group—θ_R for the rich and θ_P for the poor, with $\theta_R > \theta_P$.

The firm's incentive compatibility across states becomes central. The profit in state γ is

$$\Pi(\gamma) = \gamma\big[p_R(\gamma) - c(q_R(\gamma))\big] + (1-\gamma)\big[p_P(\gamma) - c(q_P(\gamma))\big] + t(\gamma),$$

where equilibrium qualities, prices and the transfer are now expressed as a function of the demand state γ. The realization γ then determines the firm's preferences over the following *mark-up difference*:

$$r(\gamma) = \Big(p_R(\gamma) - c\big(q_R(\gamma)\big)\Big) - \Big(p_P(\gamma) - c\big(q_P(\gamma)\big)\Big).$$

The mark-up difference $r(\gamma)$ captures the extra mark-up the firm extracts from a rich relative to a poor consumer. Intuitively, the more there are rich consumers in the market (i.e. the higher γ), the greater the firm's value for a high mark-up difference.

A standard Myersonian argument then yields:

Lemma 2. *In the binary-WTP framework, the firm's incentive compatibility* (IC-F) *is satisfied if and only if for all $\gamma \in \Gamma$, the envelope condition*

$$\Pi(\gamma) = \Pi(\underline{\gamma}) + \int_{\underline{\gamma}}^{\gamma} r(\gamma')d\gamma'. \tag{8}$$

holds and $r(\cdot)$ is non-decreasing on Γ.

Unless $r \equiv 0$, the envelope (8) implies that the firm earns positive information rents at some states and the regulator cannot force $\Pi(\gamma) = 0$ for all γ as in Section 3.1. Eliminating all information rents would require $r(\gamma) \equiv 0$, corresponding to *cost-plus pricing* in every state, i.e., each quality priced at cost plus (possibly) a common fixed fee.

The mark-up difference $r(\gamma)$ also closely relates to the rich consumers' rent:

$$u(\theta_R, \gamma) - u(\theta_P, \gamma) = S_R(q_R(\gamma)) - S_P(q_P(\gamma)) - r(\gamma). \tag{9}$$

Thus, increasing the mark-up difference $r(\gamma)$ (holding the qualities fixed) decreases the rich consumers' rent while increasing the rent of the firm in high states. This creates a potential trade-off for regulation.

Using (8) and (9), the regulator's payoff can be written in terms of the quality allocation, the regulatory transfer t and the mark-up difference r.

Lemma 3. *In the binary-WTP framework, the regulator's expected payoff is*

$$\int_{\underline{\gamma}}^{\overline{\gamma}} \Big\{ \underbrace{\mathbb{E}[\omega_i \mid \gamma]\big[\gamma S_R(q_R(\gamma)) + (1-\gamma) S_P(q_P(\gamma))}_{\text{avg. consumer weight} \times \text{total surplus}} + t(\gamma)\big]$$

$$- \underbrace{\big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\, \gamma \big[S_R(q_R(\gamma)) - S_P(q_P(\gamma)) - r(\gamma)\big]}_{\text{within-consumer redistributive motive} \times \text{rich rent}} \tag{10}$$

$$- \underbrace{\big(\mathbb{E}[\omega_i \mid \gamma] - \omega_F\big)\Pi(\underline{\gamma})}_{\substack{\text{across-side redistributive motive} \\ \times \text{firm's base rent}}} - \underbrace{\big(\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F\big)\frac{1-G(\gamma)}{g(\gamma)}\, r(\gamma)}_{\text{across-side redistributive motive} \times \text{firm's info rent}} - k\big(t(\gamma)\big)\Big\}\, dG(\gamma).$$

The first line shows the gains from trade in the market and the transfer to the firm, weighted by the consumers' average welfare weight. The second line subtracts the rich consumers' rent, weighted by gap $\mathbb{E}[\omega_i \mid \gamma] - \omega_R$ that reflects the redistributive motive within the consumer side. In contrast to the benchmark where the firm had no private information (analyzed in Section 3.1), it is not clear that the optimal regulation minimizes the rich consumers' rent by making their IC constraint bind; this would imply a high mark-up difference $r(\gamma)$, leading to information rents to the firm. Those information rents are subtracted on the third line and weighted by difference $\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F$, reflecting the redistributive motive across the two sides of the market.

Optimal regulation. The following theorem shows that an optimal regulation mechanism has a cut-off at which cost-plus pricing is switched to inequality-aware pricing. The cut-off is a zero of function δ defined as

$$\delta(\gamma) := \gamma(1-\gamma)(\omega_P - \omega_R) - (\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F)\frac{1-G(\gamma)}{g(\gamma)}. \tag{11}$$

The function δ is the derivative of the integrand in (10) with respect to the mark-up difference $r(\gamma)$. The first term in the formula for δ captures the benefit of increasing the mark-up difference in terms of consumer-side redistribution. The second term reflects the cost of increasing the mark-up difference in terms of the rent captured by the firm in high states γ.

Theorem 1. *In the binary-WTP framework, there exists an optimal regulation mechanism with a cut-off $\gamma^* > \underline{\gamma}$ satisfying $\delta(\gamma^*) = 0$, such that*

(i) Transfers satisfy $k'(t(\gamma)) = \mathbb{E}[\omega_i \mid \gamma]$ for all $\gamma \in \Gamma$.

(ii) Cost-plus pricing is implemented for all $\gamma \leq \gamma^*$:

$$q_R(\gamma) = \arg\max_q S_R(q), \qquad q_P(\gamma) = \arg\max_q S_P(q);$$

$$p_R(\gamma) = c(q_R(\gamma)) - t(\gamma), \quad p_P(\gamma) = c(q_P(\gamma)) - t(\gamma).$$

(iii) Inequality-aware pricing is implemented for all $\gamma > \gamma^*$:

$$q_R(\gamma) = \arg\max_q S_R(q), \quad q_P(\gamma) \leq \arg\max_q S_P(q);$$

$q_P(\gamma)$ is decreasing—if locally strictly decreasing[7], it satisfies

$$\mathbb{E}[\omega_i \mid \gamma](1-\gamma)S_P'(q_P(\gamma)) - \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\gamma\big(S_R'(q_P(\gamma)) - S_P'(q_P(\gamma))\big)$$
$$+ \big(\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F\big)\frac{1-G(\gamma)}{g(\gamma)}S_R'(q_P(\gamma)) = 0;$$

the rich consumers' IC constraint binds, which together with (8) fixes p_R and p_P.

The main part of the proof shows that some mechanism where (i) transfers are set according to $k'(t(\gamma)) = \mathbb{E}[\omega_i \mid \gamma]$, (ii) cost-plus pricing is implemented for all $\gamma \leq \gamma^*$, and (iii) inequality-aware pricing is implemented for all $\gamma > \gamma^*$ improves upon any mechanism that does not satisfy these properties on some positive-measure interval. The existence of an optimal mechanism follows from compactness of the set of mechanisms satisfying these properties and the continuity of the objective.

Part (i) of the theorem shows how the optimal regulation mechanism screens the average value of allocating money to consumers in the market: the transfer satisfies $k'(t(\gamma)) = \mathbb{E}[\omega_i \mid \gamma]$. The firm's incentives can be affected either through direct payments $t(\gamma)$ or through the overall price level, and the trade-off between these instruments depends on the consumers' average weight in the market. When many consumers are poor, the regulator optimally subsidizes them by requiring low prices and compensates the firm for this with high direct transfers $t(\gamma)$. When many consumers are rich, high prices are allowed and the firm's transfer is reduced.

Part (ii) of the theorem shows that for low demand realizations $\gamma \leq \gamma^*$, the regulator's incentive to limit redistribution from consumers to the firm dominates the incentive to redistribute within the consumer side; inequality-aware pricing would give large rents to the firm. Cost-plus pricing is therefore implemented: the firm's IR constraint binds, the rich consumers' IC constraint

[7]The decreasing function q_P is said to be strictly decreasing at γ if $q_P(\gamma+\varepsilon) - q_P(\gamma-\varepsilon) < 0$ for all $\varepsilon > 0$.

does not bind, and the quality allocation is efficient.

Part (iii) of the theorem shows that for high demand realizations $\gamma > \gamma^*$, the motive to redistribute on the consumer side dominates. Inequality-aware pricing is implemented: the firm gets positive rents, the rich consumers' IC constraint binds, and low quality $q_P(\gamma)$ is distorted downward to extract relatively more from rich consumers. Yet, the presence of the firm's private information weakens the quality distortion relative to Section 3.1.

At the cut-off, $\delta(\gamma^*) = 0$: the marginal social gain from redistributing within the consumer side is equated with the associated firm's information rent cost. Theorem 1 does not claim that there would be a unique value of γ such that $\delta(\gamma) = 0$, or even that there would exist a uniquely optimal cut-off. The characterization of the theorem does not depend on that. However, if function δ—which is negative for small γ and satisfies $\delta(\overline{\gamma}) = 0$—happens to be strictly single-crossing on $(0, 1)$[8], then there is a unique interior γ with $\delta(\gamma) = 0$, and the cutoff is optimally set at that crossing. It is also possible that $\delta(\gamma) \leq 0$ for all $\gamma \in [0, 1]$, in which case it is optimal to always implement cost-plus pricing.

It may feel problematic that cost-plus pricing is implemented instead of inequality-aware pricing exactly when many consumers are poor and there should be an elevated concern about the poor. However, in such a market, within-market inequality is not high, and therefore within-market redistribution among consumers is not important. The optimal regulation mechanism takes the elevated concern for the poor into account by raising the regulator's spending to guarantee a low overall price level.

Comparison with unregulated market. The optimal quality schedules of Theorem 1 are illustrated in Figure 2, together with the efficient and laissez-faire (unregulated profit-maximizing monopoly screening) benchmarks. Both the regulator and the laissez-faire monopolist always allocate the efficient quality to the rich. The laissez-faire monopolist always distorts the poor's quality downward to extract surplus from rich consumers. Optimal regulation does not distort the poor's quality in the cost-plus region $\gamma < \gamma^*$. In inequality-aware pricing region $\gamma \geq \gamma^*$, optimal regulation distorts the poor's quality downward from the efficient level but raises the quality relative to the laissez-faire benchmark.

[8] That is, $\delta(\gamma') \geq 0 \Rightarrow \delta(\gamma'') > 0$ for all $\gamma'' > \gamma'$.

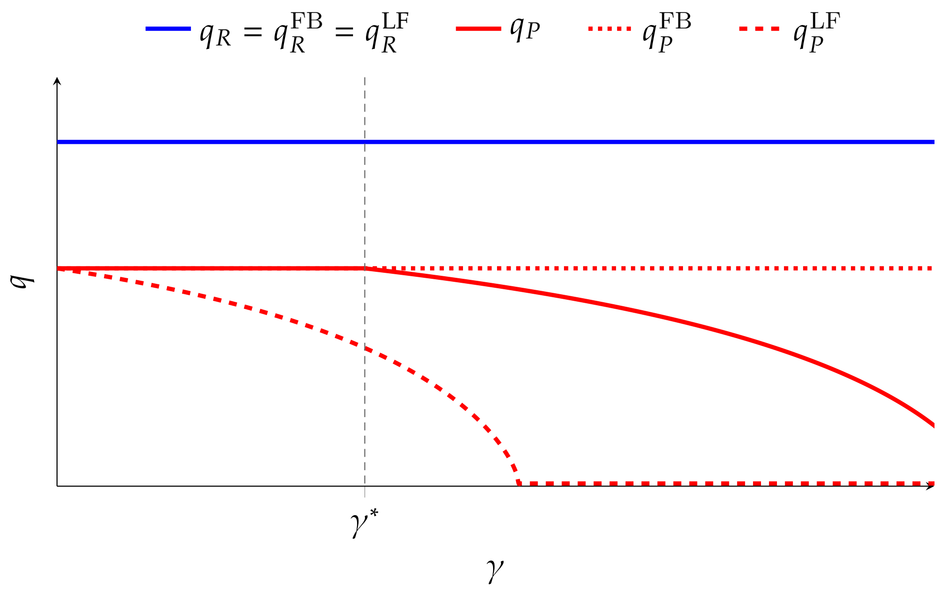

Figure 2: Qualities (q_R, q_P) in the optimal regulation mechanism when G is uniform on $[0, 1]$, $\theta_P = 1$, $\theta_R = 2$, $\omega_P = 1$, $\omega_R = 1/2$, $\omega_F = 1/2$, $v(q) = \sqrt{q}$, and $c(q) = q^2$. Benchmarks: efficient allocation q^{FB} and laissez-faire allocation q^{LF} for the poor.

Remark 3. *Relative to the laissez-faire benchmark, optimal regulation always weakly increases the poor's quality.*

Simple implementation. Analogous to Section 3.1, implementing inequality-aware pricing does not require directly controlling the whole menu that the firm offers in the market.

Remark 4. *Inequality-aware pricing options admit a simple implementation: for each γ in the region $\gamma > \gamma^*$, it suffices to mandate that the firm includes the option $(q_P(\gamma), p_P(\gamma))$ in its menu and receives the transfer $t(\gamma)$. No further menu restrictions are required.*

The proof of Remark 4 is slightly more involved than that of corresponding Remark 2 in the known-γ case. This is because the proof must also rule out profitable "double deviations" under the simple implementation, in which the firm would select the option intended for type γ' and then reallocate quality in a manner inconsistent with the optimal regulation for γ'. In fact, *conditional on* choosing the option intended for high γ', a low-γ firm may wish to allocate higher quality than $q_P(\gamma')$ to the poor; nonetheless, the proof shows that this double deviation as a whole is unprofitable for the low-γ firm, ensuring that the simple implementation gives the intended outcome of Theorem 1. The result

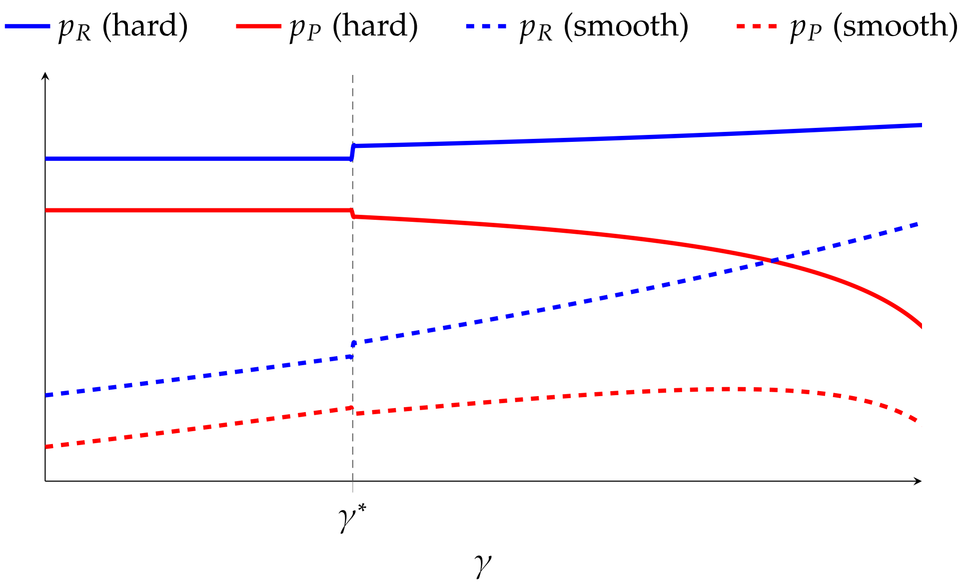

Figure 3: Prices under the optimal regulation mechanism in the parametric example of Figure 2, under a hard budget constraint 0 (solid curves) and under a smoothly increasing marginal cost of public funds, $k(t) = (1/3)e^t$ (dashed curves).

of Remark 4 is desired because it also implies that the regulator does not have to restrict the firm from offering "side options" in addition to those required by the regulator.

Delegation-type regulation. In many regulatory environments, the regulator is unable to make transfers to the regulated firm; for example, the regulator may have no access to public funds or may face legal constraints. In such settings, regulation takes the form of *delegation*: the regulator provides the firm with a set of allowed menus among which the firm chooses, without any associated transfers between the regulator and the firm.[9]

Delegation-type regulation is optimal in the current framework if the regulator faces a hard budget constraint 0 without explicit cost of public funds. The hard budget constraint can be represented by function k that explodes for $t > 0$, which arises as a limiting case of admissible k. Under the hard budget constraint, Theorem 1 continues to apply, with part (i) replaced by $t(\gamma) = 0$

[9]Canonical models in regulatory economics typically allow transfers between the regulator and the firm, while real-world regulatory policies such as quality standards and price caps often do not involve such payments, as noted by Armstrong and Sappington (2007). The seminal work in the theory of delegation is by Holmström (1978); for a more recent, influential contribution with an application to monopoly regulation, see Alonso and Matouschek (2008).

for all γ. Parts (ii)–(iii) still characterize the quality allocation; in pricing, cost-based pricing (i.e., cost-plus pricing with no "plus") is applied for $\gamma \leq \gamma^*$, and inequality-aware pricing is applied for $\gamma > \gamma^*$.

The firm then chooses between pricing at cost and subsidizing the low-quality option to charge a mark-up for a premium-quality good. When γ is small, expected demand for the premium quality is too low to recoup losses on the basic option, so the firm self-selects into cost-based pricing. When γ is large, the firm instead provides the basic quality at a below-cost price to gain flexibility in premium pricing.

The regulator's budget flexibility—captured by the smoothness of k—therefore affects price levels but not the quality allocation. Under delegation, the price paid by the poor for basic quality, $p_P(\gamma)$, is (weakly) decreasing in γ, since inequality-aware pricing is chosen only when high demand for high quality makes cross-subsidization attractive. As k becomes smoother, the regulator can compress the overall price level in markets with many poor consumers, and the monotonicity of p_P consequently fails, as in Figure 3, making $p_P(\gamma)$ increase on some subintervals even as the price gap remains increasing:

Remark 5. *In the optimal regulation mechanism of Theorem 1:*

(i) The price gap $p_R(\gamma) - p_P(\gamma)$ is weakly increasing in γ on $[\underline{\gamma}, \overline{\gamma}]$.

(ii) Under a hard budget constraint (corresponding to delegation-type regulation), $p_P(\cdot)$ is weakly decreasing on $[\underline{\gamma}, \overline{\gamma}]$.

Weights and the distribution of rent. The optimal regulation and the implied distribution of rent depend critically on the weight structure.

Remark 6. *Fix an optimal regulation mechanism of Theorem 1 with cutoff γ^*.*

(i) Holding $\mathbb{E}[\omega_i]$ fixed, a larger (smaller) spread $\omega_P - \omega_R$ weakly lowers (raises) the cutoff in every optimal mechanism.

(ii) A larger (smaller) ω_F weakly lowers (raises) the cutoff in every optimal mechanism.

Part (i) of the remark shows that strengthening the consumer-side redistributive motive makes the regulator implement inequality-aware pricing more often (as the cut-off decreases): it both increases the social gain from redistributing

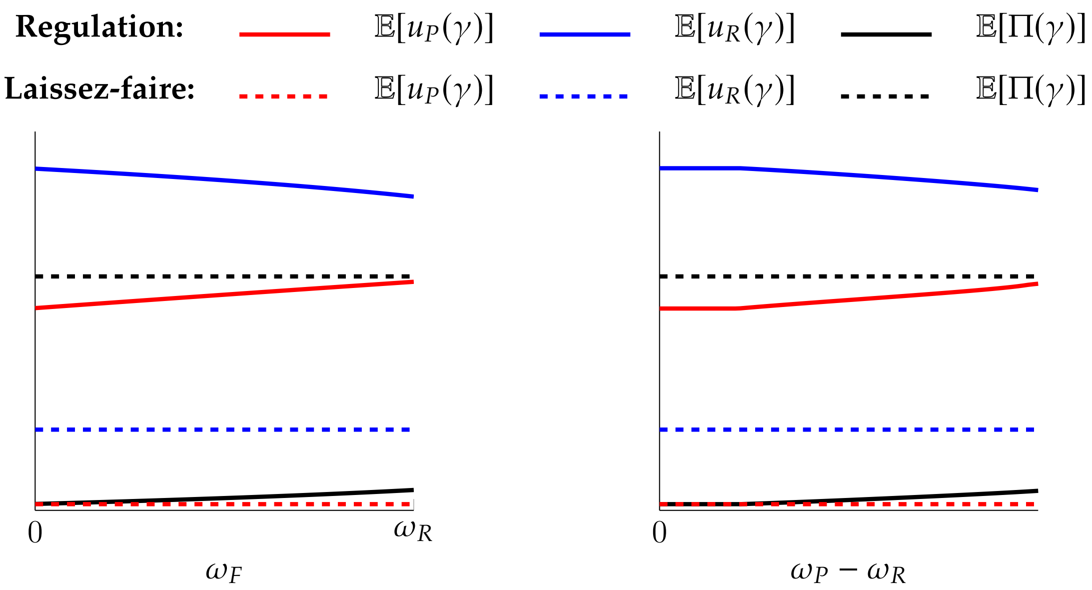

Figure 4: Expected surplus distribution under optimal regulation (solid) and unregulated monopoly screening (dashed) for different weight structures. Both panels assume G uniform on $[0,1]$, $\theta_P = 1$, $\theta_R = 3/2$, $v(q) = \sqrt{q}$, $c(q) = q^2$, and a hard budget so $t \equiv 0$. Left figure: fix $\omega_R = 1/2, \omega_P = 1$ and vary ω_F. Right figure: fix $\omega_F = 1/4$ and $\mathbb{E}[\omega_i] = 3/4$ and vary $\omega_P - \omega_R$.

from the rich to the poor and lowers the perceived social cost of the firm's information rents.[10] Part (ii) shows that weakening the across-side redistributive motive likewise makes the regulator switch to inequality-aware pricing at lower values of γ.

Figure 4 further illustrates how the rents of market participants vary with the weight structure under optimal regulation (solid) and unregulated monopoly screening (dashed). Regulation redistributes from the firm to consumers—most strongly so if the within-consumer redistributive motive is weak (i.e., $\omega_P - \omega_R$ is small) and ω_F is small. The gap between the payoffs of the rich and the poor is higher under optimal regulation than under unregulated monopoly screening (since regulation weakly raises the quality consumed by the poor), but less so if $\omega_P - \omega_R$ and ω_F are high.

[10] Using $\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] = \mathbb{E}[\omega_i] + \left(\mathbb{E}[\gamma'] - \mathbb{E}[\gamma' \mid \gamma' \geq \gamma]\right)(\omega_P - \omega_R)$, it is easy to observe that spreading the difference $\omega_P - \omega_R$, while keeping $\mathbb{E}[\omega_i]$, G and ω_F fixed, lowers $\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F$.

3.3 Continuous WTP type

Now consider the same model, but with the willingness to pay θ continuously distributed with full support on $[\underline{\theta}, \overline{\theta}]$ for both the rich and the poor. As before, F_R strictly first-order stochastically dominates F_P.

If the regulator observes demand state γ and hence the firm has no private information, the regulator can—and will—again set transfers and prices to exhaust the firm's profit. First-order stochastic dominance between F_R and F_P is then enough to guarantee that the regulator optimally implements a continuous-type counterpart of inequality-aware pricing; quality is distorted downward to enable lower prices for the poor without violating neither the consumers' incentive compatibility constraints nor the firm's participation constraint.

By contrast, when γ is privately known to the firm, the mechanism must also satisfy the firm's IC across demand states. Recall that if a firm with true demand γ selects the menu designed for γ', its profit is

$$\int_{\underline{\theta}}^{\overline{\theta}} (p(\theta, \gamma') - c(q(\theta, \gamma')))dF(\theta \mid \gamma) + t(\gamma'). \tag{12}$$

Using $F(\theta \mid \gamma) = \gamma F_R(\theta) + (1-\gamma)F_P(\theta)$, we can observe that the profit expression (12) is affine in γ. The firm's incentive compatibility can then be shown to be equivalent to

$$R(\gamma) = \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta, \gamma) - c(q(\theta, \gamma)))(f_R(\theta) - f_P(\theta))d\theta$$

being non-decreasing in γ, and the envelope condition

$$\Pi(\gamma) = \Pi(\underline{\gamma}) + \int_{\underline{\gamma}}^{\gamma} R(z)dz$$

holding for all γ.

In the following proposition, recall that I let $S(q, \theta) = \theta v(q) - c(q)$ denote the surplus, and the formula for function δ is

$$\delta(\gamma) := \gamma(1-\gamma)(\omega_P - \omega_R) - (\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F)\frac{1 - G(\gamma)}{g(\gamma)}.$$

Proposition 2. *Consider the continuous–WTP environment above.*

(i) If the share of rich consumers $\gamma \in (0, 1)$ is known to the regulator, then in any optimal regulation mechanism, the qualities are strictly distorted downward relative to the efficient benchmark: for almost all $\theta \in (\underline{\theta}, \overline{\theta})$,

$$q(\theta, \gamma) \ < \ \arg\max_q S(q, \theta)$$

and the price mark-up $p(\theta) - c(q(\theta))$ is increasing in θ (strictly increasing wherever $q(\theta)$ is strictly increasing), and $k'(t(\gamma)) = \mathbb{E}[\omega_i \mid \gamma]$.

(ii) If γ is privately known to the firm, then there is a cut-off $\gamma^* \geq \inf\{\gamma \in [\underline{\gamma}, \overline{\gamma}] : \delta(\gamma) \geq 0\} > \underline{\gamma}$ such that any optimal regulation mechanism has cost-plus pricing for almost all $\gamma \leq \gamma^*$, $\theta \in [\underline{\theta}, \overline{\theta}]$:

$$q(\theta, \gamma) = \arg\max_q S(q, \theta), \qquad p(\theta, \gamma) = c(q(\theta, \gamma)) - t(\gamma).$$

In any optimal regulation mechanism, the firm's profit $\Pi(\gamma)$ is weakly increasing in γ for all $\gamma \in \Gamma$, with $\Pi(\underline{\gamma}) = 0$, and $k'(t(\gamma)) = \mathbb{E}[\omega_i \mid \gamma]$.

Proposition 2 again highlights the tension between redistributing on the consumer side and limiting redistribution from consumers to the firm. Consumer-side inequality calls for inequality-aware pricing, but when the firm has private information, the regulator implements cost-plus pricing in low enough demand states.

4 Contracting with consumers: optimal subsidy design

Preliminaries of subsidy design. Public policy sometimes operates through consumer-side instruments (such as subsidies and taxes) rather than contracts with firms. For example in health care, governments do not necessarily make sophisticated contracts with firms but they may reimburse patients, which also affects profit-maximizing providers' offering.

This section studies such *contracting with consumers*. Instead of contracting with the firm, the regulator commits to a subsidy schedule $\tau : Q \to \mathbb{R}$, announced prior to the firm's choice of menu. The schedule maps a consumer's purchased quality to a (possibly negative) subsidy to the consumer, so that the equilibrium utility of consumer θ is

$$u(\theta, \gamma; q(\theta, \gamma), p(\theta, \gamma)) = \theta v(q(\theta, \gamma)) - p(\theta, \gamma) + \tau(q(\theta, \gamma)).$$

The regulator's objective, including the cost of public funds, remains as before.

The timeline of the game is summarized in Figure 5. Nature draws the demand state and consumer types. As the upstream principal, the regulator moves first and announces τ. The firm, as the downstream principal, chooses its menu after observing subsidy design τ and its private information about γ.

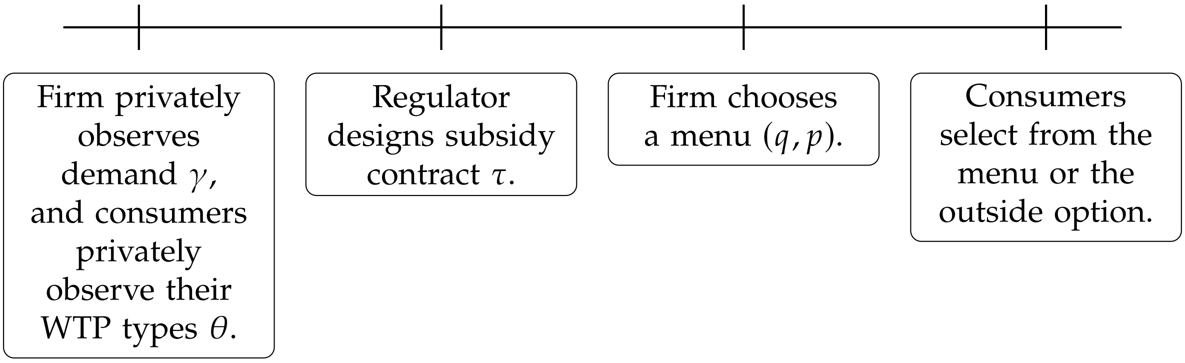

Figure 5: Contracting with consumers: timeline.

Finally, consumers select from the menu or the outside option and receive the appropriate subsidies.

For the analysis of subsidy design I work with the continuous-type framework of Section 3.3: for each group $i \in \{R, P\}$, θ is continuously distributed with full support on $[\underline{\theta}, \overline{\theta}]$. Throughout, I adopt the following additional assumptions:

Assumption 1. *(i) Regularity: $F(\cdot \mid \gamma)$ is regular for all $\gamma \in \Gamma$, i.e., the virtual value*

$$\psi(\theta \mid \gamma) := \theta - \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}$$

is strictly increasing on $(\underline{\theta}, \overline{\theta})$.

(ii) Hazard-rate dominance: for all $\theta \in (\underline{\theta}, \overline{\theta})$,

$$\frac{f_R(\theta)}{1 - F_R(\theta)} < \frac{f_P(\theta)}{1 - F_P(\theta)}.$$

Assumption 1(i) is a standard regularity condition that I impose on consumers' WTP types. Hazard-rate dominance assumption 1(ii) is slightly stronger than first-order stochastic dominance assumed in Section 3 but weaker than likelihood ratio dominance. It is straightforward to show that Assumption 1(ii) implies that for any $\gamma' > \gamma$, $F(\cdot \mid \gamma')$ strictly dominates $F(\cdot \mid \gamma)$ in the hazard-rate order.

As under firm-side regulation, I first consider the benchmark where γ is publicly known (Section 4.1), and then the case where γ is privately observed by the firm (Section 4.2).

4.1 Known demand state

Preliminary analysis. Since γ is common knowledge in this subsection, I write q and p as functions of θ only for notational ease. Consumers' incentive compatibility implies that for all θ,

$$\theta\, v\big(q(\theta)\big) - p(\theta) + \tau\big(q(\theta)\big) \;=\; \underline{u} \;+\; \int_{\underline{\theta}}^{\theta} v\big(q(z)\big)\, dz. \tag{13}$$

Substituting (13) into the expression for the firm's profit and integrating by parts yields[11]

$$\Pi(\gamma) = \int_{\underline{\theta}}^{\overline{\theta}} \Big(p(\theta) - c\big(q(\theta)\big)\Big)\, dF(\theta \mid \gamma) \tag{14}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} \Big[\underbrace{\psi(\theta \mid \gamma)\, v(q(\theta)) - c(q(\theta)) + \tau(q(\theta)) - \underline{u}}_{:=\,\pi(\theta, q(\theta))}\Big]\, dF(\theta \mid \gamma), \tag{15}$$

where I introduce notation $\pi(\theta, q(\theta))$ for the rent that the firm obtains from type θ. The firm chooses a weakly increasing $q(\cdot)$ and sets optimally the base-level utility $\underline{u} = 0$. Increases in subsidies are therefore passed through into higher firm profits, which makes redistribution to consumers difficult.

Since $\psi' > 0$ by Assumption 1, and $v' > 0$, the rent $\pi(\theta, q)$ has strictly increasing differences in (θ, q) for any subsidy design τ. By monotone comparative statics (Milgrom and Shannon, 1994), $q(\theta) \in \arg\max_q \pi(\theta, q)$ is then weakly increasing in θ, so the firm's monotonicity constraint does not bind; the firm chooses $q(\theta)$ to pointwise maximize $\pi(\theta, q(\theta))$. Therefore, by the envelope theorem,

$$\pi(\theta, q(\theta)) \;=\; \pi(\underline{\theta}, q(\underline{\theta})) \;+\; \int_{\underline{\theta}}^{\theta} \psi'(z \mid \gamma)\, v(q(z))\, dz. \tag{16}$$

Lemma 4. *In the continuous-WTP environment with known γ, a quality allocation $q : [\underline{\theta}, \overline{\theta}] \longrightarrow \boldsymbol{Q}$ is implementable by some subsidy schedule τ if and only if q is weakly increasing. The firm's profit equals*

$$\Pi \;=\; \pi(\underline{\theta}, q(\underline{\theta})) \;+\; \int_{\underline{\theta}}^{\overline{\theta}} \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, \psi'(\theta \mid \gamma)\, v\big(q(\theta)\big)\, dF(\theta \mid \gamma),$$

[11] Since trading quality 0 at price 0 and subsidy 0 is payoff-equivalent with no trade, it is without loss to focus on full coverage on the equilibrium path.

and an implementing subsidy satisfies, for all $\theta \in \Theta$,

$$\tau\big(q(\theta)\big) \;=\; \pi\big(\underline{\theta}, q(\underline{\theta})\big) \;-\; \psi(\theta \mid \gamma)\, v\big(q(\theta)\big) \;+\; c\big(q(\theta)\big) \;+\; \int_{\underline{\theta}}^{\theta} \psi'(z \mid \gamma)\, v\big(q(z)\big)\, dz. \tag{17}$$

Thus consumer-side subsidies can implement any monotone q, including the efficient allocation. The proof of Lemma 4 shows that for any weakly increasing allocation q, a subsidy schedule defined by (17) for all $q \in q[\Theta]$ (i.e. on the image) and $\tau(q) = \underline{\tau}$ for all $q \in Q \setminus q[\Theta]$ is well-defined and indeed incentivizes the firm to provide the desired allocation when $\underline{\tau}$ is low enough.

As seen in formula (15), the effect of the subsidy on the firm's incentives is similar to a modification of the cost function. Lemma 4 is therefore technically close to stating that any weakly increasing quality allocation is profit-maximizing for some cost function (when negative costs are allowed).

The regulator can adjust the base-level rent $\pi(\underline{\theta})$,[12] but is limited by the firm's option not to serve some consumers, which implies $\pi(\underline{\theta}) \geq 0$. When the rent from trading with low-WTP consumers is minimized, the firm still derives positive "obedience rents" from trades with high-WTP consumers. This contrasts with the firm-regulation framework, where the regulator can cross-subsidize across consumers by contracting over the entire menu and, when the firm has no private information, enforce zero profit.

Optimal subsidy design with known demand state. By using Lemma 4 and the consumers' envelope condition (13), each component of the regulator's payoff—the firm's profit, the regulator's spending, and consumer utilities—can be written as a functional of the allocation rule q and the base rent $\pi(\underline{\theta})$. The regulator's problem can therefore be reduced to choosing non-decreasing q and $\pi(\underline{\theta}) \geq 0$. A first-order approach yields the following characterization for the optimal choices, where

$$\kappa \;:=\; k'\bigg(\int_{\Theta} \tau\big(q(\theta)\big)\, dF(\theta \mid \gamma)\bigg)$$

denotes the induced marginal cost of public funds.

Proposition 3. *In the continuous-WTP environment with known γ, an optimal subsidy design exists. For almost every θ at which q is strictly increasing,*[13] *the allocation*

[12]For brevity, write $\pi(\underline{\theta})$ instead of $\pi(\underline{\theta}, q(\underline{\theta}))$.

[13]A nondecreasing function q is *strictly increasing at θ* if $q(\theta + \varepsilon) - q(\theta - \varepsilon) > 0$ for all $\varepsilon > 0$.

satisfies

$$\begin{aligned}
0 \;=\; & \kappa \left[\theta\, v'(q(\theta)) - c'(q(\theta))\right] && \text{(efficiency)} \\
& - \left(\kappa - \mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma]\right) \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, v'(q(\theta)) && \text{(consumers' rent)} \\
& - \left(\kappa - \omega_F\right) \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, \psi'(\theta \mid \gamma)\, v'(q(\theta)), && \text{(firm's rent)}
\end{aligned}$$

and the marginal subsidy satisfies

$$\tau'\big(q(\theta)\big) = \frac{v'(q(\theta))}{\kappa} \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \Big(\mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma] - \left(\kappa - \omega_F\right) \psi'(\theta \mid \gamma)\Big). \tag{18}$$

Finally, either $\kappa = \omega_F$ and $\pi(\underline{\theta}) \geq 0$, or $\kappa > \omega_F$ and $\pi(\underline{\theta}) = 0$.

Proposition 3 decomposes the regulator's first-order condition into: (i) an *efficiency* term (marginal money-metric surplus weighted by κ), (ii) a *consumers' rent* term, and (iii) a *firm's rent* term. The last two terms justify distortions relative to the efficient benchmark. They capture that increasing quality raises high-θ consumers' and the firm's rents, which the regulator prefers to limit when the weights on those consumers and the firm are low.

The formula (18) for the marginal subsidy shows that the optimal subsidy may be decreasing or increasing. For example, if the regulator has no redistributive motive—set $\kappa = \omega_R = \omega_P = \omega_F$—the formula reduces to

$$\tau'\big(q(\theta)\big) \;=\; v'(q(\theta)) \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)},$$

so implementing the efficient allocation requires an increasing subsidy to correct the monopolist's incentive to underprovide quality.

However, efficiency is not generally optimal under the inequality-aware objective. A decreasing subsidy schedule becomes optimal when: (i) high WTP θ is a strong signal of being rich and the planner assigns low weight ω_R to the rich; (ii) the planner places relatively low weight ω_F on the firm; and (iii) the virtual value $\psi(\theta \mid \gamma)$ has a large derivative.[14]

4.2 Unknown demand state

Preliminary analysis. I now return to the environment where the firm privately knows its demand state γ. The regulator commits to a single subsidy schedule $\tau : Q \to \mathbb{R}$. In Section 4.1, I showed that for any fixed τ, the firm's

[14]The large derivative implies that the firm has strong incentives to underprovide quality.

profit-maximizing selling mechanism satisfies

$$q(\theta,\gamma) \in \arg\max_{q\geq 0} \left\{\psi(\theta \mid \gamma)\, v(q) - c(q) + \tau(q)\right\}$$

for any (θ,γ), and it follows directly from Lemma 4 that an appropriately chosen subsidy schedule can implement any non-decreasing allocation $q(\cdot,\gamma)$ for a fixed demand state γ.

A key technical observation of this section is that for any (θ,γ), there exists some θ' such that $\psi(\theta \mid \gamma) = \psi(\theta' \mid \overline{\gamma})$. Then for any schedule τ,

$$\arg\max_{q\geq 0} \left\{\psi(\theta \mid \gamma)\, v(q) - c(q) + \tau(q)\right\} = \arg\max_{q\geq 0} \left\{\psi(\theta' \mid \overline{\gamma})\, v(q) - c(q) + \tau(q)\right\},$$

so the firm's profit-maximizing allocation to type θ in state γ coincides with the profit-maximizing allocation to type θ' in state $\overline{\gamma}$. Therefore, once the quality allocation in state $\overline{\gamma}$ is fixed, the quality allocation for every other state is pinned down by virtual–value matching.[15] This implies that with consumer subsidies, the regulator does not have similar state-contingent flexibility as when contracting directly with the firm (as in Section 3).

Lemma 5. *Any non-decreasing quality allocation $\overline{q}(\cdot) := q(\cdot,\overline{\gamma})$ in the highest demand state is implementable by consumer-side subsidies. Moreover, the schedule $\overline{q}(\cdot)$ determines $q(\theta,\gamma)$ for all γ and almost every θ, according to*

$$q(\theta,\gamma) = \overline{q}(\psi^{-1}(\psi(\theta \mid \gamma) \mid \overline{\gamma})),$$

and together with base-level rent $\pi(\underline{\theta},\overline{\gamma}) \geq 0$, it pins down all payoffs.

Optimal subsidy design with privately known demand state. Let $L(\theta,\gamma)$ be the WTP type to which the firm allocates the same quality in demand state γ as it allocates to WTP type θ in demand state $\overline{\gamma}$; such type can be found for all $\theta \in [\underline{\theta},\overline{\theta}]$, $\gamma \in [\hat{\gamma}(\theta),\overline{\gamma}]$, where $\hat{\gamma}(\theta) := (\psi(\underline{\theta}|\cdot))^{-1}(\min\{\psi(\theta|\overline{\gamma}), \psi(\underline{\theta}|\underline{\gamma})\})$ and $(\psi(\underline{\theta}|\cdot))^{-1}$ denotes the inverse in the γ-argument. Then

$$L(\theta,\gamma) := \psi^{-1}\big(\psi(\theta \mid \overline{\gamma}) \mid \gamma\big) \text{ for all } \theta \in [\underline{\theta},\overline{\theta}], \gamma \in [\hat{\gamma}(\theta),\overline{\gamma}].$$

Furthermore, let $\kappa(\gamma) := k'\Big(\int_\Theta \tau\big(q(\theta,\gamma)\big)\, dF(\theta \mid \gamma)\Big)$ denote the marginal cost of public funds.

Proposition 4. *Fix the environment above. The regulator's problem can be reduced to choosing a weakly increasing $\overline{q} : \Theta \to \mathcal{Q}$ and base-level rent $\pi(\underline{\theta},\overline{\gamma}) \geq 0$. An optimal*

[15] A slight caveat is that $\arg\max_q \psi(\theta \mid \gamma)v(q) - c(q) + \tau(q)$ is not necessarily a singleton, but the proof of Lemma 5 shows that nevertheless, for any γ, allocation $q(\cdot \mid \overline{\gamma})$ pins down $q(\theta,\gamma)$ for almost every θ.

design exists. For almost every θ where $\overline{q}$ is strictly increasing, an optimal choice of $\overline{q}$ satisfies[16]

$$
\begin{aligned}
0 = \int_{\hat{\gamma}(\theta)}^{\overline{\gamma}} \Big\{ & \kappa(\gamma)\,\left[Lv'(\overline{q}) - c'(\overline{q})\right] \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'(L \mid \gamma)} && \text{(efficiency)} \\
& - (\kappa(\gamma) - \mathbb{E}[\omega_i \mid \theta' \geq L, \gamma]) \frac{1 - F(L|\gamma)}{f(L|\gamma)} \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'(L \mid \gamma)} v'(\overline{q}) && \text{(consumers' rent)} \\
& - (\kappa(\gamma) - \omega_F) \frac{1 - F(L|\gamma)}{f(L|\gamma)} \psi'(\theta|\overline{\gamma})\, v'(\overline{q}) \Big\} f(L|\gamma) dG(\gamma) \\
& && \text{(firm's obedience rent)} \\
& - \int_{\underline{\gamma}}^{\hat{\gamma}(\theta)} (\kappa(\gamma) - \omega_F) \psi'(\theta \mid \overline{\gamma}) v'(\overline{q}) dG(\gamma). && \text{(firm's information rent)}
\end{aligned}
$$

and elsewhere, $\overline{q}$ is locally constant. Moreover, either $\int \kappa(\gamma)dG(\gamma) = \omega_F$ and $\pi(\underline{\theta}, \overline{\gamma}) \geq 0$, or $\int \kappa(\gamma)dG(\gamma) > \omega_F$ and $\pi(\underline{\theta}, \overline{\gamma}) = 0$.

Analogous to the known-demand benchmark (Proposition 3), the expression for the optimal policy in Proposition 4 consists of an efficiency component, a consumers' rent component, and a firm's rent component. The regulator's trade-offs are quite similar; the left panel of Figure 6 illustrates how the optimal subsidy varies with redistributive preferences. When redistributive concerns are strong (case i in Figure 6), the optimal subsidy again decreases with quality, limiting rents accruing to the firm and to high-θ consumers. When such concerns are weak (case ii), the optimal subsidy increases with quality to counteract the monopolist's tendency to underprovide quality, despite the induced rents. A difference from the known–demand case is that the trade-offs are now averaged across demand states through the aggregator L, so the schedule need not be optimal in any particular realized state.

A further difference from the known–demand case is that two distinct rent components for the firm arise. The firm's *obedience rent* has a counterpart in the known-demand benchmark, whereas the firm's *information rent* is new and arises for the following reason. Under consumer subsidies, the firm must earn a nonnegative rent π on any consumer it serves. The regulator can set the rent from the lowest consumer type to zero in the highest demand state, $\pi(\underline{\theta}, \overline{\gamma}) = 0$. In any lower demand state $\gamma < \overline{\gamma}$, however, the firm has stronger incentives to serve consumer type $\underline{\theta}$, and trade with type $\underline{\theta}$ can still be strictly profitable, $\pi(\underline{\theta}, \gamma) > 0$. The information rent component arises because the quality allocation affects $\pi(\underline{\theta}, \gamma)$.

[16] To be concise, I use notation $\overline{q}$ for $\overline{q}(\theta)$ and notation L for $L(\theta, \gamma)$.

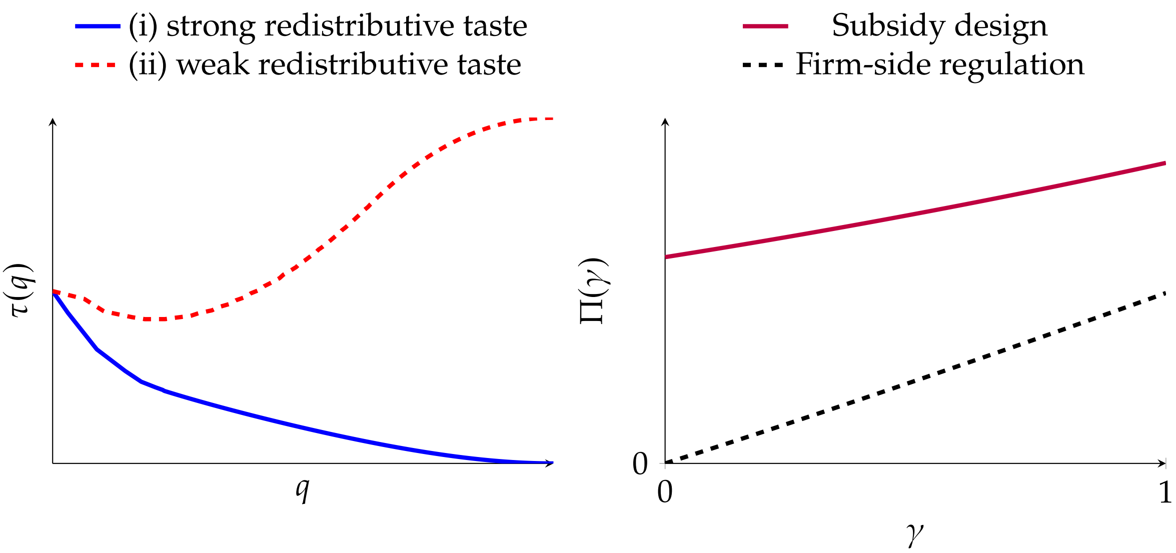

Figure 6: Left: Optimal subsidy $s(q)$ under two cases of redistributive tastes: (i) strong (ω_P=2, ω_R=1, ω_F=0) and (ii) weak ($\omega_P=\frac{5}{3}, \omega_R=\frac{4}{3}, \omega_F=\frac{2}{3}$). Primitives: $G \sim U[0,1]$, $f_P(\theta)=\frac{5}{2}-\theta$, $f_R(\theta)=\theta-\frac{1}{2}$ for $\theta \in [1,2]$, $v(q)=\sqrt{q}$, $c(q)=\frac{1}{2}q^2$, and k linear with slope 1 on the relevant subset of the domain. Right: Firm's profit $\Pi(\gamma)$ when the regulator implements the same quality allocation $q : \Theta \times \Gamma \to Q$ corresponding to part (i) of the left figure, with either consumer subsidies or firm regulation.

Comparison with firm-side regulation. The analysis thus far has highlighted two forces that favor contracting with the firm over contracting with consumers: (i) consumer-side subsidies cannot implement cross-subsidization, so the firm typically earns positive profit even absent firm-side private information; and (ii) subsidies offer limited scope for screening the firm's private information and for conditioning allocations on realized demand.

Nevertheless, I do not establish global dominance of firm-side regulation over consumer-side subsidies. The following channel can work in favor of subsidies. Because realized sales are not contractible, transfers under firm-side regulation cannot depend directly on the demand state. Under consumer-side contracting, by contrast, the regulator's spending varies with realized demand. Let $\Pi_C(\gamma)$ and $\Pi_F(\gamma)$ denote the firm's profit when the same allocation $q : \Theta \times \Gamma \to Q$ is implemented via consumer-side subsidies or via firm-side regulation, respectively. Under differentiability,

$$\Pi'_C(\gamma) - \Pi'_F(\gamma) = \int_\Theta \tau\big(q(\theta,\gamma)\big)\big(f_R(\theta) - f_P(\theta)\big)\, d\theta,$$

which is negative when subsidies to the poor exceed those to the rich (case (i) of Figure 6). In that case, consumer-side contracting reduces the sensitivity of profits to γ and thereby leaves less relative information rent to a high-γ firm (right panel of Figure 6). This benefit of consumer-side contracting is small when there is little uncertainty about market demand (cf. Section 4.1) or the motive for consumer-side redistribution is small (either because $\omega_P - \omega_R$ is small or because willingness-to-pay θ is only weakly informative about $i \in \{R, P\}$); in these cases, contracting with the firm outperforms consumer-side subsidies.[17]

The comparison above also raises the question of what is achievable when both instruments are available. With a mechanism that maps the firm's report into (i) a price–quality menu and (ii) a consumer-subsidy schedule, the regulator can require cost-based pricing to enforce zero profit and use subsidies to implement optimal redistribution among consumers subject to consumer IC. An optimal mechanism is then outcome-equivalent to optimal firm-side regulation under known demand (see Propositions 1 and 2(i)). In practice, regulators often face constraints on instruments, and whether firm-side regulation, consumer-side subsidies, or both are feasible is application-specific. A university procuring on-campus food service typically contracts with the provider without subsidizing downstream customers, whereas a central government designing redistributive policy in consumer markets (e.g., transport or food) may find it more feasible to rely on consumption taxes and subsidies than to regulate individual firms' offerings.

Policy instruments in subsidy design. The regulator may benefit from augmenting the subsidy scheme with a subsidy for those who do not purchase from the firm. An alternative interpretation for this additional element is that the regulator can offer a zero-quality public option—which consumers may choose instead of participating in the private market—and design a price for this option. By raising this outside option by Δ and simultaneously increasing all purchase-contingent subsidies by Δ, the regulator keeps the firm's allocation incentives unchanged while redistributing Δ from its own budget to all consumers. The design of the outside option therefore serves as a lump-sum redistributive instrument between consumers and the regulator (absent this design, type $\underline{\theta}$ always receives payoff 0). The level of the outside option is optimally chosen so that the expected marginal cost of public funds equals the average consumer weight, $\mathbb{E}[\kappa(\gamma)] = \mathbb{E}[\omega_i]$, but the rest of the subsidy design remains characterized by Proposition 4.

[17] I have not identified a numerical example in which contracting with consumers would strictly outperform contracting with the firm.

A second issue regarding the available instruments is that throughout this section, subsidies have been assumed to depend on allocation q. While this is a natural starting point, an alternative formulation would be to let the subsidy be conditioned on the transaction price, $\tau_p : \mathcal{P} \to \mathbb{R}$.[18] Unlike in a competitive environment, price-based and quality-based subsidies can imply different sets of implementable allocations and distribution of rent.[19] Finally, the subsidies could potentially be even more complicated, conditioning on both price and quality. A comprehensive analysis of different consumer-side instruments in a monopoly market is left for future research.

5 Conclusion

The two perhaps most important forces that misalign a profit-maximizing firm's decisions with social objectives are lack of competition and redistributive concerns. I study optimal regulation in that environment. Private information on both sides of the market makes policy design a nontrivial mechanism design problem. Consumers' private information makes it optimal to use consumption behavior to screen consumers' welfare weights, rationalizing inequality-aware pricing. The firm's private information makes it optimal to distort the mechanism to limit the firm's information rents, rationalizing cost-plus pricing. Both instruments are widespread in practice.

On an applied level, I build a bridge between two well-developed literatures. Market power and optimal regulatory solutions have long been studied in industrial organization literature, whereas redistribution has long been studied in public economics. Yet there is relatively little research that tries to understand the interaction of imperfect competition and economic inequality. This motivates further theoretical and empirical work.

On a more technical level, I develop a tractable sequential framework with an upstream and a downstream principal, where the downstream principal holds private information. The model is natural for studying the interaction

[18]Conditioning the subsidy on price is especially attractive if the quality of service is hard to specify in a contract, as discussed in Hart et al. (1997).

[19]This can be confirmed in a linear model, where $v(q) = q$ and $c(q) = cq$, with q restricted to a compact interval. To analyze optimal design of price subsidies in this setting, represent the envelope as $\frac{u(\theta,\gamma)}{\theta} = \frac{u(\underline{\theta},\gamma)}{\underline{\theta}} - \int_{\underline{\theta}}^{\theta} \frac{\tau_p(p(z,\gamma)) - p(z,\gamma)}{z^2} dz$, write the firm's profit as a functional of p, and proceed using an approach analogous to this paper's analysis of quality-based subsidies. More details regarding the analysis of price-based subsidies are available upon request.

of a regulator, a firm, and consumers, but the structure also seems suitable for studying, e.g., multiple levels or generations of government or vertical relationships among firms.

An important assumption in the current work—and an avenue for future work—is the extent of regulatory commitment. This paper assumes full commitment (as quite standard), but in practice commitment may be more limited (e.g., due to political cycles and other reasons for renegotiation). Limited commitment can lead to ratchet effects (see, e.g., Freixas et al., 1985). In a similar vein, an interesting variation of the subsidy design framework of Section 4 would be one where the moves of the regulator and the firm are reversed. For example, if the firm moves first, anticipating the regulator's desire to support the poor through subsidies, the firm might find it optimal to design an offering that leaves the poor very needy.

Appendix

A Proofs

Proofs are presented in the order in which the corresponding results appear in the text:

A.1 Proof of Lemma 1

Proof. I will first prove the only-if direction: the constraints (IR-B), (IC-B), (IR-F), and (IC-F) are necessary to be satisfied in equilibrium. The necessity of (IR-B) and (IR-F) is immediate from the availability of the outside options for consumers and the firm. In any state γ, a consumer of type θ can choose any option $(q(\theta', \gamma), p(\theta', \gamma))$, hence (IC-B) is necessary. The necessity of (IC-F) follows since the deviation profit of firm γ from choosing the equilibrium menu of firm γ' is

$$\Pi\big(\gamma; q(\cdot, \gamma'), p(\cdot, \gamma'), t(\gamma')\big) = \int_{\Theta} \big(p(\theta, \gamma') - c(q(\theta, \gamma'))\big)\, dF(\theta \mid \gamma) + t(\gamma'),$$

as each consumer type's choice after the deviation must be the same as in state γ'.

I will then prove the if direction: Suppose that the constraints (IR-B), (IC-B), (IR-F), and (IC-F) hold for (q, p, t), and I will prove that there exists some transfer function $\hat{t} : \mathcal{M} \to \mathbb{R}$ such that (q, p, t) is implemented in the continuation game.

Let $M(\gamma) = \{(q, p) : (q, p) = (q(\theta, \gamma), p(\theta, \gamma)) \text{ for some } \theta \in \Theta\}$ be the menu chosen by the firm in state γ. The firm's equilibrium profit, excluding the regulator's transfer, is bounded above by some finite $\overline{S}$.[20] Let

$$\hat{t}(M) := \begin{cases} t(\gamma) & \text{if } M = M(\gamma) \text{ for some } \gamma, \\ -\overline{S} & \text{otherwise.} \end{cases}$$

This is a well-defined function because if $M(\gamma) = M(\gamma')$ then $t(\gamma) = t(\gamma')$ for constraint (IC-F) to be satisfied. Given that (IC-B) and (IR-B) are satisfied, consumers act sequentially rationally. A firm with any demand γ does not prefer exiting by (IR-F). Furthermore, the firm does not have a strictly profitable deviation to an off-path menu M' such that $\hat{t}(M') = -\overline{S}$, and it does not have a strictly profitable deviation to an on-path menu $M(\gamma')$ by (IC-F). □

[20]E.g., define $\overline{S} := \sup_{\gamma \in \Gamma} \int_{\Theta} \sup_{q \in Q} \big(\theta v(q) - c(q)\big) dF(\theta|\gamma)$; given the Inada-type assumptions on v and c, there is a unique finite $\sup_{q \in Q} \big(\theta v(q) - c(q)\big)$ for all θ.

A.2 Proof of Proposition 1

Proof. Since the firm has no private information, the regulator chooses $q_R, q_P \in \mathbb{R}_{\geq 0}, p_R, p_P, t \in \mathbb{R}$ to maximize

$$\gamma\omega_R(\theta_R v(q_R) - p_R) + (1-\gamma)\omega_P(\theta_P v(q_P) - p_P)$$
$$+\omega_F(\gamma(p_R - c(q_R)) + (1-\gamma)(p_P - c(q_P))) - k(t)$$

subject to the firm's IR constraint

$$\gamma(p_R - c(q_R)) + (1-\gamma)(p_P - c(q_P)) + t \geq 0, \tag{19}$$

the consumers' IC constraints

$$\theta_R v(q_R) - p_R \geq \theta_R v(q_P) - p_P \tag{20}$$
$$\theta_P v(q_P) - p_P \geq \theta_P v(q_R) - p_R, \tag{21}$$

and the consumers' IR constraints

$$\theta_R v(q_R) - p_R \geq 0 \tag{22}$$
$$\theta_P v(q_P) - p_P \geq 0. \tag{23}$$

According to a standard argument, the IR constraint of the poor (23) and the IC constraint of the rich (20) imply the IR constraint of the rich (22).

The firm's IR constraint (19) binds in the solution as otherwise p_R and p_P could be decreased by small $dp > 0$, which would not violate any constraint and would increase the regulator's payoff by $(\mathbb{E}[\omega_i] - \omega_F)dp$, which is positive by assumption.

Moreover, the rich consumers' IC constraint (20) binds as otherwise the regulator could increase p_R by some small $dp > 0$ and decrease p_P by $\frac{\gamma}{1-\gamma}dp$ without violating any constraints, and this would increase the regulator's payoff by $\gamma dp(\omega_P - \omega_R)$ which is positive by assumption.

Given that the two constraints bind, we may solve for p_R and p_P as a function of q_R, q_P, and t and plug that into the regulator's objective to write the objective as

$$\mathbb{E}[\omega_i \mid \gamma]\Big(\gamma\big[\theta_R v(q_R) - c(q_R)\big] + (1-\gamma)\big[\theta_P v(q_P) - c(q_P)\big] + t\Big) \tag{24}$$
$$- (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P)v(q_P) - k(t).$$

To obtain part (i) of the proposition, note that given the assumptions about function k, there is an optimal t which solves $k'(t) = \mathbb{E}[\omega_i \mid \gamma]$ and satisfies $t > 0$.

To obtain part (ii) of the proposition, note that the marginal effect of q_R on

(24) is

$$\mathbb{E}[\omega_i \mid \gamma]\,\gamma\left[\theta_R v'(q_R) - c'(q_R)\right].$$

Given that v is a continuously differentiable, strictly concave function such that $\lim_{q\to\infty} v'(q) = 0$ and $\lim_{q\to 0} v'(q) = \infty$ and c is a continuously differentiable, increasing and convex function, there is a uniquely optimal q_R which is interior and solves $\theta_R v'(q_R) - c'(q_R) = 0$, proving part ii of the proposition.

To obtain part (iii) of the proposition, note first that

$$\gamma\frac{\omega_P - \omega_R}{\omega_P} < \frac{\theta_P}{\theta_R} \tag{25}$$

$$\implies \mathbb{E}[\omega_i \mid \gamma](1-\gamma)\theta_P - (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P) > 0 \tag{26}$$

where $\mathbb{E}[\omega_i \mid \gamma] = \gamma\omega_R + (1-\gamma)\omega_P$ is used.

Therefore, under condition (25), and since v is concave and c convex, we have

$$\mathbb{E}[\omega_i \mid \gamma](1-\gamma)\theta_P v''(q) - (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P)v''(q) < 0 \tag{27}$$

$$\implies \mathbb{E}[\omega_i \mid \gamma](1-\gamma)(\theta_P v''(q) - c''(q)) - (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P)v''(q) < 0 \tag{28}$$

for all $q \in \mathbb{R}_+$, so the first-order condition for q_P,

$$\mathbb{E}[\omega_i \mid \gamma](1-\gamma)\left[\theta_P v'(q_P) - c'(q_P)\right] - (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P)v'(q_P) = 0, \tag{29}$$

is sufficient for optimality. Note that $v'(q)$ goes to infinity as $q \to 0$, so under condition (26), the LHS of (29) is positive for low enough positive value of q_P. Furthermore, since the LHS of (29) is negative for large enough value of q_P (since $v'(q) \to 0$ as $q \to \infty$), there is a (unique) value of q_P that solves (29).

On the other hand,

$$\gamma\frac{\omega_P - \omega_R}{\omega_P} \geq \frac{\theta_P}{\theta_R} \tag{30}$$

$$\implies \mathbb{E}[\omega_i \mid \gamma](1-\gamma)\theta_P - (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P) \leq 0$$

$$\implies \mathbb{E}[\omega_i \mid \gamma](1-\gamma)\left[\theta_P v'(q) - c'(q)\right] - (\mathbb{E}[\omega_i \mid \gamma] - \omega_R)\gamma(\theta_R - \theta_P)v'(q) < 0\ \forall q \in \mathbb{R}_{\geq 0},$$

so under condition (30), it is uniquely optimal to set $q_P = 0$, which finalizes the proof.

Note that the assumptions guarantee that there is a unique "efficient" quality for the poor, $q_P^{FB} > 0$ such that $\theta_P v'(q_P^{FB}) - c'(q_P^{FB}) = 0$, and the regulator's optimal choice of q_P is lower than q_P^{FB} regardless of whether it satisfies (29) or $q_P = 0$.

Furthermore, the price satisfies

$$p_P = c(q_P) - \gamma\big[(\theta_R v(q_R) - c(q_R)) - (\theta_R v(q_P) - c(q_P))\big] - t$$

where the expression inside the square brackets is positive when q_R is chosen efficiently, i.e. to maximize $\theta_R v(q_R) - c(q_R)$. So then, since $t > 0$ in the optimal policy, $p_P < c(q_P)$. Since $\theta_P v(q_P) > c(q_P)$ for all $q_P \leq q_P^{FB}$, then $\theta_P v(q_P) - c(q_P) > 0$ so the IR constraints are also clearly satified in the solution characterized in Proposition 1.

□

A.3 Proof of Remark 1

Proof. Consider the problem of an unregulated monopolist. Standard arguments imply that the monopolist sets poor IR and rich IC binding, so profit maximization over (q_P, q_R) is equivalent to maximizing

$$\gamma\Big(\theta_R v(q_R) - (\theta_R - \theta_P)v(q_P) - c(q_R)\Big) + (1-\gamma)\big(\theta_P v(q_P) - c(q_P)\big).$$

The first-order condition for an interior $q_P^M > 0$ is

$$(1-\gamma)\big(\theta_P v'(q_P^M) - c'(q_P^M)\big) = \gamma(\theta_R - \theta_P)v'(q_P^M). \tag{31}$$

In the regulator's optimum (Proposition 1), an interior $q_P^\star > 0$ satisfies

$$(1-\gamma)\big(\theta_P v'(q_P^\star) - c'(q_P^\star)\big) = \frac{\mathbb{E}[\omega_i \mid \gamma] - \omega_R}{\mathbb{E}[\omega_i \mid \gamma]}\,\gamma(\theta_R - \theta_P)v'(q_P^\star). \tag{32}$$

Since

$$\frac{\mathbb{E}[\omega_i \mid \gamma] - \omega_R}{\mathbb{E}[\omega_i \mid \gamma]} < 1,$$

it is straightforward that a solution to (32) is greater than a solution to (31). Furthermore, if the regulator's solution is at the corner $q_P^\star = 0$, so that we have $(1-\gamma)\theta_P \leq \frac{\mathbb{E}[\omega_i|\gamma]-\omega_R}{\mathbb{E}[\omega_i|\gamma]}\gamma(\theta_R - \theta_P)$, then also $(1-\gamma)\theta_P \leq \gamma(\theta_R - \theta_P)$, so that the monopolist's solution is also at the corner, $q_P^M = 0$. Hence in all cases $q_P^\star \geq q_P^M$.

□

A.4 Proof of Remark 2

Proof. Let $(q_P^\star, p_P^\star, q_R^\star, p_R^\star, t^\star)$ denote the regulator's optimal mechanism from Proposition 1. Consider the following regulatory instrument: the firm must include the option

$$(\underline{q}, \overline{p}) = (q_P^\star, p_P^\star)$$

in any posted menu; the regulator pays the transfer $t^\star$ irrespective of the rest of the menu. The firm then freely chooses any additional options to maximize profit.

The firm's problem is then effectively to choose (q_P, p_P, q_R, p_R) to maximize its profit

$$\gamma(p_R - c(q_R)) + (1-\gamma)(p_P - c(q_P)) + t \tag{33}$$

subject to the constraint that both the rich and the poor obtain at least the utility they obtain from the mandated option:

$$\theta_R v(q_R) - p_R \geq \theta_R v(\underline{q}) - \overline{p}, \tag{34}$$

$$\theta_P v(q_P) - p_P \geq \theta_P v(\underline{q}) - \overline{p}, \tag{35}$$

and the rich IC constraint

$$\theta_R v(q_R) - p_R \geq \theta_R v(q_P) - p_P. \tag{36}$$

Clearly, (35) must bind (otherwise the firm would want to increase p_P), and either (34) or (36) must bind (otherwise the firm would want to increase p_R).

If $q_P > \overline{q}$, then (36) binds and (34) does not, and the marginal benefit from increasing q_P is

$$(1-\gamma)\big(\theta_P v'(q_P) - c'(q_P)\big) \; - \; \gamma(\theta_R - \theta_P)v'(q_P). \tag{37}$$

which is negative when $q_P > \overline{q} = q_P^\star$ (see the proof of Remark 1), so that it is profitable to decrease q_P.

On the other hand, if $q_P < \overline{q}$, then (34) binds and (36) does not, and the marginal benefit from increasing q_P is

$$(1-\gamma)\big(\theta_P v_P'(q_P) - c_P'(q_P)\big)$$

which is positive when $q_P < \overline{q} = q_P^\star$, so it is profitable to increase q_P.

It follows that the firm maximizes profit by setting $(q_P, p_P, q_R, p_R) = (q_P^\star, p_P^\star, q_R^\star, p_R^\star)$. ☐

A.5 Proof of Lemma 2

Proof. By the envelope theorem, a necessary condition for incentive compatibility is

$$p_P(\gamma) - c(q_P(\gamma)) + t(\gamma) = \Pi(\underline{\gamma}) - \gamma r(\gamma) + \int_{\underline{\gamma}}^{\gamma} r(\gamma')d\gamma'.$$

We can now write the difference between the profits that a firm with demand type γ obtains from a menu designed to type γ and a menu designed to type $\hat{\gamma}$ as

$$
\begin{aligned}
&\gamma r(\gamma) + p_P(\gamma) - c(q_P(\gamma)) + t(\gamma) - \Big(\gamma r(\hat{\gamma}) + p_P(\hat{\gamma}) - c(q_P(\hat{\gamma})) + t(\hat{\gamma})\Big) \\
=& \int_{\underline{\gamma}}^{\gamma} r(\gamma')d\gamma' - \Big(\gamma r(\hat{\gamma}) - \hat{\gamma} r(\hat{\gamma}) + \int_{\underline{\gamma}}^{\hat{\gamma}} r(\gamma')d\gamma'\Big) \\
=& \int_{\hat{\gamma}}^{\gamma} (r(\gamma') - r(\hat{\gamma}))d\gamma'.
\end{aligned}
$$

The first equality uses the envelope formula and the second equality is housekeeping. The final expression is clearly non-negative for all $\gamma, \hat{\gamma} \in \Gamma$ if and only if $r(\gamma)$ is non-decreasing in γ. □

A.6 Proof of Lemma 3

Proof. The regulator's payoff is given by

$$
\int_{\underline{\gamma}}^{\overline{\gamma}} \Big(\gamma \omega_R u(\theta_R, \gamma) + (1-\gamma)\omega_P u(\theta_P, \gamma) + \omega_F \Pi(\gamma) - k(t(\gamma))\Big)\, dG(\gamma). \tag{38}
$$

First, using the definitions of u, S and r, we can write

$$
u(\theta_R, \gamma) - u(\theta_P, \gamma) = S_R\big(q_R(\gamma)\big) - S_P\big(q_P(\gamma)\big) - r(\gamma). \tag{39}
$$

There is a "resource constraint" such that

$$
\gamma u(\theta_R, \gamma) + (1-\gamma)u(\theta_P, \gamma) = \gamma S_R\big(q_R(\gamma)\big) + (1-\gamma)S_P\big(q_P(\gamma)\big) + t(\gamma) - \Pi(\gamma). \tag{40}
$$

We then obtain

$$
\begin{aligned}
&\gamma \omega_R u(\theta_R, \gamma) + (1-\gamma)\omega_P u(\theta_P, \gamma) \\
=&\mathbb{E}[\omega_i \mid \gamma]\big[\gamma u(\theta_R, \gamma) + (1-\gamma)u(\theta_P, \gamma)\big] - \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\, \gamma\, (u(\theta_R, \gamma) - u(\theta_P, \gamma)) \\
=&\mathbb{E}[\omega_i \mid \gamma]\big[\gamma S_R\big(q_R(\gamma)\big) + (1-\gamma)S_P\big(q_P(\gamma)\big) + t(\gamma) - \Pi(\gamma)\big] \qquad (41)\\
&- \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\, \gamma \Big[S_R\big(q_R(\gamma)\big) - S_P\big(q_P(\gamma)\big) - r(\gamma)\Big].
\end{aligned}
$$

where the first equality simply uses $\mathbb{E}[\omega_i \mid \gamma] = \gamma\omega_R + (1-\gamma)\omega_P$, and the second equality uses (39) and (40).

Furthermore,

$$
\int_{\underline{\gamma}}^{\overline{\gamma}} (\mathbb{E}[\omega_i \mid \gamma] - \omega_F)\, \Pi(\gamma)\, dG(\gamma) = \int_{\underline{\gamma}}^{\overline{\gamma}} (\mathbb{E}[\omega_i \mid \gamma] - \omega_F)\Big(\Pi(\underline{\gamma}) + \int_{\underline{\gamma}}^{\gamma} r(\gamma')d\gamma'\Big)dG(\gamma) \tag{42}
$$

$$= \int_{\underline{\gamma}}^{\overline{\gamma}} \Big((\mathbb{E}[\omega_i \mid \gamma] - \omega_F)\Pi(\underline{\gamma}) + \big(\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F\big)\frac{1 - G(\gamma)}{g(\gamma)} r(\gamma) \Big) dG(\gamma)$$

where the first equality uses Lemma 2 and the second uses integration by parts.

We can then use results (41) and (42) to write the regulator's payoff (38) as

$$\int_{\underline{\gamma}}^{\overline{\gamma}} \Big\{ \mathbb{E}[\omega_i \mid \gamma] \big[\gamma S_R(q_R(\gamma)) + (1-\gamma) S_P(q_P(\gamma)) + t(\gamma) \big]$$
$$- \ \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\, \gamma \big[S_R(q_R(\gamma)) - S_P(q_P(\gamma)) - r(\gamma) \big]$$
$$- \ \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_F\big)\Pi(\underline{\gamma}) - \big(\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F\big) \frac{1 - G(\gamma)}{g(\gamma)}\, r(\gamma) \ - \ k\big(t(\gamma)\big) \Big\}\, dG(\gamma).$$

□

A.7 Proof of Theorem 1

I use shorthand notations

$$\bar{\omega}_i(\gamma) := \mathbb{E}[\omega_i \mid \gamma] = \gamma\omega_R + (1-\gamma)\omega_P, \qquad \bar{\omega}^+(\gamma) := \frac{\int_{\gamma}^{\overline{\gamma}} \bar{\omega}_i(\gamma')\, dG(\gamma')}{1 - G(\gamma)}.$$

Let

$$S_R(q) := \theta_R v(q) - c(q), \qquad S_P(q) := \theta_P v(q) - c(q),$$

and denote the (unique) efficient qualities by

$$q_R^{FB} \in \arg\max_q S_R(q), \qquad q_P^{FB} \in \arg\max_q S_P(q).$$

The firm's mark-up difference is

$$r(\gamma) \ := \ \big(p_R(\gamma) - c(q_R(\gamma))\big) - \big(p_P(\gamma) - c(q_P(\gamma))\big),$$

so rich consumers' IC is

$$r(\gamma) \ \leq \ S_R(q_R(\gamma)) - S_R(q_P(\gamma)). \tag{43}$$

Using Lemma 3, we know that the regulator chooses $(q_R, q_P, r, t, \Pi(\underline{\gamma}))$ to maximize

$$\int_{\underline{\gamma}}^{\overline{\gamma}} \Phi(\gamma; q_R, q_P, r, t, \Pi(\underline{\gamma}))\, dG(\gamma),$$

where

$$\Phi(\gamma; \cdot) = \mathbb{E}[\omega_i \mid \gamma] \big[\gamma S_R(q_R(\gamma)) + (1-\gamma) S_P(q_P(\gamma)) + t(\gamma) \big] \tag{44}$$
$$- \ \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\, \gamma \big[S_R(q_R(\gamma)) - S_P(q_P(\gamma)) - r(\gamma) \big]$$
$$- \ \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_F\big)\Pi(\underline{\gamma}) - \big(\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F\big) \frac{1 - G(\gamma)}{g(\gamma)}\, r(\gamma) \ - \ k\big(t(\gamma)\big),$$

subject to r nondecreasing, (43), and firm IR

$$\Pi(\underline{\gamma}) + \int_{\underline{\gamma}}^{\gamma} r(\gamma')\, d\gamma' \;\geq\; 0. \tag{45}$$

I ignore the consumers' IR constraints and the poor consumers' IC constraint in the analysis, but it is easy to check that they are satisfied in the solution that I find.

Finally, define

$$\begin{aligned}\delta(\gamma) :=& \big(\bar{\omega}(\gamma) - \omega_R\big)\gamma - \big(\bar{\omega}^+(\gamma) - \omega_F\big)\frac{1 - G(\gamma)}{g(\gamma)} && (46)\\ =& \gamma(1-\gamma)(\omega_P - \omega_R) - (\bar{\omega}^+(\gamma) - \omega_F)\frac{1 - G(\gamma)}{g(\gamma)}. && (47)\end{aligned}$$

Lemma 6. *From any feasible mechanism $(q_R, q_P, r, t, \Pi(\underline{\gamma}))$ one can obtain another feasible mechanism that weakly improves the regulator's payoff and satisfies:*

(a) $k'(t(\gamma)) = \bar{\omega}(\gamma)$ *for all* $\gamma \in \Gamma$*;*

(b) $q_R(\gamma) = q_R^{\text{eff}}$ *for all* $\gamma \in \Gamma$*;*

(c) $\Pi(\underline{\gamma}) = 0$ *and* $r(\gamma) \geq 0$ *for all* $\gamma \in \Gamma$*;*

(d) There exists a cut-off $\gamma^* \in [\gamma_*, \overline{\gamma}]$*, where* $\gamma_* := \inf\{\gamma \in [\underline{\gamma}, \overline{\gamma}] :\; \delta(\gamma) \geq 0\} > \underline{\gamma}$*, such that* $q_P(\gamma) = q_P^{\text{eff}}$ *and* $r(\gamma) = 0$ *for all* $\gamma \leq \gamma^*$*, whereas* $q_P(\gamma) \leq q_P^{\text{eff}}$ *and rich-IC binds (i.e.* $r(\gamma) = S_R(q_R^{\text{eff}}) - S_R(q_P(\gamma))$*) for all* $\gamma > \gamma^*$*.*

Proof. Property (a). Consider a feasible mechanism with transfers t. Since $t(\gamma)$ enters the integrand (44) only through $\bar{\omega}(\gamma)t(\gamma) - k(t(\gamma))$ and does not enter the constraints, replacing $t(\gamma)$ by $\tilde{t}(\gamma) \in \arg\max_{t\in\mathbb{R}} \bar{\omega}(\gamma)t - k(t)$ for all $\gamma \in \Gamma$ is feasible and strictly increases the objective if $t(\gamma) \notin \arg\max_{t\in\mathbb{R}} \bar{\omega}(\gamma)t - k(t)$ for a positive-measure subset of Γ. Since k is continuously differentiable and convex, with $k'(0) \leq \omega_R$ and $k'(t) \to \infty$ as $t \to \infty$, there is a solution to problem $\max_{t\in\mathbb{R}} \bar{\omega}(\gamma)t - k(t)$ which satisfies the first-order condition $k'(t(\gamma)) = \bar{\omega}(\gamma)$.

Property (b). The quality of the rich, $q_R(\gamma)$, enters the integrand (44) only through term $\omega_R\gamma\, S_R(q_R)$, where the coefficient is strictly positive. Therefore, replacing $q_R(\gamma)$ by efficient q_R^{eff} for each γ strictly increases the integrand if $q_R(\gamma) \neq q_R^{\text{eff}}$ and preserves feasibility: the IC constraint of the rich consumers, (43), becomes slackened because $S_R(q_R)$ rises, and r and its monotonicity are unchanged. If in the original mechanism, $q_R(\gamma) \neq q_R^{\text{eff}}$ for a positive-measure subset of Γ, the objective strictly increases.

Property (c). Here, take first any feasible mechanism that satisfies properties (a)–(b) in the lemma. By feasibility, in this original mechanism, (45) is satisfied and r is non-decreasing in γ. Hence there is a cut-off $\hat{\gamma} \in [\underline{\gamma}, \overline{\gamma}]$ such that $r(\gamma) < 0$ for all $\gamma < \hat{\gamma}$ and $r(\gamma) \geq 0$ for all $\gamma > \hat{\gamma}$. Modify the mechanism so that in the new mechanism, the lowest firm type's profit is

$$\tilde{\Pi}(\underline{\gamma}) = 0,$$

and modified mark-up difference $\tilde{r}$ is introduced such that

$$\tilde{r}(\gamma) := \begin{cases} 0, & \gamma \leq \hat{\gamma}, \\ r(\gamma), & \gamma > \hat{\gamma}, \end{cases}$$

without any changes to qualities and transfers t. In the new mechanism, the monotonicity constraint, the firm's IR constraint and the rich consumers' IC constraints remain satisfied[21]. The pointwise value of the integrand (44) increases for all γ:

- The second line of (44) strictly increases for all $\gamma < \hat{\gamma}$ and weakly increases for all $\gamma \geq \hat{\gamma}$.
- The third line of (44) weakly increases for all $\gamma \in \Gamma$: the modified mechanism minimizes the profit of the firm with any demand $\gamma \leq \hat{\gamma}$, and the profit of every firm with demand $\gamma > \hat{\gamma}$ decreases by $\Pi(\hat{\gamma}) \geq 0$.

So, modifying the mechanism in this way improves the regulator's payoff without violating the constraints.

Property (d). Start from a mechanism satisfying (a)–(c). For each γ, move $q_P(\gamma)$ closer to q_P^{eff} (keeping $r(\gamma)$ fixed) until $q_P(\gamma) = q_P^{\text{eff}}$ or the rich consumers' IC constraint binds. Because the coefficient of $S_P(q_P)$ in (44) is positive and S_P is strictly concave, this weakly improves the objective and preserves feasibility. After this step, for all γ, either $q_P(\gamma) = q_P^{FB}$, or the rich consumers' IC constraint holds as an equality.

Let

$$r_0 := S_R(q_R^{\text{eff}}) - S_R(q_P^{\text{eff}}) > 0, \qquad \hat{\gamma} := \inf\{\gamma : \ r(\gamma) > 0\} \ \in [\underline{\gamma}, \overline{\gamma}].$$

By monotonicity of r, we have $r(\gamma) < r_0$ for all $\gamma < \hat{\gamma}$ and $r(\gamma) \geq r_0$ for all $\gamma > \hat{\gamma}$. On $[\underline{\gamma}, \hat{\gamma})$, we have $0 \leq r(\gamma) < r_0$ and $q_P(\gamma) = q_P^{\text{eff}}$.

[21] For the last claim, it is important that the original mechanism satisfies property (b) in the lemma: $q_R(\gamma) = q_R^{\text{eff}}$ for all $\gamma \in \Gamma$.

Consider the set of alternative mark-up difference schedules

$$\tilde{R} := \left\{\tilde{r} : [\underline{\gamma}, \overline{\gamma}] \to \mathbb{R}_+ : \tilde{r} \text{ nondecreasing}, \tilde{r}(\gamma) \in [0, r_0] \, \forall \gamma < \hat{\gamma}, \, \tilde{r}(\gamma) = r(\gamma) \, \forall \gamma \geq \hat{\gamma}\right\}.$$

This set is nonempty, convex, and compact in L^1 and contains r. Over $\tilde{R}$ the objective (44) is affine in $\tilde{r}$. By Bauer's maximum principle, the maximum is attained at an extreme point of $\tilde{R}$, which is $\{0, r_0\}$-valued on $[\underline{\gamma}, \hat{\gamma})$ and hence of the form

$$\tilde{r}(\gamma) = \begin{cases} 0, & \gamma < \gamma^*, \\ r_0, & \gamma \in [\gamma^*, \hat{\gamma}), \\ r(\gamma), & \gamma \geq \hat{\gamma}, \end{cases}$$

for some $\gamma^* \in [\underline{\gamma}, \hat{\gamma}]$. After this step, the mechanism has a cut-off structure such that $q_P(\gamma) = q_P^{FB}$ and $r(\gamma) = 0$ for all $\gamma \leq \gamma^*$, whereas $q_P(\gamma) \leq q_P^{\text{eff}}$ and the rich consumers' IC binds for all $\gamma > \gamma^*$.

Finally, since the coefficient of $r(\gamma)$ in the integrand (44) is $\delta(\gamma) < 0$ for all $\gamma < \gamma_* := \inf\{\gamma : \delta(\gamma) \geq 0\}$, choosing $r(\gamma) = 0$ for all $\gamma < \gamma_*$ increases the regulator's objective relative to any without violating any constraints. Hence, choosing $\gamma^* \geq \gamma_*$ improves the objective relative to any $\gamma^* < \gamma_*$. Note that $\gamma_* > \underline{\gamma}$ by continuity of δ and $\delta(\underline{\gamma}) < 0$. $\square$

Lemma 7. *In the class of feasible mechanisms satisfying Lemma 6(a)–(d), there exists a mechanism that maximizes the regulator's payoff.*

Proof. By Lemma 6, the free choices are a cutoff $\gamma^* \in [\gamma_*, \overline{\gamma}]$ and a non-increasing schedule q_P with $q_P(\gamma) = q_P^{\text{eff}}$ for $\gamma \leq \gamma^*$ and $q_P(\gamma) \in [0, q_P^{\text{eff}}]$ for $\gamma > \gamma^*$. For a fixed γ^* define

$$Q_P(\gamma^*) := \left\{q_P : [\underline{\gamma}, \overline{\gamma}] \to [0, q_P^{\text{eff}}] \text{ non-increasing} : \; q_P(\gamma) = q_P^{\text{eff}} \; \forall \gamma \leq \gamma^*\right\}.$$

$Q_P(\gamma^*)$ is (sequentially) compact in L^1 and nonempty. With $t^*(\gamma)$ pinned down by $k'(t^*(\gamma)) = \bar{\omega}(\gamma)$, the regulator's payoff reduces to

$$\begin{aligned} W(q_P, \gamma^*) := &\int_{\underline{\gamma}}^{\gamma^*} \Big\{\bar{\omega}_i(\gamma)\Big(\gamma S_R(q_R^{FB}) + (1-\gamma) S_P(q_P^{FB}) + t^*(\gamma)\Big) \\ &\qquad - (\bar{\omega}_i(\gamma) - \omega_R)\gamma\big(S_R(q_R^{FB}) - S_P(q_P^{FB})\big) - k\big(t^*(\gamma)\big)\Big\} dG(\gamma) \\ &+ \int_{\gamma^*}^{\overline{\gamma}} \Big\{\bar{\omega}_i(\gamma)\Big(\gamma S_R(q_R^{FB}) + (1-\gamma) S_P(q_P(\gamma)) + t^*(\gamma)\Big) \\ &\qquad - (\bar{\omega}_i(\gamma) - \omega_R)\gamma\big(S_R(q_P(\gamma)) - S_P(q_P(\gamma))\big) \\ &\qquad - (\bar{\omega}^+(\gamma) - \omega_F)\frac{1 - G(\gamma)}{g(\gamma)}\big(S_R(q_R^{FB}) - S_R(q_P(\gamma))\big) - k\big(t^*(\gamma)\big)\Big\} dG(\gamma), \end{aligned} \tag{48}$$

which is continuous in q_P under L^1 (the integrand is continuous in $q_P(\gamma)$ and dominated by integrable $\bar{\omega}_i(\gamma)\big(\gamma S_R(q_R^{FB})+(1-\gamma)S_P(q_P^{FB})+t^*(\gamma)\big)-k(t^*)$). Hence, for each fixed γ^*, $W(\cdot,\gamma^*)$ attains a maximum on $\mathcal{Q}_P(\gamma^*)$ (Weierstrass). By Berge's maximum theorem, the regulator's value function $\gamma^* \mapsto \max_{q_P\in\mathcal{Q}_P(\gamma^*)} W(q_P,\gamma^*)$ is continuous on the compact set $[\gamma_*,\overline{\gamma}]$ and therefore attains its supremum. □

Lemma 8. *In any optimal mechanism satisfying Lemma 6(a)–(d), the cut-off γ^* must satisfy $\delta(\gamma^*) = 0$.*

Proof. Let $(q_R, q_P, r, t, \Pi(\underline{\gamma}))$ satisfying Lemma 6(a)–(d) be optimal with cut-off γ^*. By Lemma 6(b)–(d),

$$q_R(\gamma) = q_R^{\text{eff}} \text{ for all } \gamma, \qquad q_P(\gamma) = q_P^{\text{eff}} \text{ and } r(\gamma) = 0 \text{ for all } \gamma \le \gamma^*,$$

and for $\gamma > \gamma^*$ the rich–type IC binds so that

$$r(\gamma) = S_R\big(q_R^{\text{eff}}\big) - S_R\big(q_P(\gamma)\big).$$

Recall

$$r_0 := S_R\big(q_R^{\text{eff}}\big) - S_R\big(q_P^{\text{eff}}\big).$$

Using $q_P(\gamma) \le q_P^{\text{eff}} < q_R^{\text{eff}}$ for $\gamma > \gamma^*$, we have

$$r(\gamma) = S_R\big(q_R^{\text{eff}}\big) - S_R\big(q_P(\gamma)\big) \;\ge\; S_R\big(q_R^{\text{eff}}\big) - S_R\big(q_P^{\text{eff}}\big) = r_0 \tag{49}$$

for $\gamma > \gamma^*$.

First, suppose by contradiction $\delta(\gamma^*) > 0$. For small $\varepsilon > 0$, define $\tilde{r}(\gamma) = r(\gamma)$ for $\gamma \notin [\gamma^* - \varepsilon, \gamma^*)$ and $\tilde{r}(\gamma) = r_0$ on $[\gamma^* - \varepsilon, \gamma^*)$.[22] All the constraints remain satisfied (in particular, $\tilde{r}$ is monotonic and satisfies (43)). The change in the objective equals

$$\int_{\gamma^*-\varepsilon}^{\gamma^*} \delta(\gamma)\,(r_0 - 0)\,dG(\gamma) \;>\; 0$$

for ε small by continuity of δ and $\delta(\gamma^*) > 0$, contradicting optimality.

Second, suppose by contradiction $\delta(\gamma^*) < 0$. For small $\varepsilon > 0$, define $\tilde{r}(\gamma) = r(\gamma)$ for $\gamma \notin [\gamma^*, \gamma^* + \varepsilon)$ and $\tilde{r}(\gamma) = 0$ for $\gamma \in [\gamma^*, \gamma^* + \varepsilon)$.[23] All the constraints remain satisfied. The change in the objective equals

$$-\int_{\gamma^*}^{\gamma^*+\varepsilon} \delta(\gamma)\,r(\gamma)\,dG(\gamma) \;>\; 0,$$

[22]For small enough $\varepsilon > 0$, $\gamma^* - \varepsilon \in (\underline{\gamma}, \overline{\gamma})$, as we know $\gamma^* > \underline{\gamma}$.

[23]For small enough $\varepsilon > 0$, $\gamma^* + \varepsilon \in (\underline{\gamma}, \overline{\gamma})$, since we must have $\gamma^* < \overline{\gamma}$ given that $\delta(\overline{\gamma}) = 0$.

for small ε by continuity of δ and $\delta(\gamma^*) < 0$, again contradicting optimality. ☐

Lemma 9. *In any optimal mechanism, for (almost) every $\gamma \in [\gamma^*, \overline{\gamma}]$, either $q_P(\gamma)$ is locally constant, or it is strictly decreasing and satisfies*

$$
\begin{aligned}
0 = \bar{\omega}(\gamma)(1-\gamma)\big[\theta_P v'(q_P(\gamma)) - c'(q_P(\gamma))\big] - \big(\bar{\omega}(\gamma) - \omega_R\big)\gamma(\theta_R - \theta_P)v'(q_P(\gamma)) \\
+ \ \big(\bar{\omega}(\gamma) - \omega_F\big)\frac{1-G(\gamma)}{g(\gamma)}\big[\theta_R v'(q_P(\gamma)) - c'(q_P(\gamma))\big]. \qquad (50)
\end{aligned}
$$

Proof. By Lemmas 6 and 7, we know that an optimal regulation mechanism exists, satisfies properties (a)–(e) in Lemma 6 for almost every $\gamma \in \Gamma$, and the regulator's optimal payoff is given by (48). Optimality conditions for the problem of choosing decreasing q_P on $(\gamma^*, \overline{\gamma}]$ are presented, e.g., in Hellwig (2008). If the monotonicity constraint does not bind, optimality requires $q_P(\gamma)$ to satisfy the first-order condition (50) for almost every $\gamma > \gamma^*$; if the monotonicity constraint binds, $q_P(\gamma)$ is locally constant. ☐

Theorem 1 follows from simply combining Lemmas 6, 7, 8, and 9.

A.8 Proof of Remark 6

Proof. From $\mathbb{E}[\omega_i] = \mathbb{E}[\gamma]\omega_R + (1-\mathbb{E}[\gamma])\omega_P$, we obtain:

$$
\omega_R = \mathbb{E}[\omega_i] - (1-\mathbb{E}[\gamma'])(\omega_P - \omega_R), \qquad \omega_P = \mathbb{E}[\omega_i] + \mathbb{E}[\gamma'](\omega_P - \omega_R),
$$

and therefore,

$$
\begin{aligned}
\mathbb{E}[\omega_i \mid \gamma] &= \mathbb{E}[\omega_i] + (\mathbb{E}[\gamma'] - \gamma)\,(\omega_P - \omega_R), \\
\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] &= \mathbb{E}[\omega_i] + \big(\mathbb{E}[\gamma'] - \mathbb{E}[\gamma' \mid \gamma' \geq \gamma]\big)\,(\omega_P - \omega_R). \qquad (51)
\end{aligned}
$$

Then

$$
\delta(\gamma) = (\omega_P - \omega_R)\Big[\gamma(1-\gamma) + \big(\mathbb{E}[\gamma' \mid \gamma' \geq \gamma] - \mathbb{E}[\gamma']\big)\frac{1-G(\gamma)}{g(\gamma)}\Big] - \big(\mathbb{E}[\omega_i] - \omega_F\big)\frac{1-G(\gamma)}{g(\gamma)},
$$

so holding $\mathbb{E}[\omega_i]$ fixed, increasing (decreasing) the spread $\omega_P - \omega_R$ strictly increases (strictly decreases) $\delta(\gamma)$ for all $\gamma \in (\underline{\gamma}, \overline{\gamma})$. Similarly, increasing (decreasing) ω_F strictly increases (strictly decreases) $\delta(\gamma)$ for all $\gamma \in (\underline{\gamma}, \overline{\gamma})$.

Now, given an increase in spread $\omega_P - \omega_R$ or in ω_F, let δ_a and δ_b be the pre- and post-change function, so that $\delta_b(\gamma) > \delta_a(\gamma)$ for all $\gamma \in (\underline{\gamma}, \overline{\gamma})$ by the earlier observation. Suppose by contradiction that after the change, an optimal mechanism has a higher cut-off than before: $\gamma_b^* > \gamma_a^*$. Therefore, in the pre-change mechanism, $r_a(\gamma) \geq r_0$ for all $\gamma \in (\gamma_a^*, \gamma_b^*)$, and in the post-change mechanism, $r_b(\gamma) = 0$ for all $\gamma \in (\gamma_a^*, \gamma_b^*)$.[24] Since the pre-change mechanism is optimal, decreasing r by r_0 on (γ_a^*, γ_b^*), without any changes in the quality allocation, can't strictly increase the regulator's payoff as such a change would not violate the constraints. Therefore,

$$0 \leq r_0 \int_{\gamma_a^*}^{\gamma_b^*} \delta_a(\gamma) d\gamma < r_0 \int_{\gamma_a^*}^{\gamma_b^*} \delta_b(\gamma) d\gamma \tag{52}$$

where the strict inequality follows since $\delta_b(\gamma) > \delta_a(\gamma)$ for all $\gamma \in (\gamma_a^*, \gamma_b^*)$. This implies that in the post-change mechanism, increasing r from 0 to r_0 on $(\gamma_a^*, \gamma_b^*]$ increases the regulator's payoff, and as it is also feasible, this contradicts the optimality of the post-change mechanism.

The analysis regarding a *decrease* in spread $\omega_P - \omega_R$ or in ω_F is completely analogous.

□

A.9 Proof of Remark 3

Proof. Fix any state $\gamma \in (0, 1)$. Under laissez-faire (LF), the monopolist knows γ and chooses $q_R^{LF}(\gamma) = \arg\max S_R(q)$ and either an interior poor quality $q_P^{LF}(\gamma) > 0$ that solves

$$(1-\gamma)\big(\theta_P v'(q) - c'(q)\big) \;=\; \gamma(\theta_R - \theta_P)\, v'(q), \tag{53}$$

or the corner $q_P^{LF}(\gamma) = 0$ if no interior solution exists. The LF solution is downward distorted relative to the poor's efficient quality $q_P^{\text{eff}} = \arg\max_q S_P(q)$, i.e. $q_P^{LF}(\gamma) < q_P^{\text{eff}}$, and $q_P^{LF}(\gamma)$ is (weakly) decreasing in γ.

Consider the optimal regulation mechanism $(q_P^\star, p_P^\star, q_R^\star, p_R^\star, t^\star)$. For $\gamma \leq \gamma^*$ (cost-plus region), $q_P^\star(\gamma) = q_P^{\text{eff}} \geq q_P^{LF}(\gamma)$ by the previous observation. It remains to compare the two for $\gamma > \gamma^*$ (inequality-aware region).

From Theorem 1(iii), when $q_P^\star(\gamma)$ is strictly decreasing, it satisfies

$$\mathbb{E}[\omega_i \mid \gamma](1-\gamma)S_P'(q) \;-\; \big(\mathbb{E}[\omega_i \mid \gamma] - \omega_R\big)\gamma\big(S_R'(q) - S_P'(q)\big) \tag{54}$$

$$+ \big(\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F\big)\frac{1 - G(\gamma)}{g(\gamma)} S_R'(q) = 0, \tag{55}$$

that is,

$$(1-\gamma)\big(\theta_P v'(q) - c'(q)\big) \;-\; \frac{\mathbb{E}[\omega_i \mid \gamma] - \omega_R}{\mathbb{E}[\omega_i \mid \gamma]}\gamma(\theta_R - \theta_P)\, v'(q) \tag{56}$$

[24] Recall that $r_0 := S_R(q_R^{\text{eff}}) - S_R(q_P^{\text{eff}}) > 0$.

$$+\frac{\mathbb{E}[\omega_i \mid \gamma' \geq \gamma] - \omega_F}{\mathbb{E}[\omega_i \mid \gamma]} \frac{1 - G(\gamma)}{g(\gamma)} (\lambda_R v'(q) - c'(q)) = 0, \tag{57}$$

Since $\frac{\mathbb{E}[\omega_i|\gamma]-\omega_R}{\mathbb{E}[\omega_i|\gamma]} < 1$ and $q_P^\star(\gamma) \leq q_P^{FB} < q_R^{FB}$, it is straightforward that a solution to (53) is greater than a solution to (56).

If $q_P^\star(\gamma)$ is constant on an interval $[\gamma_a, \gamma_b]$ (ironing or $q_P^\star(\gamma) \equiv 0$), then by Theorem 1(iii) and continuity the interior FOC holds at γ_a, hence $q_P^\star(\gamma_a) \geq q_P^{LF}(\gamma_a)$. Because $q_P^\star(\cdot)$ is weakly decreasing in γ on the inequality-aware region and $q_P^{LF}(\cdot)$ is weakly decreasing in γ by (53), for every $\gamma \in [\gamma_a, \gamma_b]$,

$$q_P^\star(\gamma) = q_P^\star(\gamma_a) \;\geq\; q_P^{LF}(\gamma_a) \;\geq\; q_P^{LF}(\gamma).$$

This completes the comparison in all cases. □

A.10 Proof of Remark 4

I will first prove a helpful lemma.

Lemma 10. *In the optimal regulation mechanism of Theorem 1, $u_P(\gamma) - t(\gamma)$ is weakly increasing in γ on $[\underline{\gamma}, \overline{\gamma}]$.*

Proof. On $[\underline{\gamma}, \gamma^*]$ (cost–plus region), $q_P(\gamma) = \arg\max_q S_P(q)$ and $r(\gamma) \equiv 0$, so

$$u_P(\gamma) - t(\gamma) = \theta_P v\big(q_P(\gamma)\big) - p_P(\gamma) - t(\gamma) = S_P\big(q_P(\gamma)\big)$$

is constant. It remains to show weak monotonicity on $[\gamma^*, \overline{\gamma}]$.

For $\gamma > \gamma^*$, rich IC binds against the poor option, hence

$$u_R(\gamma) - u_P(\gamma) = \big(\theta_R - \theta_P\big)\, v\big(q_P(\gamma)\big) = S_R\big(q_P(\gamma)\big) - S_P\big(q_P(\gamma)\big).$$

Substituting this into the resource constraint

$$\gamma S_R\big(q_R(\gamma)\big) + (1-\gamma) S_P\big(q_P(\gamma)\big) + t(\gamma) = (1-\gamma) u_P(\gamma) + \gamma u_R(\gamma) + \Pi(\gamma),$$

we obtain

$$u_P(\gamma) - t(\gamma) = \gamma S_R\big(q_R(\gamma)\big) + (1-\gamma) S_P\big(q_P(\gamma)\big) - \gamma\Big(S_R\big(q_P(\gamma)\big) - S_P\big(q_P(\gamma)\big)\Big) - \Pi(\gamma). \tag{58}$$

For almost all $\gamma \geq \gamma^*$, $\Pi'(\gamma) = r(\gamma)$ with $r(\gamma) = S_R\big(q_R(\gamma)\big) - S_R\big(q_P(\gamma)\big)$, and $q_R(\gamma)$ is constant in γ. Differentiating (58) at such γ's gives

$$\big(u_P(\gamma) - t(\gamma)\big)' = \big[(1-\gamma) S_P'(q_P(\gamma)) - \gamma\big(S_R'(q_P(\gamma)) - S_P'(q_P(\gamma))\big)\big] \cdot q_P'(\gamma).$$

By Theorem 1(iii), $q_P(\cdot)$ is weakly decreasing on $[\gamma^*, \overline{\gamma}]$, so $q_P'(\gamma) \leq 0$. By Remark 3, $q_P(\gamma) \geq q_P^{LF}(\gamma)$ for all γ, which is equivalent to

$$(1-\gamma) S_P'\big(q_P(\gamma)\big) - \gamma\big(S_R'\big(q_P(\gamma)\big) - S_P'\big(q_P(\gamma)\big)\big) \;\leq\; 0.$$

Therefore $u_P'(\gamma) - t'(\gamma) \geq 0$, finalizing the proof. □

Let us then prove the actual remark.

Remark 4. *Inequality-aware pricing options admit a simple implementation: for each γ in the region $\gamma > \gamma^*$, it suffices to mandate that the firm includes the option $\big(q_P(\gamma), p_P(\gamma)\big)$ in its menu and receives the transfer $t(\gamma)$. No further menu restrictions are required.*

Proof. Fix the optimal mechanism of Theorem 1, $(q_P^\star, p_P^\star, q_R^\star, p_R^\star, t^\star)$ and its simple implementation: the inequality-aware option designed for any $\gamma > \gamma^*$ only requires the firm to include the option $(q_P^\star(\gamma), p_P^\star(\gamma))$ in its menu and is associated with transfer $t^\star(\gamma)$; otherwise the firm's menu is unrestricted.

Let us consider the incentives of a firm with true state γ choosing the simple-implementation option for $\gamma' > \gamma^*$. Given that the firm chooses this option, it faces the problem of choosing $(q_P(\gamma'), p_P(\gamma'), q_R(\gamma'), p_R(\gamma'))$ to maximize its profit to maximize

$$\gamma\big(p_R(\gamma') - c(q_R(\gamma'))\big) + (1-\gamma)\big(p_P(\gamma') - c(q_P(\gamma'))\big) + t^\star(\gamma') \tag{59}$$

subject to the constraint that both the rich and the poor obtain at least the utility they obtain from the mandated option:

$$\theta_R v(q_R(\gamma')) - p_R(\gamma') \geq \theta_R v(q_P^\star(\gamma')) - p_P^\star(\gamma'), \tag{60}$$

$$\theta_P v(q_P(\gamma')) - p_P(\gamma') \geq \theta_P v(q_P^\star(\gamma')) - p_P^\star(\gamma'). \tag{61}$$

and the rich IC constraint

$$\theta_R v(q_R(\gamma')) - p_R(\gamma') \geq \theta_R v(q_P(\gamma')) - p_P(\gamma'). \tag{62}$$

In this problem, (61) must bind (otherwise the firm would want to increase $p_P(\gamma')$), and either (60) or (62) must bind (otherwise the firm would want to increase $p_R(\gamma')$).

If $q_P(\gamma') > q_P^\star(\gamma')$, then (62) binds and (60) does not, and the marginal benefit from increasing $q_P(\gamma')$ is

$$(1-\gamma)\big(\theta_P v'(q_P(\gamma')) - c'(q_P(\gamma'))\big) \;-\; \gamma(\theta_R - \theta_P)v'(q_P(\gamma')). \tag{63}$$

which is negative if $q_P(\gamma') > q_P^{LF}(\gamma)$ and positive if $q_P(\gamma') < q_P^{LF}(\gamma)$ (where q_P^{LF} corresponds to the quality that the laissez-faire monopolist assigns to the poor).

On the other hand, if $q_P(\gamma') < q_P^\star(\gamma')$, then (60) binds and (62) does not, and the marginal benefit from increasing $q_P(\gamma')$ is

$$(1-\gamma)\big(\theta_P v_P'(q_P(\gamma')) - c_P'(q_P(\gamma'))\big)$$

which is positive when $q_P(\gamma') < q_P^\star(\gamma') \le q_P^{FB}$, so it is profitable to increase q_P.

Hence:

(i) If $q_P^{LF}(\gamma) \le q_P^\star(\gamma')$, the deviator allocates the mandate $q_P^\star(\gamma')$ to the poor.

(ii) If $q_P^{LF}(\gamma) > q_P^\star(\gamma')$, the deviator allocates the laissez-faire quality $q_P^{LF}(\gamma)$ to the poor and chooses $p_P(\gamma')$ to keep the constraint (61) binding.

Because $q_P^\star(\cdot)$ is weakly decreasing on $(\gamma^*, \overline{\gamma}]$ (Theorem 1(iii)), the inequality $q_P^{LF}(\gamma) \lesseqgtr q_P^\star(\gamma')$ defines a cutoff $\hat{\gamma}(\gamma) \in [\gamma^*, \overline{\gamma}]$ so that

- For $\gamma' \in [\gamma^*, \hat{\gamma}(\gamma)]$ we have case (i),
- For $\gamma' > \hat{\gamma}(\gamma)$ we have case (ii).

For $\gamma' \le \hat{\gamma}(\gamma)$, the deviator's best response coincides with the mandated poor option, so the induced menu equals the intended γ'-menu; standard single-crossing across states then implies no profitable deviation within this region (already shown in Lemma 2).

For $\gamma' > \hat{\gamma}(\gamma)$, substituting $q_P(\gamma') = q_P^{LF}(\gamma)$, the deviator's payoff simplifies to

$$\Pi^{LF}(\gamma) - u_P^\star(\gamma') + t(\gamma')$$

where $\Pi^{LF}(\gamma)$ is the laisse-faire profit (note that the firm must provide the poor with utility $u_P^\star(\gamma')$ according to constraint (61)). Since $u_P^\star(\gamma') - t(\gamma')$ is increasing in γ' by Lemma 10, the deviation payoff is decreasing in γ', and the result that the firm does not want to deviate to any option γ' follows. □

A.11 Proof of Remark 5

Proof. Let us first prove that the price gap $p_R(\gamma) - p_P(\gamma)$ is monotonic. For $\gamma \le \gamma^*$, Theorem 1 gives cost-plus pricing, hence

$$p_R(\gamma) - p_P(\gamma) = c\big(q_R^{\text{eff}}\big) - c\big(q_P^{\text{eff}}\big).$$

For $\gamma > \gamma^*$, the rich consumers' IC constraint binds, so

$$p_R(\gamma) - p_P(\gamma) = \theta_R\Big(v\big(q_R^{\text{eff}}\big) - v\big(q_P(\gamma)\big)\Big).$$

Since $q_P(\cdot)$ is decreasing on $(\gamma^*, \overline{\gamma}]$ and v is strictly increasing, the right-hand side is increasing in γ on $(\gamma^*, \overline{\gamma}]$. Furthermore, for any $q_P(\gamma) \le q_P^{\text{eff}}$, we have

$$c\big(q_R^{\text{eff}}\big) - c\big(q_P^{\text{eff}}\big) \;<<\; \theta_R\Big(v\big(q_R^{\text{eff}}\big) - v\big(q_P(\gamma)\big)\Big),$$

because $S_R(q_R^{\text{eff}}) >> S_R(q_P^{\text{eff}}) \geq \theta_R v\big(q_P(\gamma)\big) - c\big(q_P^{\text{eff}}\big)$. Hence, there is an upward jump in the price gap at γ^*.

Let us then prove that under an ex-post budget balance requirement for the regulator, $t(\gamma) = 0$ for all $\gamma \in \Gamma$, $p_P(\cdot)$ is weakly decreasing for all $\gamma \in \Gamma$. By contradiction, if $p_P(\gamma'') > p_P(\gamma')$ for some $\gamma'' > \gamma'$, then we have by the first proposition also $p_R(\gamma'') > p_R(\gamma')$, and furthermore by Theorem 1 also $c(q_P(\gamma'')) \leq c(q_P(\gamma'))$. Therefore, it cannot be incentive-compatible for any firm to choose the menu designed for firm γ'. □

A.12 Proof of Proposition 2

Lemma 11. *Fix the environment of the text; in particular, suppose that F_R strictly first-order stochastically dominates F_P: $F_R(\theta) < F_P(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta})$. Then for every $\theta \in (\underline{\theta}, \overline{\theta})$,*

$$\mathbb{E}[\omega_i \mid \gamma] \;>\; \mathbb{E}[\omega_i \mid \theta' \geq \theta,\ \gamma].$$

Proof. The expected weight ω_i conditional on γ only can be written as

$$\mathbb{E}[\omega_i \mid \gamma] = \gamma\,\omega_R + (1-\gamma)\,\omega_P$$

while the expected weight ω_i conditional on both γ and event $\theta' \geq \theta$ can be written as

$$\mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma] = \frac{\omega_R \gamma\,[1 - F_R(\theta)] + \omega_P (1-\gamma)\,[1 - F_P(\theta)]}{\gamma\,[1 - F_R(\theta)] + (1-\gamma)\,[1 - F_P(\theta)]}$$

So we get:

$$\mathbb{E}[\omega_i \mid \gamma] - \mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma] \tag{64}$$

$$= \frac{\Big(\gamma\,[1 - F_R] + (1-\gamma)\,[1 - F_P]\Big)\mathbb{E}[\omega_i \mid \gamma] - \omega_R\gamma\,[1 - F_R] - \omega_P(1-\gamma)\,[1 - F_P]}{\gamma\,[1 - F_R] + (1-\gamma)\,[1 - F_P]} \tag{65}$$

$$= \frac{\gamma(1-\gamma)\,(F_P(\theta) - F_R(\theta))\,(\omega_P - \omega_R)}{\gamma\,[1 - F_R] + (1-\gamma)\,[1 - F_P]}, \tag{66}$$

which is positive for all $\theta \in (\underline{\theta}, \overline{\theta})$ by the FOSD property and $\omega_P > \omega_R$. □

Lemma 12. *In the continuous-WTP environment of Section 3.3, if the share of rich consumers $\gamma \in (0, 1)$ is known to the regulator, then in any optimal regulation mechanism, the qualities are strictly distorted downward relative to the efficient benchmark: for almost all $\theta \in (\underline{\theta}, \overline{\theta})$,*

$$q(\theta, \gamma) \;<\; \arg\max_q S(q, \theta)$$

and the price mark-up $p(q(\theta)) - c(q(\theta))$ is increasing in θ and strictly increasing in quality $q(\theta)$.

Proof. Since the firm has no private information, for a given value of γ, the regulator can just choose quality allocation $q : \Theta \to Q$ and price allocation $p : \Theta \to \mathbb{R}$ and the transfer $t \in \mathbb{R}$ to the firm to maximize

$$\int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta, \gamma](\theta v(q(\theta)) - p(\theta))dF(\theta \mid \gamma) - k(t) \tag{67}$$

$$\omega_F\Big(t + \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta) - c(q(\theta)))dF(\theta \mid \gamma)\Big) \tag{68}$$

(where the first line has the consumers' weighted payoffs and the cost of the regulator's spending, and the second line has the firm's weighted profit), subject to the firm's IR constraint

$$t + \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta) - c(q(\theta)))dF(\theta \mid \gamma) \geq 0, \tag{69}$$

the consumers' IC constraints

$$\theta v(q(\theta)) - p(\theta) \in \arg\max_{\theta'} \theta v(q(\theta')) - p(\theta')$$

for all θ, and the consumers' IR constraints

$$\theta v(q(\theta)) - p(\theta) \geq 0$$

for all θ.

The firm's IR constraint (19) binds in the solution as otherwise all prices could be decreased by small $dp > 0$ without violating any constraint, which would increase the regulator's payoff by $(\mathbb{E}[\omega_i \mid \gamma] - \omega_F)dp$, which is positive by assumption.

Furthermore, a standard Myersonian result implies that incentive-compatibility is equivalent with q being non-decreasing and envelope condition

$$p(\theta) = \theta v(q(\theta)) - \int_{\underline{\theta}}^{\theta} v(q(z))dz - \underline{u}. \tag{70}$$

being satisfied for all θ and some constant $\underline{u}$.

Using the binding firm's IR constraint and (70), we can then write the regulator's payoff as a function of the allocation:

$$\int_{\underline{\theta}}^{\overline{\theta}} \Big\{\bar{\omega}_i(\gamma)\Big[\theta v(q(\theta)) - c(q(\theta)) + t\Big] - k(t) \tag{71}$$

$$-(\bar{\omega}_i(\gamma)-\mathbb{E}[\omega_i \mid \theta' \geq \theta,\gamma])v(q(\theta))\frac{1-F(\theta\mid\gamma)}{f(\theta\mid\gamma)}\Big\}dF(\theta\mid\gamma),$$

where $\bar{\omega}_i(\gamma) := \mathbb{E}[\omega_i \mid \gamma]$ as before. The regulation problem is to choose non-decreasing allocation q and transfer $t \in \mathbb{R}$ to maximize (71).

Given the properties of function k, there is an optimal t which solves $k'(t) = \bar{\omega}_i(\gamma)$ and satisfies $t > 0$.

At almost every $\theta \in \Theta$ where optimal $q(\theta)$ is strictly increasing, it should solve the pointwise first-order condition

$$0 = \bar{\omega}_i(\gamma)\Big[\theta v'(q(\theta)) - c'(q(\theta))\Big] - \underbrace{(\bar{\omega}_i(\gamma)-\mathbb{E}[\omega_i \mid \theta' \geq \theta,\gamma])\frac{1-F(\theta\mid\gamma)}{f(\theta\mid\gamma)}}_{B} v'(q(\theta)). \tag{72}$$

Lemma 11 implies that expression B in (72) is positive for all $\theta \in (\underline{\theta},\overline{\theta})$, and therefore quality $q(\theta)$ that solves the pointwise FOC (72) is strictly lower than the efficient quality $q^{FB}(\theta)$ for which $\theta v'(q^{FB}(\theta)) - c'(q^{FB}(\theta)) = 0$.

Then if in the optimal solution, $q(\theta) \geq q^{FB}(\theta)$ for some $\theta \in (\underline{\theta},\overline{\theta})$, it must be because the monotonicity constraint binds locally and hence q is constant around θ for some interval $[\theta_a,\theta_b]$ such that $\theta \in [\theta_a,\theta_b]$. But since the optimal allocation is continuous, the FOC (72) should hold at $\theta = \theta_a$, implying $q(\theta_a) < q^{FB}(\theta_a)$ which is a contradiction with $q^{FB}(\theta_a) < q^{FB}(\theta) \leq q(\theta) = q(\theta_a)$. This proves the result that for almost all $\theta \in (\underline{\theta},\overline{\theta})$,

$$q(\theta,\gamma) \ < \ \arg\max_q S(q,\theta).$$

Finally, let me prove the monotonicity of the mark-up $p(q(\theta)) - c(q(\theta))$. We have

$$p(q(\theta)) - c(q(\theta)) = \theta v(q(\theta)) - \int_{\underline{\theta}}^{\theta} v(q(z))dz - c(q(\theta)). \tag{73}$$

Note that q is absolutely continuous and hence almost everywhere differentiable, and at any point of differentiability, we have by (73),

$$(p'(q(\theta)) - c'(q(\theta)))q'(\theta) = q'(\theta)(\theta v'(q(\theta)) - c'(q(\theta))) \tag{74}$$

Derivative (74) is non-negative as by incentive compatibility, $q'(\theta) \geq 0$, and since $q(\theta) < q^{FB}(\theta)$, $\theta v'(q(\theta)) - c'(q(\theta)) > 0$. Integrating then yields that for all $\theta' > \theta$,

$$p(q(\theta')) - c(q(\theta')) \geq p(q(\theta)) - c(q(\theta)),$$

with strict inequality if $q(\theta') > q(\theta)$.

$\square$

Lemma 12 proves part (i) of Proposition 2.

Let us then proceed to part (ii) to study the regulator's problem of designing the regulation mechanism when γ is privately known to the firm.

The firm's profit is

$$\Pi(\gamma) = \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta,\gamma) - c(q(\theta,\gamma)))dF(\theta \mid \gamma) \tag{75}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta,\gamma) - c(q(\theta,\gamma)))(\gamma f_R(\theta) + (1-\gamma)f_P(\theta))d\theta \tag{76}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta,\gamma) - c(q(\theta,\gamma)))f_P(\theta))d\theta \tag{77}$$

$$+ \gamma \int_{\underline{\theta}}^{\overline{\theta}} (p(\theta,\gamma) - c(q(\theta,\gamma)))(f_R(\theta) - f_P(\theta))d\theta \tag{78}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} ((\theta - \frac{1-F_P(\theta)}{f_P(\theta)})v(q(\theta,\gamma)) - c(q(\theta,\gamma)))dF_P(\theta)) \tag{79}$$

$$+ \gamma \underbrace{\int_{\underline{\theta}}^{\overline{\theta}} (\theta v(q(\theta,\gamma)) - c(q(\theta,\gamma)))(f_R(\theta) - f_P(\theta)) + (F_R(\theta) - F_P(\theta))v(q(\theta,\gamma))d\theta,}_{:=R(q(\cdot,\gamma))} \tag{80}$$

where the final equality uses consumers' IC condition, and I introduce notation $R(q)$. By a standard Myersonian result, incentive compatibility for the firm is equivalent with R being non-decreasing in γ and

$$\Pi(\gamma) = \Pi(\underline{\gamma}) + \int_{\underline{\gamma}}^{\gamma} R(\gamma')d\gamma',$$

and consumer incentive compatibility further requires that q is non-decreasing in θ.

Using similar techniques as before, the regulator's payoff can again be written as

$$\int_{\underline{\gamma}}^{\overline{\gamma}} \Bigg\{ \int_{\underline{\theta}}^{\overline{\theta}} \Bigg[\bar{\omega}_i(\gamma)\Big[\theta v(q(\theta,\gamma)) - c(q(\theta,\gamma)) + t(\gamma)\Big]$$

$$- (\bar{\omega}_i(\gamma) - \mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma])v(q(\theta,\gamma))\frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\Bigg] dF(\theta \mid \gamma) \tag{81}$$

$$- (\bar{\omega}(\gamma) - \omega_F)\Pi(\underline{\gamma}) - (\bar{\omega}^+(\gamma) - \omega_F)\frac{1-G(\gamma)}{g(\gamma)}R(q(\cdot,\gamma)) - k(t(\gamma))\Bigg\} dG(\gamma),$$

and the regulator chooses an allocation $q : \Theta \times \Gamma$ and $\Pi(\underline{\gamma})$ to maximize (81) subject to the constraint that q is non-decreasing in θ and $\overline{R}$ is non-decreasing in γ, and the firm's IR constraint is satisfied: $\Pi(\underline{\gamma}) + \int_{\underline{\gamma}}^{\gamma} R(\gamma')d\gamma' \geq 0$ for all γ.

Lemma 13. *Denote*

$$W(\gamma, q) := \int_{\underline{\theta}}^{\overline{\theta}} \Bigg[\bar{\omega}_i(\gamma)\Big[\theta v(q(\theta,\gamma)) - c(q(\theta,\gamma))\Big] - (\bar{\omega}_i(\gamma) - \mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma])v(q(\theta,\gamma))\frac{1-F(\theta\mid\gamma)}{f(\theta\mid\gamma)}\Bigg]dF(\theta\mid\gamma), \quad (82)$$

Then $R(\gamma, q) < 0$ *implies* $W(\gamma, q) \leq W(\gamma, q^{FB})$.

Proof. Note that $F(\theta) = \gamma F_R(\theta) + (1-\gamma)F_P(\theta)$ and

$$\mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma] = \frac{\omega_R\gamma\,[1-F_R(\theta)] + \omega_P(1-\gamma)\,[1-F_P(\theta)]}{\gamma\,[1-F_R(\theta)] + (1-\gamma)\,[1-F_P(\theta)]}.$$

Therefore, we can write

$$W(\gamma, q) = \int_{\underline{\theta}}^{\overline{\theta}} \bar{\omega}(\gamma)\, S(q(\theta,\gamma),\theta)\, f(\theta\mid\gamma)\, d\theta - (\omega_P - \omega_R)\,\gamma(1-\gamma)\int_{\underline{\theta}}^{\overline{\theta}} (F_P - F_R)\, v(q)\, d\theta.$$

Furthermore,

$$W(\gamma, q) - W(\gamma, q^{\text{eff}}) = \bar{\omega}(\gamma)\int_{\underline{\theta}}^{\overline{\theta}} (S(q(\theta),\theta) - S(q^{FB}(\theta),\theta))f(\theta\mid\gamma)\, d\theta \quad (83)$$

$$-(\omega_P - \omega_R)\,\gamma(1-\gamma)\int_{\underline{\theta}}^{\overline{\theta}} (F_P - F_R)\,\big(v(q(\theta)) - v(q^{FB}(\theta))\big)\, d\theta.$$

Note that

$$\int_{\underline{\theta}}^{\overline{\theta}} (F_P - F_R)\,\big(v(q(\theta)) - v(q^{FB}(\theta))\big)\, d\theta = \int_{\underline{\theta}}^{\overline{\theta}} (S(q(\theta),\theta) - S(q^{FB}(\theta),\theta))(f_R(\theta) - f_P(\theta))d\theta - R(\gamma, q)$$

and substitute this to (83) to obtain

$$W(\gamma, q) - W(\gamma, q^{\text{eff}}) = \gamma\omega_R \int_{\underline{\theta}}^{\overline{\theta}} [S(q(\theta),\theta) - S(q^{FB}(\theta),\theta)]dF_R(\theta) \quad (84)$$

$$+ (1-\gamma)\omega_P \int_{\underline{\theta}}^{\overline{\theta}} [S(q(\theta),\theta) - S(q^{FB}(\theta),\theta)]dF_P(\theta) \quad (85)$$

$$+ (\omega_P - \omega_R)\,\gamma(1-\gamma)\,R(\gamma, q), \quad (86)$$

where the first two lines on the right-hand side are clearly negative, so

$$W(\gamma, q) - W(\gamma, q^{\text{eff}}) \;\leq\; (\omega_P - \omega_R)\,\gamma(1-\gamma)\,R(\gamma, q).$$

The statement of the lemma follows. ◻

Lemma 14. *In the continuous-WTP environment of Section 3.3, if the share of rich consumers $\gamma \in (0,1)$ is privately known to the firm, then in an optimal regulation mechanism, $R(\gamma) \geq 0$ for all γ and $\Pi(\underline{\gamma}) = 0$.*

Proof. Take a mechanism that satisfies the IC and IR constraints. Since incentive compatibility requires R to be non-decreasing. Hence, there must be a cut-off $\hat{\gamma} \in [\underline{\gamma}, \overline{\gamma}]$ such that at $R(\gamma) < 0$ for all $\gamma < \hat{\gamma}$ and $R(\gamma) \geq 0$ for all $\gamma > \hat{\gamma}$.

Now modify the mechanism to introduce cost-plus pricing at the bottom: in the new mechanism, $\tilde{q}(\theta, \gamma) = q^{FB}(\theta)$ for all $\gamma \leq \hat{\gamma}$ and all θ, and the lowest firm type's profit is

$$\tilde{\Pi}(\underline{\gamma}) = 0;$$

the allocation for $\gamma > \hat{\gamma}$ remains the same. Note that then $\tilde{R}(\gamma) = 0$ for all $\gamma \leq \hat{\gamma}$. The arising mechanism clearly satisfies the condition that the quality allocation must be non-decreasing in θ (consumers' IC). It also satisfies the firm's IC constraint that $\hat{R}$ is non-decreasing as $\hat{R}(\gamma) = R(\gamma) \geq 0$ for $\gamma > \hat{\gamma}$, and the firm's IR constraint remains satisfied as well. By Lemma 13, the modification increases the two first lines in the regulator's payoff (81), and it clearly improves the third line (as the firm's profit is minimized for all $\gamma \leq \hat{\gamma}$, while for all $\gamma > \hat{\gamma}$, $\Pi(\gamma) - \Pi(\hat{\gamma}) = \tilde{\Pi}(\gamma) - \tilde{\Pi}(\hat{\gamma})$.

□

Since

$$\mathbb{E}[\omega_i \mid \gamma] - \mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma] = \frac{\gamma(1-\gamma)\,(F_P(\theta) - F_R(\theta))\,(\omega_P - \omega_R)}{1 - F(\theta \mid \gamma)},$$

(as derived in the proof of Lemma 11), and $\Pi(\underline{\gamma}) \geq 0$ (as derived in 14) we can rewrite the regulator's objective (81) as

$$\begin{aligned}
&\int_{\underline{\gamma}}^{\overline{\gamma}} \int_{\underline{\theta}}^{\overline{\theta}} \Bigg[\bar{\omega}_i(\gamma) \Big[S(q(\theta,\gamma),\theta) + t(\gamma) \Big] f(\theta \mid \gamma) - k(t(\gamma)) \\
&- \gamma(1-\gamma)\,(F_P(\theta) - F_R(\theta))\,(\omega_P - \omega_R) v(q(\theta,\gamma)) \qquad (87) \\
&- (\bar{\omega}^+(\gamma) - \omega_F) \frac{1 - G(\gamma)}{g(\gamma)} \Big[S(q(\theta,\gamma),\theta)(f_R(\theta) - f_P(\theta)) + (F_R(\theta) - F_P(\theta)) v(q(\theta,\gamma)) \Big] \Bigg] d\theta dG(\gamma),
\end{aligned}$$

Using Lemma 14, the regulator's problem is to choose $t : \Gamma \to \mathbb{R}$ and $q : \Theta \times \Gamma \to \mathcal{Q}$ to maximize (88) so that q is non-decreasing in θ and

$$R(q(\cdot,\gamma)) := \int_{\underline{\theta}}^{\overline{\theta}} \Big[S(q(\theta,\gamma),\theta)(f_R(\theta) - f_P(\theta)) + (F_R(\theta) - F_P(\theta)) v(q(\theta,\gamma)) \Big] d\theta$$

is non-decreasing in γ and non-negative. Clearly any optimal solution must again have $k'(t(\gamma)) = \bar{\omega}_i(\gamma)$ for almost all γ.

In the following lemma, recall the definition

$$\delta(\gamma) := \gamma(1-\gamma)(\omega_P - \omega_R) - (\bar{\omega}^+(\gamma) - \omega_F)\frac{1-G(\gamma)}{g(\gamma)}.$$

Lemma 15. *In the continuous-WTP environment of Section 3.3, take γ such that $\delta(\gamma) < 0$. Consider the "relaxed slice problem" of choosing $q : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}_+$ to maximize*

$$\begin{aligned} A(q) := \int_{\underline{\theta}}^{\overline{\theta}} \Big[& \bar{\omega}_i(\gamma)\Big[S(q(\theta),\theta)\Big] f(\theta \mid \gamma) \\ & - \gamma(1-\gamma)\left(F_P(\theta) - F_R(\theta)\right)(\omega_P - \omega_R) v(q(\theta)) \qquad (88) \\ & - (\bar{\omega}^+(\gamma) - \omega_F)\frac{1-G(\gamma)}{g(\gamma)}\Big[S(q(\theta),\theta)(f_R(\theta) - f_P(\theta)) + (F_R(\theta) - F_P(\theta))v(q(\theta))\Big]\Big] d\theta \qquad (89) \end{aligned}$$

subject to constraint

$$\underbrace{\int_{\underline{\theta}}^{\overline{\theta}} \Big[S(q(\theta),\theta)(f_R(\theta) - f_P(\theta)) + (F_R(\theta) - F_P(\theta))v(q(\theta))\Big] d\theta}_{:=R(q)} \geq 0. \qquad (90)$$

In any optimal solution, $q(\theta) = q^{FB}(\theta) := \arg\max_q S(q,\theta)$ for almost all $\theta \in [\underline{\theta}, \overline{\theta}]$.

Proof. Denote

$$\Delta F(\theta) := F_R(\theta) - F_P(\theta), \quad \Delta f(\theta) := f_R(\theta) - f_P(\theta), \quad S^{\text{eff}}(\theta) := S(q^{FB}(\theta),\theta).$$

and notice that the objective is denoted by W and the constraint function by R. Recall that we have the first-order stochastic dominance $\Delta F(\theta) < 0$ for all $\theta \in (\underline{\theta}, \overline{\theta})$.

By the envelope theorem, for almost all θ,

$$(S^{\text{eff}})'(\theta) = \frac{\partial}{\partial\theta} S(q^{FB}(\theta),\theta) = v(q^{FB}(\theta)), \qquad (91)$$

and note that

$$\begin{aligned} R(q^{FB}) &= \int_{\underline{\theta}}^{\overline{\theta}} S^{\text{eff}}(\theta)\,\Delta f(\theta)\,d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \Delta F(\theta)\,v(q^{FB}(\theta))\,d\theta \\ &= \int_{\underline{\theta}}^{\overline{\theta}} S^{\text{eff}}(\theta)\,\Delta f(\theta)\,d\theta + \int_{\underline{\theta}}^{\overline{\theta}} \Delta F(\theta)\,(S^{\text{eff}})'(\theta)\,d\theta \\ &= 0, \end{aligned}$$

where the second equality follows from (91) and the third equality follows from integration by parts (using $\Delta F(\underline{\theta}) = \Delta F(\overline{\theta}) = 0$). So, as intuitively clear, constraint (90) binds at $q = q^{FB}$.

The Lagrangian of the problem is

$$\mathcal{L}(q, \mu) = A(q) + \mu R(q).$$

with KKT multiplier $\mu \geq 0$. Define

$$\mu^* = -\delta(\gamma) > 0,$$

where the inequality follows by assumption. We have

$$\mathcal{L}(q, \mu^*) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\gamma \omega_R f_R(\theta) + (1-\gamma)\omega_P f_P(\theta)\right] S(q(\theta), \theta)\, d\theta.$$

For any feasible q,

$$A(q) \leq \mathcal{L}(q, \mu).$$

Since $R(q^{FB}) = 0$, we have $A(q^{FB}) = \mathcal{L}(q^{FB}, \mu)$. Hence

$$\begin{aligned}
A(q^{FB}) - A(q) &\geq \mathcal{L}(q^{FB}, \mu^*) - \mathcal{L}(q, \mu^*) \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left[\gamma \omega_R f_R(\theta) + (1-\gamma)\omega_P f_P(\theta)\right] \left(S^{\text{eff}}(\theta) - S(q(\theta), \theta)\right) d\theta \\
&\geq 0
\end{aligned}$$

where the last inequality is strict unless $q(\theta) = q^{FB}(\theta)$ for almost all θ. The claim of the lemma follows.

□

Lemma 16. *If γ is privately known to the firm, then there is a cut-off $\gamma^* \geq \inf\{\gamma \in [\underline{\gamma}, \overline{\gamma}] : \ \delta(\gamma) \geq 0\} > \underline{\gamma}$ such that any optimal regulation mechanism has cost-plus pricing for almost all $\gamma \leq \gamma^*$, $\theta \in [\underline{\theta}, \overline{\theta}]$:*

$$q(\theta, \gamma) = \arg\max_q S(q, \theta), \qquad p(q(\theta, \gamma)) = c(q(\theta, \gamma)) - t(\gamma).$$

Proof. The regulator's problem is to choose $q : \Theta \times \Gamma \to \mathcal{Q}$ to maximize (88) so that q is non-decreasing in θ and $R(q(\cdot, \gamma))$ is non-negative and non-decreasing for all γ.

Recall $\delta = \gamma(1-\gamma)(\omega_P - \omega_R) - (\bar{\omega}^+(\gamma) - \omega_F)\frac{1-G(\gamma)}{g(\gamma)}$, and let

$$\gamma^* := \inf\{\gamma \in [\underline{\gamma}, \overline{\gamma}] : \ \delta(\gamma) \geq 0\}.$$

Note that $\gamma^* > \underline{\gamma}$ by continuity of δ and since $\delta(\underline{\gamma}) < 0$.

Consider a relaxed problem that ignores the monotonicity constraints. By Lemma 15, for almost all $\gamma \le \gamma^*$ and $\theta \in [\underline{\theta}, \overline{\theta}]$, the solution must satisfy $q(\theta, \gamma) = q^{FB}(\theta)$. Furthermore, the choice $q(\theta, \gamma) = q^{FB}(\theta)$ for all $\gamma \le \gamma^*$ clearly satisfies the monotonicity constraints for all $\gamma \le \gamma^*$, and any choice of allocation $q(\cdot, \gamma)$ for all $\gamma > \gamma^*$ that can ever be part of a feasible solution remains feasible with the efficient allocation for $\gamma \le \gamma^*$. Therefore, any optimal regulation mechanism must have efficient allocation for almost all $\gamma \le \gamma^*$, which together with $\Pi(\underline{\gamma}) = 0$ implies cost-plus pricing. □

Part (ii) of Proposition 2 follows directly from Lemmas 14 and 16.

A.13 Proof of Lemma 4

Proof. As noted in the text, the firm's selling problem can be written as

$$\max_{q(\cdot) \text{ non-decreasing}} \int_{\underline{\theta}}^{\overline{\theta}} \Big[\underbrace{\psi(\theta \mid \gamma)\, v(q(\theta)) - c(q(\theta)) + \tau(q(\theta))}_{:= \pi(\theta, q(\theta))} \Big] \, dF(\theta \mid \gamma). \tag{92}$$

Since

$$\frac{\partial^2}{\partial \theta\, \partial q} \pi(\theta, q) \;=\; \psi'(\theta \mid \gamma)\, v'(q) \;>\; 0,$$

π has strictly increasing differences in (θ, q). By Topkis's theorem (Milgrom and Shannon, 1994; Topkis, 1998), the pointwise argmax correspondence $Q^*(\theta) = \arg\max_{q \in Q} \pi(\theta, q)$ is nondecreasing in the strong set order. It follows that the monotonicity constraint in (92) does not bind: a non-decreasing $q(\cdot)$ with $q(\theta) \in Q^*(\theta)$ for all θ solves problem (92).

By the envelope theorem for monotone choice (Milgrom and Segal, 2002), given that the firm chooses q to maximize profit,

$$\pi\big(\theta, q(\theta)\big) \;=\; \pi\big(\underline{\theta}, q(\underline{\theta})\big) \;+\; \int_{\underline{\theta}}^{\theta} \psi'(z \mid \gamma)\, v\big(q(z)\big)\, dz. \tag{93}$$

Integrating (93) with respect to $F(\cdot \mid \gamma)$ and using integration by parts gives

$$\begin{aligned}
\Pi(\gamma) &= \int_{\underline{\theta}}^{\overline{\theta}} \pi\big(\theta, q(\theta)\big)\, dF(\theta \mid \gamma) \\
&= \pi\big(\underline{\theta}, q(\underline{\theta})\big) \;+\; \int_{\underline{\theta}}^{\overline{\theta}} \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, \psi'(\theta \mid \gamma)\, v\big(q(\theta)\big)\, dF(\theta \mid \gamma).
\end{aligned} \tag{94}$$

Rearranging (93) yields the formula

$$\tau\big(q(\theta)\big) \;=\; \pi\big(\underline{\theta}, q(\underline{\theta})\big) \;-\; \psi(\theta \mid \gamma)\, v\big(q(\theta)\big) \;+\; c\big(q(\theta)\big) \;+\; \int_{\underline{\theta}}^{\theta} \psi'(z \mid \gamma)\, v\big(q(z)\big)\, dz. \tag{95}$$

To see that *any* weakly increasing q is implementable, fix such a q. Define τ on the image $q[\Theta]$ via (95), and set $\tau(q) = \underline{\tau}$ for $q \in \mathcal{Q} \setminus q[\Theta]$. This is well-defined: if $q(\theta) = q(\theta')$ with $\theta' < \theta$, then q is constant on $[\theta', \theta]$, and by the fundamental theorem of calculus,

$$\tau\big(q(\theta)\big) - \tau\big(q(\theta')\big) = \big(\psi(\theta') - \psi(\theta)\big)v\big(q(\theta)\big) + \int_{\theta'}^{\theta} \psi'(z \mid \gamma)\, v\big(q(\theta)\big)\, dz = 0.$$

Finally, for $\underline{\tau}$ sufficiently small, deviating to $q' \notin q[\Theta]$ is unprofitable:

$$\underline{\tau} \;\le\; \pi\big(\underline{\theta}, q(\underline{\theta})\big) - \sup_{(\theta,q)\in[\underline{\theta},\overline{\theta}]\times\mathcal{Q}} \{\psi(\theta \mid \gamma)v(q) - c(q)\}$$

ensures $\pi(\theta, q(\theta)) \ge \pi(\theta, q')$. If the deviation is to some $q(\theta') \in q[\Theta]$, then

$$\pi\big(\theta, q(\theta)\big) - \pi\big(\theta, q(\theta')\big) = \int_{\theta'}^{\theta} \psi'(z \mid \gamma)\, \big[v\big(q(z)\big) - v\big(q(\theta')\big)\big]\, dz \;\ge\; 0,$$

using that $\psi' \ge 0$, v and q are nondecreasing. Thus q is the firm's profit-maximizing schedule under τ. □

A.14 Proof of Proposition 3

Proof. Express the regulator's objective in terms of $(q, \pi(\underline{\theta}))$. By Lemma 4,

$$\Pi(\gamma) \;=\; \pi(\underline{\theta}, q(\underline{\theta})) \;+\; \int_{\underline{\theta}}^{\overline{\theta}} \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, \psi'(\theta \mid \gamma)\, v\big(q(\theta)\big)\, dF(\theta \mid \gamma).$$

Integrating (17) over $F(\cdot \mid \gamma)$ gives the regulator's total spending

$$\begin{aligned} T(q, \pi(\underline{\theta})) &:= \int_{\underline{\theta}}^{\overline{\theta}} \tau\big(q(\theta)\big)\, dF(\theta \mid \gamma) \\ &= \pi(\underline{\theta}) \;+\; \int_{\underline{\theta}}^{\overline{\theta}} \Big[\frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, \psi'(\theta \mid \gamma)\, v(q(\theta)) - \psi(\theta \mid \gamma)\, v(q(\theta)) + c(q(\theta))\Big]\, dF(\theta \mid \gamma). \end{aligned}$$

Consumer utility contributes

$$\begin{aligned} \int \omega\, u(\theta)\, d\tilde{F}(\theta, \omega) &= \int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta] \int_{\underline{\theta}}^{\theta} v\big(q(z)\big)\, dz\, dF(\theta \mid \gamma) \\ &= \int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta' \ge \theta]\, \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, v\big(q(\theta)\big)\, dF(\theta \mid \gamma). \end{aligned}$$

Hence the regulator chooses nondecreasing q, $\pi(\underline{\theta}) \geq 0$, and $T \in \mathbb{R}$ to maximize

$$\int_{\underline{\theta}}^{\overline{\theta}} \Big\{ \mathbb{E}[\omega_i \mid \theta' \geq \theta] \, \frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \, v(q(\theta)) + \omega_F \Big[\pi(\underline{\theta}) + \frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \, \psi'(\theta \mid \gamma) \, v(q(\theta)) \Big] - k(T) \Big\} dF(\theta \mid \gamma)$$

subject to

$$T \;=\; \int_{\underline{\theta}}^{\overline{\theta}} \Big[\pi(\underline{\theta}) + \frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \, \psi'(\theta \mid \gamma) \, v(q(\theta)) - \psi(\theta \mid \gamma) \, v(q(\theta)) + c(q(\theta)) \Big] \, dF(\theta \mid \gamma). \tag{96}$$

By standard maximum principles for variational problems with an integral constraint (e.g., Hellwig, 2008), at almost every θ where q is strictly increasing the stationarity condition is

$$\begin{aligned} 0 \;=\; & \mathbb{E}[\omega_i \mid \theta' \geq \theta] \, \frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \, v'(q(\theta)) + \omega_F \, \frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \, \psi'(\theta \mid \gamma) \, v'(q(\theta)) \\ & - \kappa \Big[\frac{1-F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \, \psi'(\theta \mid \gamma) \, v'(q(\theta)) - \psi(\theta \mid \gamma) \, v'(q(\theta)) + c'(q(\theta)) \Big], \end{aligned} \tag{97}$$

where $\kappa = k'(T)$. Rearrangement gives the three-term decomposition in the proposition. The FOC for T yields $\kappa = k'(T)$; the FOC for $\pi(\underline{\theta})$ yields $\omega_F - \kappa + \mu = 0$ with complementary slackness $\mu \cdot \pi(\underline{\theta}) = 0$, hence either $\kappa = \omega_F$ and $\pi(\underline{\theta}) \geq 0$, or $\kappa > \omega_F$ and $\pi(\underline{\theta}) = 0$.

At almost every point where q is strictly increasing, it is differentiable; differentiating (17) in θ gives

$$\tau'\big(q(\theta)\big) \, q'(\theta) \;=\; -\psi(\theta \mid \gamma) \, v'(q(\theta)) \, q'(\theta) + c'(q(\theta)) \, q'(\theta),$$

so (where $q'(\theta) \neq 0$)

$$\tau'\big(q(\theta)\big) \;=\; -\psi(\theta \mid \gamma) \, v'(q(\theta)) \;+\; c'(q(\theta)).$$

Substituting this into (97) yields the marginal-subsidy formula (18).

Finally, I will verify the existence of a solution to the regulator's optimization problem. First, since v is concave with $v'(q) \to 0$ as $q \to \infty$ and c is convex with $c'(q) > 0$, there exists a finite $\overline{q}(\kappa)$ such that

$$v'\Big(\kappa \, \theta - (\kappa - \mathbb{E}[\omega_i \mid \theta' \geq \theta]) \, \frac{1-F}{f} - (\kappa - \omega_F) \, \frac{1-F}{f} \Big) - \kappa c' < 0$$

for all $q \geq \overline{q}(\kappa)$ for all θ. Therefore, for any $\kappa > 0$, we may restrict attention to compact interval of qualities $[0, \overline{q}(\kappa)]$. With the uniform bound $\overline{q}(\kappa)$, the set of non-decreasing functions $q : \Theta \to [0, \overline{q}(\kappa)]$ is (sequentially) compact in L^1. Furthermore, since k satisfies $k'(T) \to \infty$ as $T \to \infty$ and $\pi(\underline{\theta}) \geq 0$, we can also restrict $\pi(\underline{\theta})$ and T to some compact intervals $[0, \overline{\pi}]$ and $[\underline{T}, \overline{T}]$, respectively.

All integrands are continuous and bounded, so by dominated convergence, the regulator's objective is continuous on a compact set. By the Weierstrass argument, the existence of a solution follows.

□

A.15 Proof of Lemma 5

Proof. Let $\overline{q}(\cdot) := q(\cdot|\overline{\gamma})$ denote the allocation in the highest demand state $\overline{\gamma}$. It follows directly from Lemma 4 in the known-γ environment that there exists a subsidy schedule τ such that $\overline{q}$ is implemented in state $\overline{\gamma}$.

Define then virtual–value matching map

$$M(\theta, \gamma) \ := \ \psi^{-1}\big(\psi(\theta \mid \gamma) \mid \overline{\gamma}\big).$$

Since ψ is continuous and strictly increasing in both θ and γ for all $[\underline{\theta}, \overline{\theta}) \times \Gamma$, with $\psi(\overline{\theta}, \gamma)$ constant in γ, $M : [\underline{\theta}, \overline{\theta}] \times \Gamma \to [\underline{\theta}, \overline{\theta}]$ is a well-defined surjective function. Note that $\psi\big(M(\theta, \gamma) \mid \overline{\gamma}\big) = \psi(\theta \mid \gamma)$.

Now take any equilibrium quality allocation q. By the proof of Lemma 4, q must be such that

$$q(\theta, \gamma) \in \arg\max_{q \geq 0} \{\psi(\theta \mid \gamma) v(q) - c(q) + \tau(q)\}$$

for almost all (θ, γ).

The argmax sets clearly coincide in the following way for all θ, γ:

$$A(\theta, \gamma) := \arg\max_{q \geq 0} \{\psi(\theta \mid \gamma) v(q) - c(q) + \tau(q)\} \ = \ \arg\max_{q \geq 0} \{\psi(M(\theta, \gamma) \mid \overline{\gamma}) v(q) - c(q) + \tau(q)\}.$$

This implies that for all (θ, γ), $\overline{q}(M(\theta, \gamma))$ is one of the profit-maximizing quality choices for the firm at (θ, γ).

However, even though the argmax sets coincide, we can have $q(\theta, \gamma) \neq q(M(\theta, \gamma), \overline{\gamma})$ if the argmax sets are not singletons. Yet, the monotone selection theorem of Milgrom and Shannon (1994) implies that since $\psi(M(\theta, \gamma) \mid \overline{\gamma}) v(q) - c(q) + \tau(q)$ has strictly increasing differences in (θ, q), *any* selection $q^*_\gamma(\theta) \in \arg\max_{q \geq 0} \psi(\theta \mid \gamma) v(q) - c(q) + \tau(q)$ is increasing in θ, so for any $\theta' < \theta < \theta''$,

$$q(\theta', \gamma) \ \leq \ \inf A(\theta, \gamma) \quad \text{and} \quad q(\theta'', \gamma) \ \geq \ \sup A(\theta, \gamma).$$

Therefore, if $A(\theta, \gamma)$ contains more than one point, allocation $q(\cdot, \gamma)$ must have a jump at θ. Since every monotone real-valued function has at most countably many points of discontinuity, the set $\{\theta : |A(\theta, \gamma)| > 1\}$ is countable. Outside this countable set, $A(\theta, \gamma)$ must be a singleton and $q(\theta, \gamma) = \overline{q}\big(M(\theta, \gamma)\big)$.

Therefore, for all γ and almost every θ,

$$q(\theta, \gamma) = \overline{q}(\psi^{-1}(\psi(\theta \mid \gamma) \mid \overline{\gamma})).$$

Finally, from Section 4.1, we know that for any fixed state γ, allocation $q(\cdot, \gamma)$ and base-level rent $\pi(\underline{\theta}, \gamma)$ determine all payoffs. We have now proved that $\overline{q}(\cdot)$ determines allocation $q(\cdot, \gamma)$ for any state. Furthermore, $\pi(\underline{\theta}, \overline{\gamma})$ together with $\overline{q}(\cdot)$ determines $\pi(\underline{\theta}, \gamma)$ for all γ, because

$$\pi(\underline{\theta}, \gamma) = \pi\big(\psi^{-1}\big(\psi(\underline{\theta} \mid \gamma) \mid \overline{\gamma}\big), \overline{\gamma}\big) = \pi(\underline{\theta}, \overline{\gamma}) + \int_{\underline{\theta}}^{\psi^{-1}\left(\psi(\underline{\theta}|\gamma)|\overline{\gamma}\right)} \psi'(z \mid \overline{\gamma})\, v\big(\overline{q}(z)\big)\, dz, \tag{98}$$

by the envelope theorem result already discussed. This proves the final claim of the lemma that the allocation $\overline{q}$ together with base-level rent $\pi(\underline{\theta}, \overline{\gamma}) \geq 0$ pins down all payoffs.

□

A.16 Proof of Proposition 4

Proof. By Lemma 5, we know that the allocation in the highest demand state $\overline{\gamma}$, $\overline{q}(\cdot) := q(\cdot|\overline{\gamma})$, together with base-level rent $\pi(\underline{\theta}, \overline{\gamma})$, fixes the whole allocation q and all the payoffs. Proposition 4 characterizes the optimal choice of $\overline{q}$ and $\pi(\underline{\theta}, \overline{\gamma})$. To derive the characterization, we must derive the components of the regulator's payoff—the firm's profit, each consumer type's payoff, and the regulator's spending in each state γ as a function of $\overline{q}$ and $\pi(\underline{\theta}, \overline{\gamma})$.

Recall that I make the following definitions in the text:

$$L(\theta, \gamma) := \psi^{-1}\big(\psi(\theta \mid \overline{\gamma}) \mid \gamma\big) \text{ for all } \theta \in [\underline{\theta}, \overline{\theta}], \gamma \in [\hat{\gamma}(\theta), \overline{\gamma}],$$
$$\hat{\gamma}(\theta) := (\psi(\underline{\theta}|\cdot))^{-1}(\min\{\psi(\theta|\overline{\gamma}), \psi(\underline{\theta}|\underline{\gamma})\}).$$

It is also helpful to make the following additional definition for this proof:

$$\theta_{\min}(\gamma) \;:=\; \psi^{-1}\big(\psi(\underline{\theta} \mid \gamma) \mid \overline{\gamma}\big).$$

By Lemma 4, for any state γ, we can write the firm's profit in state γ as

$$\Pi(\gamma) \;=\; \pi(\underline{\theta}, \gamma) \;+\; \int_{\underline{\theta}}^{\overline{\theta}} \frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)}\, \psi'(\theta \mid \gamma)\, v\big(q(\theta, \gamma)\big)\, dF(\theta \mid \gamma). \tag{99}$$

I will now perform a change variables using $\theta = L(\tilde{\theta}, \gamma)$. Note that

$$\frac{\partial L(\tilde{\theta}, \gamma)}{\partial \tilde{\theta}} \;=\; \frac{\psi'(\tilde{\theta} \mid \overline{\gamma})}{\psi'\big(L(\tilde{\theta}, \gamma) \mid \gamma\big)}. \tag{100}$$

With the change of variables, we can then write the state-γ firm's profit (99) as

$$\Pi(\gamma) = \pi(\underline{\theta}, \gamma) + \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} (1 - F(L(\tilde{\theta}, \gamma) \mid \gamma))\, \psi'(\tilde{\theta} \mid \overline{\gamma})\, v\big(\overline{q}(\tilde{\theta})\big)\, d\tilde{\theta} \tag{101}$$

$$= \pi(\underline{\theta}, \overline{\gamma}) + \int_{\underline{\theta}}^{\theta_{\min}(\gamma)} \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big)\, d\theta, \tag{102}$$

$$+ \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} (1 - F(L(\theta, \gamma) \mid \gamma))\, \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big)\, d\theta$$

where the second equality plugs in (98) to express $\pi(\underline{\theta}, \gamma)$ in terms of $\overline{q}$ and $\pi(\underline{\theta}, \overline{\gamma})$.

Let us then solve for the consumers' contribution to the regulator's payoff in state γ:

$$B(\gamma) := \int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta, \gamma]\Big(\int_{\underline{\theta}}^{\theta} v\big(q(z, \gamma)\big)\, dz\Big)\, dF(\theta \mid \gamma)$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta' \geq \theta, \gamma](1 - F(\theta|\gamma)) v\big(q(\theta, \gamma)\big) d\theta$$

$$= \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta' \geq L(\theta, \gamma), \gamma](1 - F(L(\theta, \gamma)|\gamma)) \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'\big(L(\theta, \gamma) \mid \gamma\big)} v\big(\overline{q}(\theta)\big)\, d\theta \tag{103}$$

where the second equality uses integration by parts and the third equality performs a similar change of variables as was done in (101).

Finally, the regulator's spending in state γ is

$$T(\gamma) = \int_{\underline{\theta}}^{\overline{\theta}} \tau\big(q(\theta, \gamma)\big)\, dF(\theta \mid \gamma)$$

$$= \pi(\underline{\theta}, \gamma)$$

$$+ \int_{\underline{\theta}}^{\overline{\theta}} \Big[\Big(\frac{1 - F(\theta \mid \gamma)}{f(\theta \mid \gamma)} \psi'(\theta \mid \gamma) - \psi(\theta \mid \gamma)\Big) v(q(\theta, \gamma)) + c(q(\theta, \gamma))\Big] dF(\theta \mid \gamma)$$

$$= \pi(\underline{\theta}, \overline{\gamma}) + \int_{\underline{\theta}}^{\theta_{\min}(\gamma)} \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big)\, d\theta$$

$$+ \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} \Big[\Big(\frac{1 - F(L(\theta, \gamma) \mid \gamma)}{f(L(\theta, \gamma) \mid \gamma)} \psi'(L(\theta, \gamma) \mid \gamma) - \psi(L(\theta, \gamma) \mid \gamma)\Big) v(\overline{q}(\theta)) + c(\overline{q}(\theta))\Big] \tag{104}$$

$$\frac{\psi'(\theta \mid \overline{\gamma})}{\psi'\big(L(\theta, \gamma) \mid \gamma\big)} f(L(\theta, \gamma) \mid \gamma) d\theta,$$

where the second equality uses the formula for the regulator's spending in

Lemma 4, and the third equality plugs in (98) and performs a similar change of variables as previously.

The regulator's subsidy design problem is then to choose non-decreasing allocation $\overline{q} : \Theta \to \boldsymbol{Q}$, base-level rent $\pi(\underline{\theta}, \overline{\gamma}) \geq 0$ and spending $T(\gamma)$ for each state γ to maximize

$$\int_{\underline{\gamma}}^{\overline{\gamma}} \Big[\omega_F \Pi(\gamma) + B(\gamma) - k(T(\gamma))\Big] dG(\gamma) \tag{105}$$

$$= \int_{\underline{\gamma}}^{\overline{\gamma}} \Big[\omega_F \Big(\pi(\underline{\theta}, \overline{\gamma}) + \int_{\underline{\theta}}^{\theta_{\min}(\gamma)} \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big)\, d\theta \tag{106}$$

$$+ \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} (1 - F(L(\theta, \gamma) \mid \gamma))\, \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big)\, d\theta\Big) \tag{107}$$

$$+ \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} \mathbb{E}[\omega_i \mid \theta' \geq L(\theta, \gamma), \gamma](1 - F(L(\theta, \gamma)|\gamma)) \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'\big(L(\theta, \gamma) \mid \gamma\big)} v\big(\overline{q}(\theta)\big)\, d\theta \tag{108}$$

$$- k(T(\gamma))\Big] dG(\gamma) \tag{109}$$

$$= \omega_F \pi(\underline{\theta}, \overline{\gamma}) + \int_{\underline{\theta}}^{\overline{\theta}} \Big[\omega_F \Big(\int_{\underline{\gamma}}^{\hat{\gamma}(\theta)} \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big) dG(\gamma) \tag{110}$$

$$+ \int_{\hat{\gamma}(\theta)}^{\overline{\gamma}} (1 - F(L(\theta, \gamma) \mid \gamma))\, \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big) dG(\gamma)\Big) \tag{111}$$

$$+ \int_{\hat{\gamma}(\theta)}^{\overline{\gamma}} \mathbb{E}[\omega_i \mid \theta' \geq L(\theta, \gamma), \gamma](1 - F(L(\theta, \gamma)|\gamma)) \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'\big(L(\theta, \gamma) \mid \gamma\big)} v\big(\overline{q}(\theta)\big)\, dG(\gamma) \tag{112}$$

$$- \int_{\underline{\gamma}}^{\overline{\gamma}} k(T(\gamma)) dG(\gamma)\Big] d\theta \tag{113}$$

(where the second equality changes the order of integration) subject to constraints

$$\begin{aligned} T(\gamma) =& \pi(\underline{\theta}, \overline{\gamma}) + \int_{\underline{\theta}}^{\theta_{\min}(\gamma)} \psi'(\theta \mid \overline{\gamma})\, v\big(\overline{q}(\theta)\big)\, d\theta \\ &+ \int_{\theta_{\min}(\gamma)}^{\overline{\theta}} \Big[\Big(\frac{1 - F(L(\theta, \gamma) \mid \gamma)}{f(L(\theta, \gamma) \mid \gamma)} \psi'(L(\theta, \gamma) \mid \gamma) - \psi(L(\theta, \gamma) \mid \gamma)\Big) v(\overline{q}(\theta)) + c(\overline{q}(\theta))\Big] \end{aligned} \tag{114}$$

$$\frac{\psi'(\theta \mid \overline{\gamma})}{\psi'\big(L(\theta,\gamma) \mid \gamma\big)} f(L(\theta,\gamma) \mid \gamma) d\theta,$$

for all $\gamma \in \Gamma$.

Necessary conditions for a solution to a problem of this form are presented, e.g., in Hellwig (2008). In particular, at almost every $\theta \in \Theta$, the following necessary conditions are obtained:

– Stationarity condition for $\overline{q}(\theta)$, satisfied at almost every $\theta \in \Theta$ where optimal allocation $\overline{q}$ is strictly increasing:

$$\begin{aligned} 0 = \int_{\hat{\gamma}(\theta)}^{\overline{\gamma}} \Big\{ & \kappa(\gamma) \left[Lv'(\overline{q}) - c'(\overline{q})\right] \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'(L \mid \gamma)} \\ & - (\kappa(\gamma) - \mathbb{E}[\omega_i \mid \theta' \geq L, \gamma]) \frac{1 - F(L|\gamma)}{f(L|\gamma)} \frac{\psi'(\theta \mid \overline{\gamma})}{\psi'(L \mid \gamma)} v'(\overline{q}) \\ & - (\kappa(\gamma) - \omega_F) \frac{1 - F(L|\gamma)}{f(L|\gamma)} \psi'(\theta|\overline{\gamma})\, v'(\overline{q}) \Big\} f(L|\gamma) dG(\gamma) \\ & - \int_{\underline{\gamma}}^{\hat{\gamma}(\theta)} (\kappa(\gamma) - \omega_F) \psi'(\theta \mid \overline{\gamma}) v'(\overline{q}) dG(\gamma). \end{aligned} \tag{115}$$

– First-order condition for T:

$$k'(T(\gamma)) = \kappa(\gamma)$$

– First-order condition for $\pi(\underline{\theta}, \overline{\gamma})$:

$$\omega_F - \int \kappa(\gamma) dG(\gamma) + \mu = 0 \text{ for some } \mu \geq 0$$

with complementary slackness $\mu\pi(\underline{\theta}, \overline{\gamma}) = 0$.

Hence e, either $\int \kappa(\gamma)dG(\gamma) = \omega_F$ and $\pi(\underline{\theta}, \overline{\gamma}) \geq 0$, or $\int \kappa(\gamma)dG(\gamma) > \omega_F$ and $\pi(\underline{\theta}, \overline{\gamma}) = 0$, as stated in the proposition.

Finally, I will verify the existence of a solution to the regulator's optimization problem. First, since v is concave with $v'(q) \to 0$ as $q \to \infty$ and c is convex with $c'(q) > 0$, there exists a finite $\hat{q}(\kappa)$ such that the right-hand side of (115) is negative for all $q \geq \hat{q}(\kappa)$, so that we may restrict attention to compact interval of qualities $[0, \hat{q}(\kappa)]$. With the uniform bound $\hat{q}(\kappa)$, the set of non-decreasing functions $\overline{q} : \Theta \to [0, \hat{q}(\kappa)]$ is (sequentially) compact in L^1. Furthermore, since k satisfies $k'(T) \to \infty$ as $T \to \infty$ and $\pi(\underline{\theta}, \overline{\gamma}) \geq 0$, we can also restrict $\pi(\underline{\theta}, \overline{\gamma})$ and $T(\gamma)$ to some compact intervals $[0, \overline{\pi}]$ and $[\underline{T}, \overline{T}]$, respectively. All integrands are continuous and bounded, so by dominated convergence, the regulator's objective is continuous on a compact set. By the Weierstrass argument, the existence of a solution follows.

□

References

Ahlvik, Lassi, Matti Liski, and Mikael Mäkimattila (2024) "Pigouvian income taxation."

Akbarpour, Mohammad, Piotr Dworczak, and Scott Duke Kominers (2024) "Redistributive allocation mechanisms," *Journal of Political Economy*, 132 (6), 1831–1875.

Alonso, Ricardo and Niko Matouschek (2008) "Optimal delegation," *The Review of Economic Studies*, 75 (1), 259–293.

Armstrong, Mark and David EM Sappington (2007) "Recent developments in the theory of regulation," *Handbook of industrial organization*, 3, 1557–1700.

Baron, David P and Roger B Myerson (1982) "Regulating a monopolist with unknown costs," *Econometrica: Journal of the Econometric Society*, 911–930.

Besanko, David, Shabtai Donnenfeld, and Lawrence J White (1987) "Monopoly and quality distortion: effects and remedies," *The Quarterly Journal of Economics*, 102 (4), 743–767.

Borenstein, Severin (2012) "The redistributional impact of nonlinear electricity pricing," *American Economic Journal: Economic Policy*, 4 (3), 56–90.

da Costa, Carlos E and Lucas J Maestri (2019) "Optimal Mirrleesian taxation in non-competitive labor markets," *Economic Theory*, 68 (4), 845–886.

Diamond, Peter A (1998) "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," *American Economic Review*, 83–95.

Doligalski, Paweł, G Piotr Dworczak, Mohammad Akbarpour, and Scott Kominers (2025) "Optimal redistribution via income taxation and market design,"Technical report, GRAPE Group for Research in Applied Economics.

Dworczak, Piotr, Scott Duke Kominers, and Mohammad Akbarpour (2021) "Redistribution through markets," *Econometrica*, 89 (4), 1665–1698.

Dworczak, Piotr and Ellen Muir (2024) "A mechanism-design approach to property rights," *Available at SSRN 4637366*.

Freixas, Xavier, Roger Guesnerie, and Jean Tirole (1985) "Planning under incomplete information and the ratchet effect," *The review of economic studies*, 52 (2), 173–191.

Hart, Oliver, Andrei Shleifer, and Robert W Vishny (1997) "The Proper Scope of Government: Theory and an Application to Prisons," *The Quarterly Journal of Economics*.

Hellwig, Martin F (2008) "A maximum principle for control problems with monotonicity constraints," *MPI Collective Goods Preprint* (2008/4).

Holmström, Bengt Robert (1978) *On Incentives and Control in Organizations.*: Stanford University.

Kang, Zi Yang and Mitchell Watt (2024) "Optimal In-Kind Redistribution," *arXiv preprint arXiv:2409.06112*.

Krishna, Kala (1990) "Protection and the product line: monopoly and product quality," *International Economic Review*, 87–102.

Lewis, Tracy R and David EM Sappington (1988) "Regulating a monopolist with unknown demand," *The American Economic Review*, 986–998.

Martimort, David, Jérôme Pouyet, and Carine Staropoli (2020) "Use and abuse of regulated prices in electricity markets:"How to regulate regulated prices?"," *Journal of Economics & Management Strategy*, 29 (3), 605–634.

Maskin, Eric and John Riley (1984) "Monopoly with incomplete information," *The RAND Journal of Economics*, 15 (2), 171–196.

Milgrom, Paul and Ilya Segal (2002) "Envelope theorems for arbitrary choice sets," *Econometrica*, 70 (2), 583–601.

Milgrom, Paul and Chris Shannon (1994) "Monotone comparative statics," *Econometrica: Journal of the Econometric Society*, 157–180.

Mirrlees, James A (1971) "An exploration in the theory of optimum income taxation," *The review of economic studies*, 38 (2), 175–208.

Mussa, Michael and Sherwin Rosen (1978) "Monopoly and product quality," *Journal of Economic theory*, 18 (2), 301–317.

Olmstead, Sheila M, W Michael Hanemann, and Robert N Stavins (2007) "Water demand under alternative price structures," *Journal of Environmental economics and management*, 54 (2), 181–198.

Pai, Mallesh and Philipp Strack (2022) "Taxing externalities without hurting the poor."

Pavan, Alessandro and Giacomo Calzolari (2009) "Sequential contracting with multiple principals," *Journal of Economic Theory*, 144 (2), 503–531.

Rochet, Jean-Charles and Lars A Stole (2002) "Nonlinear pricing with random participation," *The Review of Economic Studies*, 69 (1), 277–311.

Saez, Emmanuel (2001) "Using elasticities to derive optimal income tax rates," *The review of economic studies*, 68 (1), 205–229.

Schlom, Christoph (2024) "Essays on Constrained Mechanism Design and Regulation."

Stantcheva, Stefanie (2014) "Optimal income taxation with adverse selection in the labour market," *Review of Economic Studies*, 81 (3), 1296–1329.

Tobin, James (1970) "On limiting the domain of inequality," *The Journal of Law and Economics*, 13 (2), 263–277.

Topkis, Donald M (1998) *Supermodularity and complementarity*: Princeton university press.

Weitzman, Martin L (1977) "Is the price system or rationing more effective in getting a commodity to those who need it most?" *The Bell Journal of Economics*, 517–524.