When Improving the Public Option Backfires

Vasundhara Mallick*

November 20, 2026

Click here for the latest version

Abstract

In this paper, I analyze a model where a profit-maximizing monopolist competes against a free, capacity-constrained public option. My main finding is that when increasing capacity is not possible, improving quality can backfire, making every consumer worse off. The reason is that improving the quality of a public option holding its capacity fixed can have an anti-competitive effect. Rather than competing with a high quality and free public option, the monopolist finds it optimal to exploit the residual demand, namely the consumers who are rationed out of the public option. I show that this anti-competitive effect can be reversed by increasing capacity. Finally, I study how the public option changes the monopolist's menu, and the composition of her consumer base.

*Department of Economics, European University Institute. Email: vasundhara.mallick@gmail.com

1 Introduction

In markets for housing, education, and healthcare, the public sector exists alongside the private sector, offering heavily subsidized, or even free versions of these goods. The private sector offers these goods at higher prices, although it might present a wider range of options. The role of the public option is to provide a safety net for consumers. It does so by making the good accessible to those who would otherwise be priced out of the private market. It also incentivizes private firms to provide better terms to consumers who buy from them, and would pay exorbitant prices in absence of the competitive pressure exerted by the public sector.

There are two aspects of the public option that determine its effectiveness as a safety net: its capacity and quality. For example, the capacity of a public hospital is determined by the number of beds, whereas its quality is determined by its equipment, the qualifications of its doctors, etc. Thus, an improvement in the public option would involve increasing its capacity, improving its quality, or both. In the context of a hospital, this could mean using the budget to increase the number of beds and/or buying a new MRI machine.

In many situations, however, it might not be easy to increase the capacity of the public option, maybe due to the large fixed costs involved in doing so.[1] In such cases, the public option may be severely capacity-constrained and may face demand that far exceeds its capacity. For example, public healthcare systems like the NHS in the United Kingdom are overwhelmed and have long waiting times for patients. Free government schools in India often turn away students, asking them to try next year.[2] In such cases, consumers who demand the public option but fail to get it may be forced to purchase an overpriced and possibly low-quality private alternative as a last resort.

In this paper, I ask: how does the private sector respond to the limited capacity of the public option? If consumers who fail to get the public option can buy privately, private firms know that consumers who prefer the public option are not lost: they may come back to buy from them in case of rationing. So, they do not *have* to offer something better than the public option in order to sell to consumers. I study the strategic interactions emerge when a profit maximizing private firm competes against a free, capacity-constrained public option. In particular, what implications does the strategic response of the private sector have for the

[1] There could be situations where an increase in capacity involves high fixed costs, like construction of a new school or hospital building, whereas an increase in quality is easier to achieve, for e.g. by hiring better teachers or doctors, improving existing facilities, etc.

[2] These examples describe situations with waiting times, rater than a denial of service, but in many cases, a long waiting time is the same as a denial. A gravely ill person who cannot access treatment immediately through public healthcare will find it optimal to pay for private healthcare. Similarly, a student may not want to waste a school year.

role of the public option as a safety net, and what kind of public options make for better safety nets? When increasing capacity is not possible, does improving the quality of the public option always benefit consumers?

My first finding is that increasing quality can backfire, making *every* consumer worse off. The channel is through the response of the private sector. Suppose that the private sector comprises a profit-maximizing monopolist. When a low quality public option is introduced, the monopolist might lower her prices a bit to retain consumers. However, if the quality of the public option improves, lowering prices further to compete with the public option might not be optimal for the monopolist. Instead, she might revert to her original high prices and sell to *residual demand*: the consumers who tried to get the public option but failed. In this case, an increase in public quality has two adverse effects: i) it reduces the probability of getting the public option, as more consumers compete for it now, and ii) the consumers who fail to get the public option have to pay higher prices. The combination of these two effects can make every consumer worse-off.

I show this as a general result. For any capacity, I provide a sufficient condition on quality increments that make every consumer strictly worse off, and show that this effect is particularly pronounced at lower capacities. To this end, I consider a model where there is a unit mass of consumers, each with unit demand for good that can be produced at different qualities. Consumers are heterogeneous in their valuation for quality, and each consumer's valuation is her private information. The public and the private sector both provide the good. The public sector supplies the good for **free** at a fixed quality g, and has capacity $k < 1$. If demand exceeds capacity, allocation is random.[3] The private sector consists of a monopolist, who takes the public option (k, g) as given, and commits to a menu of quality-price pairs, where higher quality goods are more expensive. It is also costly for the monopolist to produce the good, and higher quality goods cost more to produce.

The important thing here is that a consumer cannot be forced to choose between the public option and the monopolist ex-ante: she can buy from the monopolist if she fails to get the public option. Every consumer has a choice: she can either buy directly from the monopolist or try to get the public option first. The choice each consumer makes depends both on the options in the monopolist's menu and the public quality g.

To see the underlying economics, consider the following simple example. There are two types of consumers: one with high valuation for quality, $\theta_h = 3$, and one with low valuation $\theta_l = 1$. Both types have equal mass. A consumer of type θ gets utility $\theta q - t$ if she buys a good of quality q and pays t. If there is no public option, the monopolist would offer only a

[3] More precisely, if mass d of consumers demands the public good, and $d > k$, then each consumer who demands the public good gets it with probability $\frac{k}{d}$.

high quality $q^* = 3$, at price $t^* = 9$, which means only high types purchase, and receive zero surplus. Now, suppose that we introduce a public option of quality $g_1 = 0.7$, and capacity 0.5: just enough to accommodate all low value consumers. The choice for the monopolist is between lowering her price to retain the high types, or continuing to charge a high price, in which case a high type would buy with a lower probability, only after failing to get the public option.[4]

I argue that improving the quality of the public option from $g_1 = 0.7$ to $g_2 = 0.8$ makes every consumer strictly worse-off. First, consider $g_1 = 0.7$. At g_1, the monopolist continues to offer quality $q^* = 3$, but finds it optimal to lower her price from 9 to $9 - 3g_1 = 6.9$, to retain high types. At price 6.9, the high types are indifferent between the public option and the monopolist, and therefore buy from the monopolist. The low types demand the public option and get it with probability one. Each type θ gets utility θg_1 here: high types get $U_h = 2.1$, and low types get $U_l = 0.7$. Therefore, at g_1 the public option performs its role as a safety net well: it provides access for low types and ensures better prices for high types.

Now, consider an increase in public quality to $g_2 = 0.8$. To retain θ_h, the monopolist would have to lower her price to $9 - 3g_2 = 6.6$. She does not find it optimal to do so and instead offers quality $q^* = 3$ at the original high price $t^* = 9$.[5] Clearly, now θ_h prefers the public option. Since *all* consumers demand the public option, each consumer succeeds only with probability 0.5. Furthermore, upon failing to get the public option, both types of consumers get zero utility from the private menu. Therefore, the expected utility of each type θ is $0.5\,\theta g_2$: θ_h gets $U_h = 1.2$, and θ_l gets $U_l = 0.4$. Observe that both types are strictly worse-off relative to g_1. The low types are worse-off because they now have to compete for the public good with the high types. The high types are worse-off because the monopolist is no longer competing for their business, so if they fail to get the public good, they get zero.[6]

So, the increase in quality, rather than exerting competitive pressure on the monopolist to further reduce her price, has an *anti-competitive effect*, where she finds it optimal to stop competing and only cater to *residual demand*. When capacity is lower, this prospect is even more tempting, because consumers fail to get the public option with a higher probability.

[4]Of course, price is not the only instrument for the monopolist, she can change the quality she offers too, in order to compete with the public option. However that would not be optimal here since the monopolist would never want to distort the high type's efficient quality. In general, the monopolist might use both instruments: price and quality, to respond to the public option. I discuss this in detail.

[5]To see why, the high types get $\theta_h g_2$ from the public option. To attract them directly, the monopolist must offer $q^* = 3$ at price $t_2 = 9 - 3g_2 = 6.6$, so that $\theta_h q^* - t_2 = \theta_h g$. The monopolist's profit in this case is $0.5(t_2 - \frac{(q^*)^2}{2}) = 0.55$. Instead, if she offers $q^* = 3$ at price $t^* = 9$, her profit is $(1-k)0.5(t^* - \frac{(q^*)^2}{2}) = 1.125$. This is obtained by substituting $k = 0.5$.

[6]It is worth pointing out that in this example, if capacity was one, **no** type would be worse off with an increase in quality. Each type would be guaranteed a surplus of θg_2, even if the monopolist does not try to attract them, because they can always just get the free public good.

My general model has a continuum of types. I set up the monopolist's problem and characterize its solution. I find that for any public option (k, g), every optimal mechanism is characterized by a cutoff c: consumer types below this cutoff try for the public option, and those above it buy directly from the monopolist. I show that keeping capacity fixed, as we increase the quality g, the cutoff rises: the monopolist competes for fewer consumer types directly, and relies more on residual demand. In particular, there is always a threshold of quality beyond which the monopolist capitulates completely: no consumer buys from her directly, and she only caters to residual demand. I also show that we can construct an interval around this threshold such that quality increments in that interval make all types strictly worse-off.

I also highlight how the monopolist *redesigns* her menu around the public option. I show that when there is an active residual market, the monopolist may introduce new products to cater specifically to residual demand. I argue that this is a consequence of the limited capacity of the public option. In the benchmark where the public option has capacity one, the introduction of a free public option simply crowds out some lower quality products that the monopolist was offering.[7] The monopolist continues to offer the same higher quality products as before, but *discounts* them, to compete with the public option. She does not introduce any new products. However, when the public option has limited capacity, the presence of residual demand changes the trade-off between earning profits from residual types and the information rents created for the higher types who buy directly. Therefore, the introduction of a capacity-constrained public option changes *what* products the monopolist offers, and not just the prices at which she offers them. I provide a sufficient condition for expansion of the monopolist's market coverage in presence of the public option, where she starts selling to the consumers she excluded when there was no public option.

Looking at the world through the lens of my model helps us understand why so many low quality private options exist alongside better quality, and subsidized public options. Examples include small private hospitals in India, with unmet safety standards and unqualified nurses (see Bhate-Deosthali, Khatri, & Wagle (2011) and Srivastav, Pant, Priya, & Coffey (2023)), and low quality private housing rented by people in the UK due to lack of social housing (see House of Commons Housing, Communities and Local Government Committee (2026)). These alternatives appear *dominated*: every consumer would either prefer a higher quality private alternative or the public option. Then why do they exist? My results suggest that such products, which are no one's first choice, may emerge to cater specifically to residual demand. They do not attempt to improve their quality because they are not competing with the public option.

[7]This benchmark has been studied in the literature; for my model, it is a special case of Jullien (2000).

Finally, I briefly talk about what happens if we could increase capacity. I explained the intuition behind an increase in quality having an anti-competitive effect and hurting consumers. I show that an increase in capacity can *reverse* this effect. In particular, if the monopolist has stopped competing for all consumers and only caters to residual demand, then an increase in capacity can make it optimal for her to offer a menu that attracts some types directly. This is because at a larger capacity, consumers who come prefer the public option come back to buy from the monopolist with a smaller probability, which makes depending on residual demand unprofitable.

The rest of the paper is organized as follows: section 1.1 reviews the related literature, section 2 sets up the model, section 3 sets up the monopolist's problem and presents important aspects of its solution, and finally, section 4 presents the main results.

1.1 Related Literature

This paper relates to a large literature on private provision of public goods with the aim of redistribution of wealth. Besley & Coate (1991) find that an inferior quality of the public good would make sure that only the poor choose it, thereby making the rich pay, through taxes, for a product they do not use, achieving the desired redistribution. Gahvari & Mattos (2007) show that combining in kind transfers with cash can help get around the inefficient low quality in Besley & Coate (1991). Nichols & Zeckhauser (1982) show that when need is not observable, and taxation is the only tool, then ordeals can help with targeting. Coate, Johnson, & Zeckhauser (1994) show that expanding the supply of the public option can lower the prices of the good, increasing welfare. These papers do study the effects of changing the quality of the public option, but the difference from my paper is that they feature a perfectly competitive private market. So, lowering quality to improve targeting is motivated by taxation and redistribution concerns, rather than capacity constraints the strategic response of the private sector.

Some recent papers on redistributive mechanism design are Akbarpour, Dworczak, & Kominers (2024), and Dworczak, Kominers, & Akbarpour (2021), however, they do not feature the strategic response of the private sector, or capacity constraints. In Kang (2023), a social planner chooses public allocation and quality jointly, and residual consumers enter a private market. The prices in the private market are affected by the design of the public option, but there are no capacity constraints and the private market is perfectly competitive.

Closer to my work is the literature on strategic interactions between the public and the private sector, where the presence of the public option causes the private sector to change what it offers. Examples include Klumpp & Su (2019), Ranasinghe & Su (2023) and Laine & Ma

(2017). However, the focus in these papers is not on limited capacity. For example, Klumpp & Su (2019) consider a trade-off between improving quality, and increasing affordability of the public option. I show that even in absence of any such trade-off, or concern about how the public good is financed, an increase in quality can make every consumer worse-off.

There is a literature that combines rationing and private sector response. The most closely related is Mekonnen (2026), who also studies a model with a monopolist competing with a capacity-constrained private option. There are two important differences from my model: i) those who choose the public option cannot buy from the private sector if they are rationed out, and ii) he focuses on the effects of increasing capacity rather than quality. He shows that an increase in capacity can never hurt low-value consumers. In contrast, I show that the possibility of selling to residual demand can lead the monopolist to respond to an increase in quality in a way that makes every consumer worse off. Also related are Grassi & Ma (2011), and Grassi & Ma (2012), who focus on the design of the optimal rationing rule, anticipating private sector responses. However, these models do not feature quality: responses are through changes in prices.

Finally, my paper is related to the monopolistic screening literature, starting from Mussa & Rosen (1978), and monopolistic screening with type dependent participation constraints, as in Jullien (2000) and Martimort & Stole (2022). In these papers, however, the monopolist cannot determine the probability with which consumers can access their outside options. Furthermore, in my paper, the monopolist may herself be an outside option or a last resort for the consumers who fail to get the public option. Which consumers treat the monopolist as the outside option, and which consumers treat the public option as the outside option, is endogenously determined by the monopolist's choice of menu.

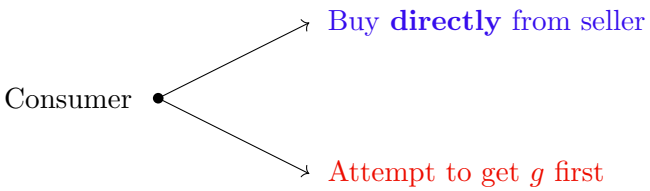
2 Model

Consider the market for an indivisible good that can be produced at different quality levels $q \in \mathbb{R}_+$. There is a unit mass of consumers, each with unit demand for the good. Each consumer has a type $\theta \in \Theta := [\underline{\theta}, \overline{\theta}]$, where $0 \leqslant \underline{\theta} < \overline{\theta} < \infty$, where θ represents her valuation for the good, and is her private information. Consumers are risk neutral and have quasilinear preferences. So, if a consumer of type θ buys a good of quality q at price t, her payoff is $\theta q - t$. The distribution of types in the market is denoted by F, where F has density f that is bounded above and bounded away from zero, so there exist $\underline{f}, \overline{f}$ such that $0 < \underline{f} \leqslant f(\theta) \leqslant \overline{f} < \infty$ for all $\theta \in \Theta$.

The good is provided by both the public and the private sector. The private sector consists of a profit maximizing monopolist who can commit to a menu of contracts $\mathcal{M}$. Each

contract in $\mathcal{M}$ is a tuple $(q,t) \in \mathbb{R}_+ \times \mathbb{R}$: if a consumer chooses contract (q,t), she gets a good of quality q at price t. I assume that the menu always contains the null contract $(0,0)$. The monopolist's cost of producing one unit of the good at quality q is $\frac{q^2}{2}$. The public sector supplies the good for **free** at a fixed quality g. It is capacity-constrained, with capacity $k \in (0,1)$. Since the market consists of a unit mass of consumers, a capacity $k < 1$ implies that if all consumers demand the public option, not all of them can get it. If mass d of consumers demand the public option, then each consumer who demands it is allocated the public option with probability $\min\{1, \frac{k}{d}\}$. I model the public sector as passive here: the capacity k and quality g are the parameters of the model, rather than chosen optimally to maximize an objective function.[8]

I assume that if a consumer attempts to get the public option, but fails due to rationing, she **can** buy from the private seller. So, every consumer has two options: she can buy directly from the seller, without attempting to get the public option, or she can attempt to get the public option first, and only buy from the seller if she fails.



Importantly, I assume that the monopolist cannot differentiate between the consumers who have come to buy from her directly, and those who are buying only after failing to get the public option. So, she cannot condition the set of contracts she offers a consumer on whether she is a direct or indirect consumer, there is just one menu $\mathcal{M}$ all consumers can choose from.

Given the seller's menu $\mathcal{M}$ and g, let $U_{\mathcal{M}}(\theta) := \max_{(q,t)\in\mathcal{M}}\{\theta q - t\}$, the payoff of type θ from $\mathcal{M}$. Since θg is the payoff that type θ gets if she is allocated the public option, therefore, if $U_{\mathcal{M}}(\theta) > \theta g$, she prefers to buy directly from the seller, and if $U_{\mathcal{M}}(\theta) < \theta g$, tries to get g first. If $U_{\mathcal{M}}(\theta) = \theta g$, the consumer is indifferent, and we assume that in this case, she follows the seller's recommendation on what to do. To sum up, the **timeline** is as follows:

- The monopolist commits to a menu $\mathcal{M}$.
- Each type θ decides whether to demand the public option or to buy directly from the seller.

[8] Later, I will conduct comparative statics with respect to them, and comment on what the optimal (k,g) might be, if the government wants to maximize some objective, for e.g. weighted consumer surplus.

- If a consumer demands the public option but fails to get it, she goes to the seller and buys her favorite option from $\mathcal{M}$.[9]

In the next section, I set up and solve the monopolist's problem.

3 Monopolist's Problem

I now set up the monopolist's problem, which is: which menu $\mathcal{M}$ should she offer to maximize her expected profits, when her menu determines which types buy directly and which types only buy after they fail to get the public option. By the Revelation Principle, we can restrict attention to direct and incentive compatible (IC) mechanisms, and individually rational (IR) mechanisms.[10] A direct mechanism is a pair of functions (q, t), where $q : \Theta \to \mathbb{R}_+$, and $t : \Theta \to \mathbb{R}$. For any type report θ, a consumer gets a good of quality $q(\theta)$ and pays $t(\theta)$ for it. Incentive compatibility means that *conditional* on buying from the monopolist, each consumer finds it optimal to report her type truthfully. Let $U(\theta) := \theta q(\theta) - t(\theta)$, the truth-telling payoff of type θ in mechanism (q, t). Then, (q, t) is IC if and only if $U(\theta) = \max_{\{\theta' \in \Theta\}} \{\theta q(\theta') - t(\theta')\}$ for every θ. Individual rationality means that conditional on buying from the monopolist, every type receives a non-negative payoff, i.e. $U(\theta) \geqslant 0$ for all θ. The following is a standard lemma from mechanism design, that characterizes IC and IR mechanisms.

Lemma 1. *A direct mechanism (q, t) is IC and IR if and only if i) q is non decreasing, and ii) $U(\theta) = \underline{u} + \int_{\theta_l}^{\theta} q(x)dx$ for all $\theta \in \Theta$, and $\underline{u} \geqslant 0$, where $U(\theta) = \theta q(\theta) - t(\theta)$, where $\underline{u} := q(\underline{\theta})\underline{\theta} - t(\underline{\theta})$.*

The characterization is independent of what g or k are, and which types buy directly and which are indirect, because the seller cannot distinguish the types who buy from her directly from the ones who don't. I now set up the monopolist's optimization problem. Fix a direct, IC, IR mechanism (q, t). Let $\pi(\theta)$ represent the seller's profit from selling to type θ, so $\pi(\theta) = t(\theta) - \frac{q(\theta)^2}{2}$, where $t(\theta) = \theta q(\theta) - U(\theta)$, and $U(\theta)$ is the utility of θ in this mechanism. For any type θ, there are three possibilities: i) $U(\theta) > \theta g$, so she strictly prefers to buy directly from the monopolist, ii) $U(\theta) < \theta g$, i.e. she strictly prefers the public option, ad ii) $U(\theta) = \theta g$, i.e. she is indifferent and follows the monopolist's recommendation. Let $D'_{(q,t)}$ be a measurable subset of $\{\theta | U(\theta) = \theta g\}$ chosen by the monopolist, such that the monopolist recommends all

[9] This could be $(0, 0)$, which represents that θ does not buy from the seller after being rationed out by the public option, because no option in $\mathcal{M}$ gives her a positive utility.

[10] For a proof of the Revelation Principle in my setting, as well as a discussion about why the characterization of IC and IR mechanism is the same as it would be in the pure monopoly case, without a public option, see the Appendix.

types in $D'_{(q,t)}$ to try for the public option, and all types in $\{\theta|U(\theta) = \theta g\} \cup (D'_{(q,t)})^c$ to buy directly. The total public demand therefore is $D_{(q,t)} := \{\theta|U(\theta) < \theta g\} \cup D'_{(q,t)}$. Let $\beta(D_{(q,t)})$ denote the probability with which types who demand the public option fail to get it, and return to buy from the seller. Then, $\beta(D_{(q,t)}) = \max\{0, \frac{k}{F\{\theta|\theta \in D_{(q,t)}\}}\}$.

The monopolist's profit from (q,t) , and the choice $D'_{(q,t)} \subseteq \{\theta|U(\theta) = \theta g\}$ is:

$$\beta(D_{(q,t)}) \int_{D_{(q,t)}} \pi(\theta) dF(\theta) + \int_{(D_{(q,t)})^c} \pi(\theta) dF(\theta) \tag{1}$$

The monopolist's problem is to maximize 1 over all IC and IR direct mechanisms (q,t), and all $D'_{(q,t)} \subseteq \{\theta|U(\theta) = \theta g\}$. I now show that **every optimal mechanism has a cutoff structure**. There is a cutoff $c \in [\underline{\theta}, \overline{\theta}]$ such that all types $\theta < c$, if any, strictly prefer to try for the public option, and all types $\theta > c$, if any, strictly prefer to buy directly from the monopolist. The cutoff type, if interior, is indifferent. The extreme cases are represented by cutoffs $\underline{\theta}$, which means that all types buy directly, and $\overline{\theta}$, which means that no type buys directly. For any interior cutoff $c \in (\underline{\theta}, \overline{\theta})$, define $\beta(c) := \max\{0, 1 - \frac{k}{F(c)}\}$. If $\theta < c$ demand the public option, then $\beta(c)$ is the probability with which consumers who demand the public option fail to get it, and buy from the monopolist. The following Lemma states the cutoff result formally:

Proposition 1. *Suppose (q^*, t^*) is an optimal mechanism. Let $U^* := \theta q^*(\theta) - t^*(\theta)$ denote the utility of type θ, $\pi^*(\theta) := t^*(\theta) - \frac{q^*(\theta)^2}{2}$ denote the monopolist's profit from selling to θ, and Π^* denote her expected profit from (q^*, t^*). Then, the following possibilities are mutually exclusive and exhaustive, and, in each case, the set $\{\theta|U^*(\theta) = \theta g\}$ consists of at most two points, and is therefore a null set.*

- *All types buy directly from monopolist: $U^*(\theta) > \theta g$ for every $\theta > 0$, and $\Pi^* = \int_{\underline{\theta}}^{\overline{\theta}} \pi^*(\theta) dF(\theta)$.*
- *Interior cutoff: There exists a $c \in (\underline{\theta}, \overline{\theta})$ such that $U^*(\theta) < \theta g$ for all $\theta < c$, and $U^*(\theta) > \theta g$ for all $\theta > c$, and $U^*(c) = cg$. All $\theta > c$ buy directly from the monopolist, and all $\theta < c$ demand the public option, and buy from the monopolist only if they fail to get it. The profit in this case is $\Pi^* = \beta(c) \int_{\underline{\theta}}^{c} \pi^*(\theta) dF(\theta) + \int_{c}^{\overline{\theta}} \pi^*(\theta) dF(\theta)$.*
- *No type buys directly: $U^*(\theta) < \theta g$ for all $\theta \in (\underline{\theta}, \overline{\theta})$. All types buy from the monopolist only if rationed out, and $\Pi^* = (1-k) \int_{\underline{\theta}}^{\overline{\theta}} \pi^*(\theta) dF(\theta)$.*

The proof of Lemma 1 is in the Appendix. The idea is that higher types value public options more. To attract consumers directly, the monopolist offers them qualities strictly higher than the public quality g. This ensures that if a type θ weakly prefers the monopolist

to the public option, then every type $\theta' > \theta$ strictly prefers the monopolist, which proves the cutoff structure.

3.1 Optimal Mechanisms

I do not explicitly solve for the optimal mechanism for every distribution F, and public option (k, g). Rather, for every cutoff $c \in [\underline{\theta}, \overline{\theta}]$, I solve the *inner problem* corresponding to this cutoff: what is the optimal mechanism *if* it induces cutoff c. These characterizations are useful for proving my main results. Also, the form of the optimal mechanism in each case sheds light on the economics of the monopolist's problem.

Before proceeding to characterize what the optimal mechanism must be in different cases, it is useful to define two benchmarks. The first is the **pure monopoly benchmark**, where there is no public option. The second is the **capacity-one benchmark**, where the monopolist operates alongside a public option of capacity one. Let (q^M, t^M) denote the unique optimal mechanism in the pure monopoly benchmark, and let Π^M denote her expected profit from it. The mechanism (q^M, t^M) solves the following problem:

$$\begin{aligned} \max_{(q,t)} \quad & \int_{\underline{\theta}}^{\overline{\theta}} (t(\theta) - \frac{q(\theta)^2}{2}) dF(\theta) \\ \text{subject to} \quad & (q, t) \text{ IC and IR} \end{aligned} \tag{P-M}$$

Lemma 1 provides the characterization of IC and IR mechanisms, which follows from standard arguments. The solution to P-M is due to Mussa & Rosen (1978). I provide a full description in the Appendix for completeness. For now, let $a^M := \inf\{\theta | q^M(\theta) > 0\}$. Thus, since q^M is non-decreasing, therefore in the pure monopoly benchmark, $\theta > a^M$ buy from the monopolist, and $\theta < a$ do not buy, and get utility zero.

In my environment, the capacity-one benchmark here is a special case of the problem solved by Jullien (2000), and I describe the optimal mechanism in the Appendix. In this case, there is no *residual* demand: consumers who strictly prefer the public option never buy from the monopolist. I point out one important feature of the optimal mechanism without describing it in detail.

Fact 1. *Let (q^1, t^1) be an optimal mechanism in the capacity-one benchmark. Then, there exists $c^1 > a^M$ such that $q^1(\theta) > 0$ if and only if $\theta \geqslant c^1$. Furthermore, $q^1(\theta) = q^M(\theta)$, and $t^1(\theta) = t^M(\theta) - c^1 g + U^M(c^1)$ for every $\theta \geqslant c^1$, where $U^M(c^1)$ is the utility of types c^1 from $)q^M, t^M)$, the pure monopoly mechanism.*

The above fact says that relative to pure monopoly, the set of types that the monopolist sells to shrinks, which is intuitive. The interesting thing is that in any optimal mechanism

with capacity one,[11] the monopolist simply offers the optimal monopoly quality q^M at a possibly discounted price to the consumers she sells to. The discounting is such that the lowest type c^1 is indifferent between the public and the private option. This ensures that all $\theta > c^1$ strictly prefer the monopolist.

Broadly speaking, when faced with a public option, the monopolist can *compete* to retain consumers in two ways. She can provide better quality products relative to the monopoly case, or offer the same products at a discount. When capacity is one, she only uses discounting. The intuition is that in the capacity-one case, every consumer either buys with probability one or does not buy. Thus, for the types that do buy, the rent extraction vs efficiency trade-off is the same as in the pure monopoly case. Thus, the monopoly qualities remain optimal.

I will argue that in my model, where the public option has limited capacity, as long as the demand for the public option is lower than the capacity, the monopolist offers discounted versions of the monopoly products. This is because limited capacity has no bite here: there is no rationing in equilibrium. This changes when the optimal mechanism induces public demand that exceeds capacity. I will show that the presence of residual demand, leads to a non-trivial redesign of the monopolist's mechanism, relative to the pure monopoly case. Therefore, the public option may change *which* products are available on the private market, and not merely their prices.

Now, I go through the different possible cases that can correspond to an optimal mechanism, and describe the form of the optimal mechanism in each case.

3.2 Optimal Mechanism with Public Demand either Zero or One

I first deal with the two extreme cases, where either **all** types buy directly from the monopolist, or **no** type buys directly, and everyone demands the public option. If the optimal mechanism corresponds to either of these two cases, then it must essentially be the monopolist's optimal mechanism from the **pure monopoly benchmark**, where there is no public option.

The following lemma describes what the optimal mechanism must be if it induces public demand of either zero or one.[12]

Proposition 2. *Let (q^*, t^*) be an optimal mechanism, and let Π^* be the monopolist's expected profit from it. Then,*

- *If (q^*, t^*) induces **all** types to buy directly from the monopolist, then, $(q^*, t^*) = (q^M, t^M - \underline{\theta} g)$, and $\Pi^* = \Pi^M - \underline{\theta} g$*

[11] There can be multiple optimal mechanisms, see Appendix.

[12] Since there is a unit mass of consumers, therefore public demand one means that all consumers demand the public option, and no one buys directly from the monopolist.

- *If* (q^*, t^*) *is such that* ***no*** *type buys directly, then* $(q^*, t^*) = (q^M, t^M)$*, and* $\Pi^* = (1-k)\Pi^M$.

The result in Proposition 2 is intuitive, because in both cases, the conditional distribution of types that buy from the monopolist is the same as the pure monopoly case. This is obvious when all types buy directly from the monopolist. When all types prefer the public option, each consumer only gets it with probability k, and returns to buy from the monopolist with probability $(1-k)$, leaving the distribution of types unchanged. The only difference is that to incentivize all types to buy directly, the monopolist offers discounted versions of the pure monopoly contracts. Every transfer is reduced by $\underline{\theta} g$, which makes the lowest type indifferent between the public and the private option.

3.3 Optimal Mechanism with Public Demand lower than Capacity

In this section, I deal with the case where the optimal mechanism induces public demand lower than capacity. Therefore, when the optimal mechanism is offered, either i) all types buy directly from the monopolist, so no consumer demands the public option, or ii) the optimal mechanism induces an interior cutoff c, such that the public demand $F(c) < k$.

I argue that if an optimal mechanism (q^*, t^*) induces public demand that is lower than capacity, then (q^*, t^*) must *also* be optimal in the full-capacity benchmark.

Proposition 3. *Fix and optimal mechanism* (q^*, t^*) *and denote its profit by* Π^**. Suppose, the optimal mechanism induces public demand lower than capacity. Then,* (q^*, t^*) *is also optimal when capacity is one, and* $\Pi^* = \Pi^1$.

Proof. Consider the optimal mechanism (q^*, t^*) at k. Let $D := \{\theta | U^*(\theta) < \theta g\}$, the set of types that demand the public option, and $F(D) < k$, so every $\theta \in D$ gets the public option with probability one, and does not return to buy from the monopolist.[13] Observe that for any θ, the decision of whether to demand the public option only depends on $U^*(\theta) = \theta q^*(\theta) - t^*(\theta)$ vs θg, and is independent of capacity. So, (q^*, t^*) is offered at $k' = 1$, again, $\theta \in D$ demand the public option, and $\theta \in D^c$ buy directly from the monopolist. So, the monopolist's expected profit remains Π^*. Suppose now that (q^*, t^*) is **not** optimal at $k' = 1$, i.e. there exists a mechanism (q^{**}, t^{**}), such that its expected profit is Π^1 at $k' = 1$ and $\Pi_1 > \Pi^*$. But this cannot be, because offering (q^{**}, t^{**}) at capacity is $k < 1$, yields an expected profit of at least Π^1: every direct type remains direct, and some types that prefer the public option may come back to buy from the monopolist due to rationing. This contradicts the optimality of (q^*, t^*) at k. □

[13] By Lemma 1, the set of indifferent types $\{\theta | U^*(\theta) = \theta g\}$ must be a null set at an optimal mechanism, so almost every type that demands the public option must strictly prefer it.

Thus, the optimal mechanisms in the full capacity problem are the only candidates for optimal mechanisms in my model that induce public demand lower than capacity.

3.4 Optimal Mechanism with Interior Cutoff and Public Demand that Exceeds Capacity

I now come to the non-trivial case, which is when a positive mass of consumers buy directly from the monopolist, *and* the demand for the public option exceeds capacity. Thus, there is an active residual market because consumers who prefer the public option may come back to buy from the monopolist. The monopolist's optimal mechanism balances rent extraction from residual consumers against the information rents they generate for the higher, direct types. This is the **only** case where the monopolist may offer different quality products relative to the pure monopoly benchmark. I first set up the monopolist's problem and then provide intuition for why the optimal mechanism can feature quality distortions relative to monopoly.

By Proposition 1, if a positive mass of consumers buy directly and public demand exceeds capacity, then it must be that the optimal mechanism induces a cutoff $c \in (\underline{\theta}, \overline{\theta})$ with $F(c) > k$. Type $\theta > c$ buy directly, types $\theta < c$ demand the public option, so public demand is $F(c)$, and type c is indifferent between the public and the private option.

I characterize what the optimal mechanism must be *if* it induces an interior cutoff $c \in (\underline{\theta}, \overline{\theta})$ with $F(c) > k$. For any such cutoff c, I solve the *fixed-cutoff problem* corresponding to c. The solution to the fixed-cutoff problem at c is the profit-maximizing IC and IR mechanism (q, t) that *induces* c, i.e. $U(\theta) \geqslant \theta g$ for all $\theta \geqslant c$, and $U(\theta) \leqslant \theta g$ for all $\theta \leqslant c$. I set up the fixed-cutoff problem formally, and sketch the strategy for solving it: the details are relegated to the Appendix. Then, I highlight through an example how the presence of residual demand can create distortions in the qualities that the monopolist offers, relative to the pure monopoly benchmark.

I now set up the fixed-cutoff problem corresponding to $c \in (\underline{\theta}, \overline{\theta})$ such that $F(c) > k$. Define $\beta_c := 1 - \frac{k}{F(c)}$: this is the probability with which each $\theta < c$ fails to get the public option, and buys from the monopolist. Suppose a mechanism (q, t) induces c. Let $\pi(\theta) := t(\theta) - \frac{q(\theta)^2}{2}$, the monopolist's profit from θ under (q, t). Then, the monopolist's profit from (q, t) is:

$$\beta_c \int_{\underline{\theta}}^{c} \pi(\theta) dF(\theta) + \int_{c}^{\overline{\theta}} \pi(\theta) dF(\theta)$$

This is because $\theta < c$ only buy with probability β_c. To solve the fixed-cutoff problem for c, we need to maximize the above objective over the set of IC and IR mechanisms (q, t) subject to the constraint that (q, t) induces c. From Lemma 1, a mechanism (q, t) is IC and IR if and

only if i) q is non decreasing, and ii) $U(\underline{\theta}) \geqslant 0$. For (q,t) to also induce c, we need $U(\theta) \geqslant \theta g$ for all $\theta \geqslant c$, and $U(\theta) \leqslant \theta g$ for all $\theta < c$. I show that this continuum of constraints can be reduced to two constraints: the indifference of the cutoff type, i.e. $U(c) = cg$, and that the lowest type prefers the public option, i.e. $U(\underline{\theta}) \leqslant \underline{\theta} g$. The following lemma summarizes this:

Lemma 2. *Fix mechanism (q,t) such that q is non-decreasing, and $U(\underline{\theta}) \geqslant 0$. Then, $U(\theta) \geqslant \theta g$ for all $\theta \geqslant c$, and $U(\theta) \leqslant \theta g$ for all $\theta \leqslant c$ holds if and only if i) $U(c) = cg$, and ii) $\underline{u} \leqslant \underline{\theta} g$*

The proof of Lemma 2 is in the Appendix, and it reduces our fixed-cutoff problem to:

$$\begin{aligned} \max_{(q,U(\underline{\theta}))} \quad & \beta_c \int_{\underline{\theta}}^{c} \pi(\theta) dF(\theta) + \int_{c}^{\overline{\theta}} \pi(\theta) dF(\theta) \\ \text{subject to} \quad & q(.) \text{ non decreasing} \\ & U(\underline{\theta}) \in [0, \underline{\theta} g] \\ & U(\underline{\theta}) + \int_{\underline{\theta}}^{c} q(s) ds = cg \end{aligned} \tag{P-fixed cutoff}$$

From the envelope formula, we have $U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(s) ds$ for every θ, so $U(\underline{\theta}) + \int_{\underline{\theta}}^{c} q(s) ds = cg$ is simply the $U(c) = cg$ constraint. Define

$$\varphi_c(\theta) := \theta - \frac{F(c) - F(\theta)}{f(\theta)}, \ \theta < c$$

and

$$\varphi(\theta) := \theta - \frac{1 - F(\theta)}{f(\theta)}, \ \theta \geqslant c$$

We can now write out objective in terms of these objects.

Lemma 3. *For any (q,t) in the feasibility set of P-fixed cutoff, the objective function in P-fixed cutoff can be written as*

$$J_c(q, U(\underline{\theta})) := \beta_c \int_{\underline{\theta}}^{c} (\varphi_c(\theta) q(\theta) - \frac{q(\theta)^2}{2}) dF(\theta) + \int_{\underline{c}}^{\overline{\theta}} (\varphi(\theta) q(\theta) - \frac{q(\theta)^2}{2}) dF(\theta) - U(\underline{\theta}) \beta_c F(c) - cg(1 - F(c)) \tag{2}$$

The above virtual value representation of the objective function is obtained by substituting $t(\theta) = \theta q(\theta) - U(\theta)$ in the objective function in P-fixed cutoff, and changing the order of integration. The market is divided into upper and lower sub-markets, and in the upper

sub-market, c is the lowest type and its rent is pinned down by the constraint $U(c) = cg$, hence for each $\theta > c$, we substitute $U(\theta) = U(c) + \int_c^\theta q(x)dx$, hence the $-cg(1 - F(c))$ term. For $\theta < c$, we substitute $U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^\theta q(s)ds$. Observe also that the virtual values in the lower and upper sub-markets are different: $\varphi(.)$ is the standard monopoly virtual value, whereas $\varphi_c(.)$ is the virtual value in the lower sub market $[\underline{\theta}, c)$: the virtual value $\varphi_c(.)$ captures the marginal value, for the seller, of supplying quality to type θ in the *lower* sub-market, *as if* information rents from this do not leak to $\theta > c$. This is why $\varphi_c(\theta)$ contains the truncated term $F(c) - F(\theta)$.

There is a discontinuous drop in the virtual value around the cutoff. Taking one-sided limits at c, $\varphi(c^-) = c$, because $|\varphi(\theta) - \theta| = \frac{F(c)-F(\theta)}{f(\theta} \leqslant \frac{F(c)-F(\theta)}{\underline{f}} \to 0$ as $\theta \uparrow c$. The right limit of $\varphi(c^+) = c - \frac{1-F(c)}{f(c^+)}$, so the drop is given by $\frac{1-F(c)}{f(c^+)}$. The intuition is again that in the lower sub-market, the type c is at the top of the sub-market, and information rents from this type do not leak to any higher type, since the rents that $\theta > c$ get from c are pinned down by $U(c) = cg$, and are not affected by the quality supplied to $\theta < c$. Therefore, for $\theta < c$, as $\theta \uparrow c$, the seller's marginal willingness to supply quality is higher than it is for $\theta > c$ as $\theta \downarrow c$, because rents from these types propagate to all higher types up to $\overline{\theta}$.

Due to this drop, and the constraint $U(c) = cg$, even if F is regular to begin with, we might need ironing. I do not go into the details of the ironing procedure here, and fully solve P-fixed cutoff in the Appendix, but we briefly outline the plan here. Basically, $U(c) = cg$ is our problematic constraint, so we it and use the multiplier $\lambda \in \mathbb{R}$ to reward or penalize its violation. Define:

$$\mathcal{L}_c(q, U(\underline{\theta}), \lambda) := J_c(q, U(\underline{\theta})) + \lambda(U(\underline{\theta}) + \int_{\underline{\theta}}^c q(\theta)d\theta - cg) \tag{3}$$

I solve the problem of maximizing 23 over non-decreasing q, and $U(\underline{\theta}) \in [0, \underline{\theta}g]$. Then, we show that there exists a λ^* such that the solution $(q_{\lambda*}, \underline{U}(\lambda^*))$ corresponding to λ^* satisfies the

$$U(c) = cg$$

constraint. Finally, I use Lagrangian Sufficiency to conclude that $(q_{\lambda*}, \underline{U}(\lambda^*))$ is the solution to P-fixed cutoff.

To end this section, I briefly provide the intuition for why, when the optimal mechanism features public demand greater than capacity, and therefore an active residual market, we can see quality distortions away from the pure monopoly qualities. We already saw that residual demand creates two sub-markets, and the virtual value, and therefore the monopolist's marginal willingness to supply quality to the lower sub -market is higher than the monopoly case. This suggests that the presence of this market changes the trade-off, relative to the

monopoly case, between extracting rents from the residual types, and information rents that this creates for the types that buy directly.

If c is the threshold type, then one strategy is to merely offer the monopoly qualities at lower prices to $\theta > c$, to attract them directly. However, this reduces what the monopolist can extract from types $\theta < c$, who constitute residual demand. This is because a lower price increases the utility of every type by the same amount. If, on the other hand, the monopolist increases the qualities and transfers of $\theta > c$, such that they want to buy directly, then the lower types gain less from this improvement, since they have lower value for quality. Therefore, the monopolist can extract more from them in this case.

4 Main Results

I now describe our main results.

4.1 Direct Demand Is Ultimately Zero

In this section, I show that for g above a certain threshold, no type buys directly from the monopolist in the optimal mechanism. Therefore, public demand becomes one. From Proposition 2, we know that that if an optimal mechanism induces **no** type to buy directly, it must be (q^M, t^M). The next theorem tells us that for any k exactly when such a mechanism is optimal. I now state the theorem.

Theorem 1. *Fix capacity k. Let $g^* := \frac{U^M(\overline{\theta})}{\overline{\theta}}$, where $U^M(\overline{\theta})$ is the utility of type $\overline{\theta}$ from the pure monopoly mechanism (q^M, t^M). Then, there is a continuous, strictly increasing function*

$$\hat{g} : (0,1) \to (g^*, \frac{\overline{\theta}}{2})$$

such that for every $k \in (0,1)$, the following holds:

- *For every $g \geqslant \hat{g}(k)$, the mechanism (q^M, t^M) is optimal, and its outcome is that **no** type buys directly from the monopolist. For $g > \hat{g}(k)$, (q^M, t^M) is the unique optimal mechanism.*
- *For $g < \hat{g}(k)$, in **every** optimal mechanism, a positive mass of consumers must buy directly.*

Moreover, we have that $\lim_{k \downarrow 0} \hat{g}(k) = g^*$, and $\lim_{k \uparrow 1} \hat{g}(k) = \frac{\overline{\theta}}{2}$.

Proof Sketch: For any $k \in (0,1)$, let $P(k)$ denote the set of public qualities g at which a mechanism where all types demand the public option is optimal. I first argue that $P(k)$ is non-empty for any $k \in (0,1)$. Then, I show that the function $\hat{g} : (0,1) \to (g^*, \frac{\overline{\theta}}{2})$ defined as $\hat{g}(k) := \inf P(k)$ for any k, is our desired function from Theorem 1. T

Fix $k \in (0,1)$. I show that all public qualities greater than $\frac{\overline{\theta}}{2}$ are in $P(k)$. This is because at $g > \frac{\overline{\theta}}{2}$, the monopolist cannot make positive profits by directly attracting any type. To see why, suppose that, in the optimal mechanism (q,t), type θ buys directly, so $U(\theta) \geqslant \theta g$. The monopolist's profit from selling to θ must be non-negative, so $t(\theta) \geqslant \frac{q(\theta)^2}{2}$, which implies that

$$U(\theta) = \theta q(\theta) - t(\theta) \leqslant \theta q(\theta) - \frac{q(\theta)^2}{2} \leqslant \frac{\theta^2}{2}$$

where the first inequality uses $t(\theta) \geqslant \frac{q(\theta)^2}{2}$, and the second inequality is just by maximizing $\theta q(\theta) - \frac{q(\theta)^2}{2}$ over $q(\theta)$. Combining $U(\theta) \leqslant \frac{\theta^2}{2}$ with $U(\theta) \geqslant \theta g$, we get $2g \leqslant \theta$: this is the condition g must satisfy for the monopolist to sell directly to θ in an optimal mechanism. If $g > \frac{\overline{\theta}}{2}$, this condition is violated for every θ. So, $(\frac{\overline{\theta}}{2}, \infty) \subseteq P$. This argument also shows that $\hat{g}(k) \leqslant \frac{\overline{\theta}}{2}$. I prove the strict inequality in the Appendix. I now sketch the rest of the proof in steps:

Step 1: I show that $\hat{g}(k) \geqslant g^*$. By Proposition 2, if an optimal mechanism induces full public demand, it must be (q^M, t^M). A corollary of Proposition 2 is that full public demand can *only* be optimal at g if (q^M, t^M) induces it, i.e. if $U^M(\theta) \leqslant \theta g$ for all θ, where $U^M(\theta) = \theta q^M(\theta) - t^M(\theta)$. At $g < g^*$, there exists $a < \overline{\theta}$ such that $U^M(\theta) > \theta g$ for all $\theta \in [a, \overline{\theta}]$. This is because the function $\frac{U^M(\theta)}{\theta}$ is non-decreasing on $[\underline{\theta}, \overline{\theta}]$ (see Lemma 17 in the Appendix). Since $g^* = \frac{U^M(\overline{\theta})}{\overline{\theta}}$, therefore at any $g < g^*$, $U(\overline{\theta}) > \overline{\theta} g$. By continuity of U^M, this strict inequality extends to a non empty interval below $\overline{\theta}$. Therefore, at any $g < g^*$, (q^M, t^M) does not induce full public demand. This implies that the outcome of an optimal mechanism cannot be that every type demands the public option.

At g^*, we cannot have an optimal mechanism with full public demand. The details are in the Appendix, but the intuition is that at g^*, if the monopolist offers (q^M, t^M), $\overline{\theta}$ is already indifferent. By scaling up the mechanism (q^M, t^M) a little, i.e. by offering slightly better qualities, the monopolist can attract positive mass of types as direct buyers, increasing her profits.

Step 2: I now show that the optimality of full public demand is absorbing. More precisely, if, at g, a mechanism where all types demand the public option is optimal, then for every $g' > g$, any optimal mechanism must be such that all types demand the public option. From ***Step*** 1, we know that if a mechanism with full public demand is optimal at g, then $g > g^*$.

Furthermore, at any $g' > g^*$, if (q^M, t^M) is offered, all types strictly prefer the public option, and the monopolist's profit is $(1-k)\Pi^M$. This is one again because $\frac{U^M(\theta)}{\theta}$ is non-decreasing, and by definition of g^*.

Even though the profit from offering (q^M, t^M) does not change with g, the profit from any other mechanism, that induces a positive mass of types to buy directly, falls strictly with g. Therefore, if (q^M, t^M) is optimal at $g > g^*$, i.e. mechanism induces a positive mass of direct demand does not beat it strictly at g, then such a mechanism cannot beat it at $g' > g$ either. To see the intuition, suppose not, i.e. (q^M, t^M) is optimal at $g > g^*$, but at $g' > g$, a mechanism that attracts all types directly is strictly better than (q^M, t^M). By Proposition 2, this mechanism must be $(q^M, t^M - \underline{\theta} g')$, and its profit, $\Pi^M - \underline{\theta} g'$ is strictly higher than $(1-k)\Pi^M$, the profit from (q^M, t^M). But then, at $g < g'$, the monopolist can offer mechanism $(q^M, t^M - \underline{\theta} g)$, attract every type directly, and make profit $\Pi^M - \underline{\theta} g' + \underline{\theta}(g' - g)$, contradicting the optimality of (q^M, t^M) at g. The cases with interior cutoffs have similar arguments, and are in the Appendix.

Step 3: It is intuitive that $\hat{g}(k)$ is strictly increasing in k. The monopolist's optimal mechanism if all types demand the public option is (q^M, t^M), irrespective of k, and its profit is $(1-k)\Pi^M$. For any k, we show that at $\hat{g}(k)$, the monopolist is indifferent between (q^M, t^M), and another mechanism that induces a positive mass of consumers to buy directly. As k increases, the profits from both mechanisms decline, but the profit from (q^M, t^M) declines more as it is entirely made up of residual demand. The other mechanism has some direct demand so it fares better in the face of increased capacity.

4.2 Non Monotonicity of Consumer Surplus in g

In this section, I show that increasing public quality g can make **every** consumer strictly worse-off. More precisely, I show that for every distribution of types F, for any capacity k, there exist two intervals of public quality, I_1 and I_2, such that the following holds: i) $g_1 < g_2$ for every $g_1 \in I_1$ and every $g_2 \in I_2$, and ii) keeping capacity fixed at k, increasing quality from g_1 to g_2 makes every type of consumer strictly worse-off.

I construct these two intervals using the threshold $\hat{g}(k)$ from Theorem 1. I show that for any g_1 and g_2 such that g_1 is slightly below $\hat{g}(k)$, and g_2 is slightly above $\hat{g}(k)$, if the difference between g_1 and g_2 is not too large, then *every* type of consumer is strictly better off at g_1. The main idea is simple. By Theorem 1, at $g_1 < \hat{g}(k)$, any optimal mechanism must have a positive mass of consumers buying directly from the monopolist. At $g_2 > \hat{g}(k)$ however, the unique optimal mechanism is such that every type demands the public option. So, if we keep capacity fixed at k, and increase quality from g_1 to g_2, congestion on the public

option increases. Further, for every type, her utility from buying from the monopolist is lower at g_2 than at g_1. This is because at g_2, the monopolist is not trying to attract *any* type directly, and has no incentive to offer favorable terms. At g_1, on the other hand, she *is* trying to attract some types directly. This not only makes the direct types better-off, but has a trickle-down effect for the types that are not buying directly, as they can imitate the types that are.

Before stating the result, I define what we mean by expected surplus for any type from the optimal mechanism. Fix (k, g). For any optimal mechanism (q^*, t^*) offered by the monopolist, I use $S^*(\theta, g, k)$ to denote the expected payoff of type θ from (q^*, t^*). As before, $U^*(\theta) = \theta q^*(\theta) - t^*(\theta)$, the payoff of type θ if she buys from the monopolist. If $U^*(\theta) \geqslant \theta g$, she prefers the monopolist to the public option, and $S^*(\theta, g, k) = U^*(\theta)$. If $U^*(\theta) < \theta g$, she prefers the public option and only buys from the monopolist if she cannot get it. If p is the probability with which she gets the public option, then $S^*(\theta, g, k) = p\theta g + (1-p)U^*(\theta)$. I now state the result.

Theorem 2. *Fix k and let $\hat{g}(k)$ be the full concession threshold at k. There exist an $\eta > 0$ such that, for every*

$$g_1 \in (\hat{g}(k) - \eta, \hat{g}(k)) \qquad \textit{and} \qquad g_2 \in (\hat{g}(k), \hat{g}(k) + \eta)$$

we have,

$$S^*(\theta, g_1, k) > S^*(\theta, g_2, k) \quad \textit{for every} \quad \theta > 0 \tag{4}$$

Proof Sketch: I want to show the existence of an interval $I := (\hat{g} - \eta, \hat{g} + \eta)$ around $\hat{g}$ such that 4 holds for every $g_1, g_2 \in I$ such that $g_1 < \hat{g} < g_2$. I leave the exact construction of the end points of the interval to the Appendix. Here, I provide a proof sketch of the following fact: if $g_1 < \hat{g}(k) < g_2$, g_1 is close enough to $\hat{g}(k)$, and $g_2 - g_1$ is small, then $S^*(\theta, g_1, k) > S^*(\theta, g_2, k)$ for every $\theta > 0$. In the Appendix, I will use this fact to constrict the required interval I.

From Theorem 1, at capacity k, for any $g > \hat{g}(k)$, the unique optimal mechanism is (q^M, t^M), and all types strictly prefer the public option. Thus, for any $g > \hat{g}(k)$, the expected surplus of any θ, across the public and private option, is $S^*(\theta, g, k) = k\theta g \; + \; (1-k)U^M(\theta)$, where $U^M(\theta) = \theta q^M(\theta) - t^M(\theta)$, the utility of type θ when (q^M, t^M) is offered and she buys from the monopolist. Theorem 1 also tells us that for every $g < \hat{g}(k)$, any optimal mechanism involves a positive mass of consumers buying directly from the monopolist. There are two possibilities: i) **all** types buy directly from the monopolist, and ii) a positive mass of types demand the public option. In both cases, I show that for g_1 slightly below $\hat{g}(k)$, and $g_2 > \hat{g}(k)$

and $g_2 - g_1$ small enough, we have $S^*(\theta, g_1, k) > S^*(\theta, g_2, k)$ for every $\theta > 0$. I now provide the proof sketch in steps.

Step 1: I first show that at g slightly below the threshold $\hat{g}(k)$, and any optimal mechanism (q^*, t^*) at g, we have $U^*(\theta) > U^M(\theta)$ for every $\theta > 0$ that buys with positive probability from the monopolist. Here, $U^*(\theta) = \theta q^*(\theta) - t^*(\theta)$. To see this, by Theorem 1, $\hat{g}(k) > g^*$, where $g^* = \frac{U^M(\overline{\theta})}{\overline{\theta}}$. If $g \in (g^*, \hat{g}(k))$, then $\theta g > U^M(\theta)$ for every $\theta > 0$. This is because by Lemma 17 in the Appendix, the function $\frac{U^M(\theta)}{\theta}$ is non-decreasing for $\theta > 0$.[14] It follows that for $g > g^*$, we must have $\theta g > U^M(\theta)$ for any $\theta > 0$. Now fix $g \in (g^*, \hat{g}(k))$, and let (q^*, t^*) be an optimal mechanism at g. If θ buys directly at (q^*, t^*), then $U^*(\theta) \geqslant \theta g > U^M(\theta)$. What about types that prefer the public option and fail to get it with positive probability? I show that the *fallback* utility of such types from (q^*, t^*), if they fail to get the public option, is also strictly higher than $U^M(\theta)$.

Lemma 4. *Fix any $g > g^*$, and an optimal mechanism (q^*, t^*) at g. Then, if (q^*, t^*) induces an interior cutoff c with $F(c) > k$, we must have that $U^*(\theta) \geqslant U^M(\theta)$ for every θ, where $U^*(\theta) = \theta q^*(\theta) - t^*(\theta)$.*

The proof is in the Appendix, but the intuition is that $U^*(\theta) > U^M(\theta)$ for the types that buy directly trickles down to the ones that don't.

Step 2: I now specify a useful inequality, and a useful lemma without proof. For any $g > g^*$, the following inequality puts a useful lower bound on the difference $\theta g - U^M(\theta)$.

$$\theta g - U^M(\theta) \geqslant (g - g^*)\theta \tag{5}$$

To see why 5 is true, $U^M(\theta) \leqslant \theta g$ for every θ by definition of g^*, and since $\frac{U^M(\theta)}{\theta}$ is non-decreasing. Subtracting both sides from θg, we get 5. The following lemma establishes that as g gets arbitrarily close to $\hat{g}(k)$, the cutoff corresponding to an optimal mechanism cannot approach $\overline{\theta}$, so the mass of types that buy directly from the monopolist is bounded away from zero.

Lemma 5. ***Top Separation Lemma****: There exist an $\eta_0 > 0$, and an $\varepsilon_0 > 0$ such that for every $g \in (\hat{g}(k) - \eta_0, \hat{g}(k))$, and for any optimal mechanism at g with an interior cutoff c, it must be that $c \leqslant \overline{\theta} - \varepsilon_0$.*

.

Step 3: I now come to the main part of the proof. I argue that if $g_1 \in (g^*, \hat{g}(k))$, $g_2 > \hat{g}(k)$, and $g_2 - g_1$ is *small*, then $S^*(\theta, g_1, k) > S^*(\theta, g_2, k)$ for every $\theta > 0$. First fix $g_1 \in (g^*, \hat{g}(k))$,

[14] In fact, $\frac{U^M(\overline{\theta})}{\overline{\theta}}$ is strictly increasing for all θ such that $q(\theta) > 0$.

and an optimal mechanism (q^*, t^*) at g_1. Suppose, at (q^*, t^*), public demand is lower than capacity. This can happen both when all types buy directly, so public demand is zero, or when the optimal mechanism induces an interior cutoff c, so $\theta < c$ demand the public option, but $F(c) < k$. I show that $S^*(\theta, g_1, k) > S^*(\theta, g_2, k)$ in both cases.

First, observe that $S^*(\theta, g_1, k) \geqslant \theta g_1$ for every θ. To see this, if θ buys directly, then she must be getting at least θg from the optimal mechanism. If she prefers the public option, she gets it with probability one, since public demand is lower than capacity. Now, fix a $g_2 > \hat{g}(k)$. Since $S^*(\theta, g_2, k) = k\theta g \ + \ (1-k)U^M(\theta)$, therefore for every θ:

$$\begin{aligned}
&S^*(\theta, g_1, k) - S^*(\theta, g_2, k) \\
&= S^*(\theta, g_1, k) - k\theta g_2 \ - (1-k)U^M(\theta) \\
&\geqslant (1-k)\theta(g_1 - U^M(\theta)) - k\theta(g_2 - g_1) \\
&\geqslant (1-k)\theta(g_1 - g^*) - k\theta(g_2 - g_1)
\end{aligned}$$

The first inequality follows from $S^*(\theta, g_1, k) \geqslant \theta g_1$. The second inequality comes from substituting 5 since $g_1 > g^*$. Thus, I have shown: $S^*(\theta, g_1, k) - S^*(\theta, g_2, k) \geqslant (1-k)\theta(g_1 - g^*) - k\theta(g_2 - g_1)$, which is strictly positive if $g_2 - g_1$ is small enough.

Step 4: Finally, suppose at g_1, the optimal mechanism (q^*, t^*) at g_1 induces public demand strictly greater than capacity. This happens when it induces cutoff c, and public demand $F(c) > k$. I now show that $S^*(\theta, g_1, k) > S^*(\theta, g_2, k)$ for this case. Since $\theta > c$ buy directly, they must get at least θg_1, so the analysis from ***Step*** 4 applies. For $\theta < c$, we have $S^*(\theta, g_1) = \frac{k}{F(c)}\theta g_1 \ + (1 - \frac{k}{F(c)})U^*(\theta)$, where $U^*(\theta)$ is θ's utility from the optimal mechanism at g_1.

The main worry here is what if $U^*(\theta)$, the *fallback* utility of type θ if she fails to get the public option, is strictly lower than $U^M(\theta)$. But Lemma 4 ensures this cannot be the case. Now consider η_0 from Lemma 5. Suppose that $g_1 > \max\{\hat{g} - \eta_0, g^*\}$. We are now ready to show that $S^*(\theta, g_1, k) > S^*(\theta, g_2, k)$.

$$\begin{aligned}
&S^*(\theta, g_1) - k\theta g_2 \ - (1-k)U^M(\theta) \\
&= \frac{k}{F(c)}\theta g_1 \ + (1 - \frac{k}{F(c)})U(\theta) - k\theta g_1 \ - \ (1-k)U^M(\theta) - k\theta(g_2 - g_1) \\
&= \frac{k(1-F(c))}{F(c)}(\theta g_1 - U^M(\theta)) + (1 - \frac{k}{F(c)})(U(\theta) - U^M(\theta)) - k\theta(g_2 - g_1)
\end{aligned}$$

$$\geqslant \frac{k(1-F(c))}{F(c)}\theta(g_1 - g^*) - k\theta(g_2 - g_1)$$
$$\geqslant \frac{k(1-F(\overline{\theta}-\varepsilon_0))}{F(\overline{\theta}-\varepsilon_0)}\theta(g_1 - g^*) - k\theta(g_2 - g_1)$$

where the second equality is just a rearrangement, the first inequality follows from 5, and Lemma 4 implies that $U(\theta) \geqslant U^M(\theta)$ for all $\theta \leqslant c$. Finally, the last inequality follows from Lemma 5. For small enough $g_2 - g_1$, the right side of the last inequality is strictly positive. This completes my proof sketch.

Looking at Theorem 4, we might worry that the surplus drop result relies on g being large enough that the monopolist stops competing with the public option entirely and direct demand drops to zero. However, the following three-type example shows that all types of consumers can become strictly worse-off as we increase g, even though at both the higher and lower values of g, a positive mass of consumers buy directly from the monopolist.

The consumers have three possible valuations for quality: high, given by $\theta_h = 1$, middle, given by $\theta_m = \frac{1}{2}$, and low, given by $\theta_l = \frac{1}{4}$. The high type consumers have mass $\mu_h = \frac{1}{5}$, and the low and medium types each have mass $\mu_m = \mu_l = \frac{2}{5}$. The capacity of the public option is fixed at $k = \frac{3}{8}$. Consider two qualities $g_1 = \frac{1}{25}$, and $g_2 = \frac{1}{20}$. At g_1, both θ_h and θ_m buy directly in the optimal mechanism, and at g_2, only θ_h buys directly. In this example, an increase from g_1 to g_2, makes **all** three types strictly worse off, even though, in both cases, a strictly positive mass of consumers buys directly from the monopolist.

The details of the example are in the Appendix, but as we increase public quality from g_1 to g_2, the following happens:

- θ_l demands the public option at both g_1 and g_2, and gets zero from the monopolist's mechanism in both cases. So, if it fails to get the public option, its utility is zero. At g_1, it is the only type demanding the public option, and gets it with probability $\frac{k}{\mu_l} = \frac{15}{16}$, but at g_2, both θ_l and θ_m demand it, so probability of success falls to $\frac{15}{32}$. The higher public quality is not enough to compensate for lower probability of success.

- θ_m buys directly at g_1, and indirectly at g_2. At g_1, the monopolist is trying to attract her directly, and she gets exactly $\theta_m g_1$. At g_2 however, she gets utility zero from the monopolist's mechanism. So, her expected payoff is $k\theta_m g_2$, which is strictly lower than her expected payoff at g_1.

- θ_h buys directly at both g_1 and g_2, but at g_2, it is the *only* direct type, so it loses out on the information rents she got at g_1, where θ_m was also buying directly. This loss in information rents overshadows the increase in g.

To summarize, the low-value consumers lose due to increased congestion. The middle-value consumers lose due losing direct status, combined with low probability of getting the public option. The high value consumers lose due to loss of information rents.

4.2.1 Capacity vs Quality

In my analysis till now, I kept capacity fixed and talked about the effects of improving quality. But what if we could increase capacity? How is it different from improving quality? In particular, suppose we want to improve the public option, with the aim of increasing (weighted) consumer surplus, and there is a trade-off between increasing capacity and improving quality. What should we do? An answer to this question would require a specification of the constraints under which we operate. Starting with any public option (k, g), what is the cost of increasing quality and capacity? What is the budget? I do not go into all that here, and leave a thorough formal analysis of which instrument is better for future work. Here, I briefly highlight how capacity and quality differ as instruments in that they affect the monopolist's incentives differently.

By Theorem 1, an increase in quality beyond a certain threshold leads the monopolist to capitulate completely. She stops competing for consumers and sells only to residual demand. Beyond the quality threshold $\hat{g}(k)$, an increase in quality has no impact on the incentives of the monopolist. By Theorem 1, for a fixed quality k, for every $g \geqslant \hat{g}(k)$, the optimal mechanism remains (q^M, t^M), and every type only buys from the monopolist it fails to get the public option. Therefore, an improvement in public quality beyond $\hat{g}(k)$ does not incentivize the monopolist to offer a "better" mechanism. Such an improvement does not alleviate the congestion on the public option. The only thing it achieves is that it increases the utility that consumers get *if* they succeed in obtaining the public option.

An increase in capacity, on the other hand, can reverse full capitulation, and thereby have a non-trivial impact on the monopolist's incentives. By Theorem 1, the full capitulation threshold $\hat{g}(k)$ is strictly increasing in k. So, starting from (k, g), where $g > \hat{g}(k)$ as we increase capacity to $k' > k$, full capitulation is reversed as long as $\hat{g}(k) < g < \hat{g}(k')$. In summary, an increase in quality makes it less attractive for the monopolist to attract consumers directly, and an increase in capacity reduces residual demand and makes competing directly more attractive.

4.3 Product Line Expansion

This section highlights how the presence of a capacity-constrained public option can strictly expand the set of consumers that the monopolist transacts with, relative to the pure monopoly

benchmark. The idea is that with the public option, the market is divided into two sub-markets: the consumers who buy directly, and those who constitute residual demand. The presence of residual demand results in a non-trivial redesign of the monopolist's menu, relative to the pure monopoly case. This is because the capacity constrained public option creates a *gap* between the two sub-markets. Those who buy directly must get high quality goods at reasonable prices. To be precise, any type θ that buys directly from the monopolist must get at least θg. Therefore, if the monopolist adds some low quality, low price products to sell to residual consumers, the higher, direct types will not be tempted to buy them. So, selling to consumers that only buy after failing to get the public option does not create additional information rents for direct types.

To understand this better, suppose in the pure monopoly case, the monopolist offers mechanism (q^M, t^M) such that $q^M(\theta) > 0$ if and only if $\theta > a$, where $a \in (\underline{\theta}, \overline{\theta})$. Now, suppose, in our model, the optimal mechanism (q^*, t^*) induces a cutoff $c > a$: types $\theta > c$ buy directly, and $\theta < c$ only buy if they fail to get the public option. So, at the optimal mechanism, we must have $U^*(c) = cg$, since the cutoff type must be indifferent. Consider $\theta = c - \varepsilon$. By simply imitating c, it gets a positive payoff. So, even though $c - \varepsilon$ does not buy directly, she would definitely buy from the monopolist if she doesn't get the public option. So, in the optimal mechanism, it must be that the monopolist sells to a positive mass of consumers below c. The next theorem shows that the the monopolist might even sell to $\theta < a$, thereby expanding her coverage relative to the monopoly case.

Assumption 1. ***Monopoly Benchmark****: We assume that the distribution F on $[\underline{\theta}, \overline{\theta}]$ has a continuous density f, that satisfies $0 < \underline{f} \leqslant f(\theta) \leqslant \overline{f} < \infty$ and the monopoly virtual value $\varphi(\theta) := \theta - \frac{1-F(\theta)}{f(\theta)}$ is continuous, strictly increasing, and there exists an $a \in (\underline{\theta}, \overline{\theta})$ such that $\varphi(a) = 0$.*

Under Assumption 1, which is satisfied, for e.g., by uniform distribution, the monopoly optimal allocation is:

$$q^M(\theta) = [\varphi(\theta)]_+ = \begin{cases} \varphi(\theta) & \text{if } \theta \geqslant a \\ 0 & \text{if } \theta < a \end{cases}$$

Define$\gamma(\theta) := \frac{U^{MR}(\theta)}{\theta}$. We know that $\frac{U^{MR}(\theta)}{\theta}$ is strictly increasing at every θ such that $q^M(\theta) > 0$, so γ is strictly increasing over $(a, \overline{\theta}]$. We now ready to state our theorem.

Theorem 3. *Suppose F satisfies Assumption 1, and that $k < F(a)$, where a satisfies $\varphi(a) = 0$. Then, there exists a $c_0 < \overline{\theta}$, such that for every public quality $g \in I := (\gamma(c_0), \gamma(\overline{\theta}))$, we have the following:*

- *Every optimal mechanism induces an interior congested cutoff.*

- *In every optimal mechanism, $q^*(\theta) > 0$, i.e., there exists a $b < a$ such that $q^*(\theta) > 0$ for all $\theta \in (b, a)$.*

Note that the interval I is independent of k, is is determined using only F. Therefore, for every $(k, g) \in (0, F(a)) \times I$, the optimal mechanism must involve strict downward expansion relative to the monopoly outcome.

5 Conclusion

In this paper, I describe a model where a profit maximizing monopolist competes against a free, but capacity-constrained public option. I show that at a fixed capacity, increasing the quality of the public option can have an anti-competitive effect. Instead of offering higher qualities or better prices to compete with the public option, the monopolist might stop competing altogether and simply offer her monopoly menu, which she would offer if there was no public option. In this case, every consumer prefers the public option, and the monopolist sells only to residual demand. Such an increase in quality can make every consumer worse off: due to increased congestion, the probability of obtaining the public option is low, and if they fail to get it, they have to buy at monopoly prices. Moreover, the quality threshold beyond which this anti-competitive effect takes over, is strictly increasing in capacity. At a low capacity, therefore, even at a low public quality, the monopolist might not compete for any type directly.

This result has implications for public policy. For example, suppose that the government's intention is to provide a safety net for low-value consumers, who would otherwise be excluded by the monopolist. Then, a very high quality public option with low capacity makes for particularly bad safety net. Due to high quality of the public option, the private sector might not attempt to compete with it and keep its prices high, knowing that a large mass of consumers will anyway be forced to buy from it as a last resort. Instead of creating a safety net for low types, all this public option does is create option value for high types, who can anyway buy from the private sector and get positive utility. For the low types, the public option was their only chance at access to the good. The combination of high quality and low capacity means that they are crowded out by high types, who were not the intended recipients of the public option.

I also show that the presence of a capacity-constrained public option creates a separate sub-market for residual demand and might cause the monopolist to introduce products to specifically cater to residual demand. In particular, surprisingly, the presence of a competitor in form of a public option, can lead the monopolist to increase her market coverage, relative to the pure monopoly case.

In future work, I would like to explore the effects of increasing capacity more fully. In particular, we know that around the full capitulation threshold, increasing capacity always makes all consumers strictly better off. But we do not know how the seller's optimal mechanism changes with changes in capacity at a small public quality. Furthermore, if the public sector is a Stackelberg leader and chooses the capacity and quality of the public option subject to budget constraints, what would it choose if it wants to maximize weighted consumer surplus? The solution to this redistributive mechanism design problem would have policy implications for which is the optimal way of using the two instruments, capacity and quality, to increase consumer surplus.

References

Akbarpour, M., Dworczak, P., & Kominers, S. D. (2024). Redistributive allocation mechanisms. *Journal of Political Economy*, *132*(6), 1831–1875. Retrieved from https://pdworczak.com/wp-content/uploads/2026/08/a6eed-redistributive-allocation-mechanisms.pdf (Authors listed in randomized order. Accessible author manuscript dated June 22, 2023) doi: 10.1086/728111

Besley, T., & Coate, S. (1991). Public provision of private goods and the redistribution of income. *American Economic Review*, *81*(4), 979–984. Retrieved from https://www.jstor.org/stable/2006658

Bhate-Deosthali, P., Khatri, R., & Wagle, S. (2011). Poor standards of care in small, private hospitals in Maharashtra, India: implications for public–private partnerships for maternity care. *Reproductive Health Matters*, *19*(37), 32–41. Retrieved from https://doi.org/10.1016/S0968-8080(11)37560-X doi: 10.1016/S0968-8080(11)37560-X

Blackorby, C., & Donaldson, D. (1988). Cash versus kind, self-selection, and efficient transfers. *American Economic Review*, *78*(4), 691–700. Retrieved from https://ideas.repec.org/a/aea/aecrev/v78y1988i4p691-700.html

Bose, S., Ghosh, P., Sardana, A., & Boda, M. (2021). *Regulation and informal market for schools in Delhi* (Working Paper No. 340). National Institute of Public Finance and Policy. Retrieved from https://www.nipfp.org.in/media/documents/WP_340_2021.pdf

Coate, S., Johnson, S., & Zeckhauser, R. (1994). Pecuniary redistribution through in-kind programs. *Journal of Public Economics*, *55*(1), 19–40. Retrieved from https://rzeckhauser.scholars.harvard.edu/sites/g/files/omnuum4441/files/rzeckhauser/files/pecuniary_redistribution_through_in-kind_programs.pdf doi: 10.1016/0047-2727(94)90078-7

Dworczak, P., Kominers, S. D., & Akbarpour, M. (2021). Redistribution through markets. *Econometrica*, *89*(4), 1665–1698. Retrieved from https://web.stanford.edu/~mohamwad/Inequality.pdf (Authors listed in randomized order) doi: 10.3982/ECTA16671

Epple, D., & Romano, R. E. (1996). Public provision of private goods. *Journal of Political Economy*, *104*(1), 57–84. Retrieved from https://doi.org/10.1086/262017 doi: 10.1086/262017

Gahvari, F., & Mattos, E. (2007). Conditional cash transfers, public provision of private goods, and income redistribution. *American Economic Review*, *97*(1), 491–502. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.97.1.491 doi: 10.1257/aer.97.1.491

Ghosh, P., & Bose, S. (2022). *Estimating the excess demand for government schools in Delhi: How much capacity creation is necessary?* (Working Paper No. 387). National Institute of Public Finance and Policy. Retrieved from https://www.nipfp.org.in/media/documents/WP_387_2022.pdf

Grassi, S., & Ma, C.-t. A. (2011). Optimal public rationing and price response. *Journal of Health Economics*, *30*(6), 1197–1206. Retrieved from https://people.bu.edu/ma/Papers_Archive/rationing-price_in-press.pdf doi: 10.1016/j.jhealeco.2011.08.011

Grassi, S., & Ma, C.-t. A. (2012). Public sector rationing and private sector selection. *Journal of Public Economic Theory*, *14*(1), 1–34. Retrieved from https://people.bu.edu/ma/rationing-selection.pdf doi: 10.1111/j.1467-9779.2011.01532.x

House of Commons Housing, Communities and Local Government Committee. (2026). *Housing conditions in the private rented sector* (Fourth Report of Session 2026–27 No. HC 42). House of Commons. Retrieved from https://publications.parliament.uk/pa/cm5902/cmselect/cmcomloc/42/report.html (Published 3 July 2026)

Jullien, B. (2000). Participation constraints in adverse selection models. *Journal of Economic Theory*, *93*(1), 1–47.

Kang, Z. Y. (2023). *The public option and optimal redistribution.* Retrieved from https://ziyangkang.com/files/jmp.pdf (Working paper, January 2023)

Kang, Z. Y., & Watt, M. (2024). *Optimal redistribution through subsidies.* Retrieved from https://ziyangkang.com/files/ORTS.pdf (Working paper, November 12)

Kang, Z. Y., & Watt, M. (2026). *Topping up and optimal redistribution.* Retrieved from https://ziyangkang.com/files/TUOR.pdf (Working paper, June 2026)

Klumpp, T., & Su, X. (2019). Price–quality competition in a mixed duopoly. *Journal of Public Economic Theory*, *21*(3), 400–432. Retrieved from https://doi.org/10.1111/jpet.12357 doi: 10.1111/jpet.12357

Laine, L. T., & Ma, C.-t. A. (2017). Quality and competition between public and private firms. *Journal of Economic Behavior & Organization*, *140*, 336–353. Retrieved from https://doi.org/10.1016/j.jebo.2017.05.012 doi: 10.1016/j.jebo.2017.05.012

Ma, C.-t. A. (2004). Public rationing and private cost incentives. *Journal of Public Economics*, *88*(1–2), 333–352. Retrieved from https://people.bu.edu/ma/Papers_Archive/JPubE_2004.pdf doi: 10.1016/S0047-2727(02)00163-9

Martimort, D., & Stole, L. A. (2022). Participation constraints in discontinuous adverse selection models. *Theoretical Economics*, *17*(3), 1145–1181. Retrieved from https://doi.org/10.3982/TE3030 doi: 10.3982/TE3030

Mekonnen, T. (2026). Luck out or outpay? competing with a public option. *American Economic Journal: Microeconomics*. Retrieved from https://arxiv.org/abs/2507.12779 (Forthcoming. July 7, 2026 manuscript; year refers to the manuscript) doi: 10.1257/mic.20250244

Milgrom, P., & Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, *70*(2), 583–601.

Mussa, M., & Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, *18*(2), 301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, *6*(1), 58–73. doi: 10.1287/moor.6.1.58

Nichols, A. L., & Zeckhauser, R. J. (1982). Targeting transfers through restrictions on recipients. *American Economic Review*, *72*(2), 372–377. Retrieved from https://www.jstor.org/stable/1802361

Oxfam Hong Kong, & Policy 21 Limited. (2012). *Research on the living conditions of tenant households who have been on the waiting list for public rental housing for over 3 years* (Report). Oxfam Hong Kong. Retrieved from https://www.oxfam.org.hk/f/news_and_publication/1404/content_16714en.pdf (December 2012)

Pankaj, A., Poornima M, & Kaur, R. (2021). *Reach and role of private schools in India* (Report). Council for Social Development. Retrieved from https://csdindia.org/wp-content/uploads/2024/05/Reach-and-role-of-private-schools-in-India.pdf (November 2021)

Ranasinghe, A., & Su, X. (2023). When social assistance meets market power: A mixed duopoly view of health insurance in the United States. *Economic Inquiry*, *61*(4), 851–869. Retrieved from https://sites.ualberta.ca/~xuejuan1/docs/social%20assistance.pdf doi: 10.1111/ecin.13158

Rochet, J.-C., & Stole, L. A. (2002). Nonlinear pricing with random participation. *The Review of Economic Studies*, *69*(1), 277–311. Retrieved from https://doi.org/10.1111/1467-937X.00206 doi: 10.1111/1467-937X.00206

Siu, T., & Jim, C. (2025). *Hong Kong struggles to improve conditions in tiny, crowded homes.* Reuters. Thomson Reuters. Retrieved from https://whtc.com/2025/01/06/hong-kong-struggles-to-improve-conditions-in-tiny-crowded-homes/ (6 January 2025; Reuters report republished by WHTC)

Srivastav, N., Pant, L., Priya, A., & Coffey, D. (2023). Understanding high mortality among private facility births in rural Uttar Pradesh. *Economic & Political Weekly*, *58*(10), 52–59. Retrieved from https://www.epw.in/journal/2023/10/special-articles/understanding-high-mortality-among-private.html (11 March 2023)

Toikka, J. (2011). Ironing without control. *Journal of Economic Theory*, *146*(6), 2510–2526. doi: 10.1016/j.jet.2011.06.003

6 Appendix

6.1 Monopolist's Problem

In this section, I first argue that the Revelation Principle holds, and the monopolist can restrict attention to direct, incentive compatible (IC), and individually rational (IR) mechanisms characterized by Lemma 1 in the main text. Then, I prove Proposition 1 that shows that every optimal mechanism has a cutoff structure. Note that I haven't shown existence of an optimal mechanism yet. I will do that in the next section.

Revelation Principle: Suppose the monopolist can commit to any arbitrary extensive form, and any strategy within that extensive form. However, she has to take two things as given. Firstly, consumers cannot be forced to choose between participating in the monopolist's extensive form, and choosing the public option: they can participate in the extensive form *after* failing to get the public option. Secondly, the monopolist cannot distinguish between consumers who come to her directly, and those who come after trying and failing to get the public option. So, she cannot condition her strategy on this fact. Any extensive form and any strategy chosen by the monopolist and the best response of the consumers, induces, for every type θ, a distribution $\mu_\theta \in \Delta(\mathbb{R}_+ \times \mathbb{R})$ over quality-price pairs (q, t). Let $U(\theta) := \mathbb{E}_{\mu_\theta}[\theta q - t]$ be type θ's expected utility from μ_θ. Because consumers are risk-neutral, and the monopolist has quadratic costs, she can earn (weakly) higher profits by simply asking every consumer

to report her type, and map report θ to $(q(\theta), t(\theta))$, where $q(\theta) := \mathbb{E}_{\mu_\theta}[q]$, and $t(\theta) := \mathbb{E}_{\mu_\theta}[t]$. This proves the Revelation Principle.

Discussion: Characterization of IC and IR mechanisms. The characterization of IC and IR mechanisms in Lemma 1 is the same as the pure monopoly case, where there is no public option. I first argue why the characterization of IC mechanisms is the same. This is because the public option only changes θ's decision, of whether to buy directly, or after failing to get the public option. It does not change how she compares two pairs (q, t) and (q', t'), *conditional* on buying from the monopolist. To see this, let $U^(\theta) := \max_{\{\theta' \in \Theta\}}\{\theta q(\theta') - t(\theta')\}$: this is θ's payoff if she buys from the monopolist, either directly, or after failing to get the public option. The comparison with the public option is determined by the sign of $U(\theta) - \theta g$. But once the consumer decides to buy from the seller, g is irrelevant. From any (q, t), type θ gets $\theta q(\theta) - t(\theta)$. It is important here that the monopolist cannot observe whether or not a consumer tried to get the public option. This means that she cannot condition what she offers the consumers for any report on this.

IR is more subtle. An IC mechanism $(q(.), t(.))$ is IR if and only if for every θ, we have $U(\theta) := \theta q(\theta) - t(\theta) \geqslant 0$, i.e. conditional on buying from the monopolist, every consumer gets a non-negative payoff. Of course, when there is a public option, if $0 \leqslant U(\theta) < \theta g$, type θ will not buy directly from the monopolist. But $U(\theta) \geqslant \theta g$ cannot be the notion of IR here, because the monopolist may not want to attract every type directly. The sign of $U(\theta) - \theta g$ is endogenous, and can be negative for some types in an optimal mechanism. However, what must always be true is that the monopolist cannot force any consumer to purchase, either before or after trying for the public option, if she gets a negative utility from buying. So, the notion of IR is the same as the monopoly case: $U(\theta) \geqslant 0$ for all θ.

The proof of Lemma 1 follows from standard arguments, and I will not provide it here. I now prove Proposition 1.

Proof of Proposition 1: To prove the cutoff structure for optimal mechanisms, first observe that in an optimal mechanism, we must have $\pi(\theta) \geqslant 0$ for every θ. If not, i.e. for any θ, we have $\pi(\theta) < 0$, we can, keeping the mapping for the other types unchanged, map θ to $\arg\max_{((q(\theta'),t(\theta'))|\pi(\theta') \geqslant 0\}}\{\theta q(\theta') - t(\theta')\}$. The following lemma then proves the cutoff structure:

Lemma 6. *Fix and optimal mechanism (q^*, t^*). Then, if $U^*(\theta) \geqslant \theta g$ for any $\theta > 0$, then i) $q^*(\theta) > g$, and ii) $U^*(\theta') > \theta' g$ for all $\theta' > 0$.*

Proof. For part i), since $U^*(\theta) \geqslant \theta g$, we have that $\theta q^*(\theta) - t^*(\theta) \geqslant \theta g$. Since this is an optimal mechanism, $\pi^*(\theta) = t^*(\theta) - \frac{q^*(\theta)^2}{2} \geqslant 0$, which implies that $t^*(\theta) \geqslant \frac{q^*(\theta)^2}{2}$. This, combined with $\theta q^*(\theta) - t^*(\theta) \geqslant \theta g$, gives us $\theta(q^*(\theta) - g) \geqslant \frac{q^*(\theta)^2}{2} \geqslant 0$. Now, $q^*(\theta)$ cannot

be 0, because that would imply $-\theta g > 0$. So, we must have $q^*(\theta) > 0$, which implies that $\theta(q^*(\theta) - g) \geqslant \frac{q^*(\theta)^2}{2} > 0$, so $q^*(\theta) > g$. Part ii) follows directly. □

Lemma 6 proves that if any type θ buys directly, then all $\theta' > \theta$ must buy directly, thereby proving the cutoff structure. I now argue that in any optimal mechanism (q^*, t^*), $\{\theta | U^*(\theta) = \theta g\}$ is an F-null set. To see this, suppose not, i.e. $U(\theta) = \theta g$ for a positive measure of types. Since $U(\theta)$ is convex, therefore $U(\theta) - \theta g$ is also convex, which means that there exists a non empty interval $[a, b]$ such that $U(\theta) = \theta g$ for all $\theta in [a, b]$. However, this implies that $(q^*(\theta), t^*(\theta)) = (g, 0)$ for all $\theta \in [a, b]$. To see this, fix $\theta' \in (a, b)$. We know that $\theta' q(\theta') - t(\theta') = \theta' g \implies \theta'(q(\theta') - g) = t(\theta')$. If $q(\theta') > g$, then θ'' slightly above θ would get a utility strictly greater than $\theta'' g$ from $(q(\theta'), t(\theta'))$. If If $q(\theta') < g$, then θ'' slightly below θ would get a utility strictly greater than $\theta'' g$ from $(q(\theta'), t(\theta'))$. This cannot be, so we must have $q(\theta') = g$ and $t(\theta') = 0$ for every $\theta' \in (a, b)$. But this violates $\pi^*(\theta) \geqslant 0$ for $\theta \in (a, b)$. So, $\{\theta | U^*(\theta) = \theta g\}$ must be an F-null set. This completes the proof of Proposition 1.

6.2 Existence of Optimal Mechanism

In this section, I show that an optimal mechanism exists. Before that, I define some notation, and a preliminary lemma.

Lemma 7. *Fix an optimal mechanism (q^*, t^*), and let $U^*(\theta) := \theta q^*(\theta) - t^*(\theta)$. Then, we must have i) $q^*(\theta) \leqslant 2\theta$ for all θ, and ii) $U^*(\underline{\theta}) \leqslant \frac{\underline{\theta}^2}{2}$.*

Proof. Let $\pi^*(\theta) = t^*(\theta) - \frac{q^*(\theta)^2}{2}$. To see why i) hold, in an optimal mechanism, $\pi^*(\theta) \geqslant 0$ for every θ, which implies that $t^*(\theta) \leqslant \frac{q^*(\theta)^2}{2}$. In addition, the mechanism must be IR, which means that $U^*(\theta) \geqslant 0 \implies t^*(\theta) \leqslant \theta q^*(\theta)$: together with $t^*(\theta) \geqslant \frac{q^*(\theta)^2}{2}$, this implies that $q^*(\theta) \leqslant 2\theta$. To see why ii) holds, for any θ, and therefore for $\underline{\theta}$, we have: $t^*(\theta) \leqslant \frac{q^*(\theta)^2}{2} \implies U^*(\theta) \leqslant \theta q^*(\theta) - \frac{q^*(\theta)^2}{2} \leqslant \frac{\theta^2}{2}$. This proves ii). □

I now define some notation. In any IC mechanism, by the envelope formula of Milgrom & Segal (2002), the transfers are completely determined by the allocation rule, up to a constant, which is $U(\underline{\theta})$, the utility of the lowest type. I use $\underline{u}$ to denote $U(\underline{\theta})$ henceforth. By Lemma 7, the space over which I maximize the monopolist's profit is:

$$\begin{aligned} \mathcal{K} = \Big\{ (q, \underline{u}) : q \in L^2(F) \text{and non-negative and non-decreasing,} \\ 0 \leqslant \underline{u} \leqslant \frac{\underline{\theta}^2}{2}, \qquad 0 \leqslant q(\theta) \leqslant 2\theta \quad F\text{-a.e.,} \\ \theta q(\theta) - U(\theta) - q(\theta)^2/2 \geqslant 0 \quad F\text{-a.e.} \Big\}. \end{aligned} \tag{6}$$

Here, the restrictions $0 \leqslant \underline{u} \leqslant \frac{\theta^2}{2}$ and $0 \leqslant q(\theta) \leqslant 2\theta$ are by Lemma 7, and $\theta q(\theta) - U(\theta) - q(\theta)^2/2 \geqslant 0$ is just $\pi(\theta) \geqslant 0$. I equip $\mathcal{K}$ with the metric $d : \mathcal{K} \times \mathcal{K} \to [0, \infty)$ defined by

$$d\big((q, \underline{u}), (\widetilde{q}, \widetilde{\underline{u}})\big) := |\underline{u} - \widetilde{\underline{u}}| + \left(\int_{\underline{\theta}}^{\overline{\theta}} |q(\theta) - \widetilde{q}(\theta)|^2 \, dF(\theta) \right)^{1/2}. \tag{7}$$

Let D denote the mass of consumers who demand the public option. If the demand for the public option is D, then the probability of getting the public option is denoted by:

$$p(k, D) = \frac{k}{\max\{k, D\}} = \begin{cases} 1, & D \leqslant k, \\ k/D, & D > k. \end{cases} \tag{8}$$

For $(q, \underline{u}) \in \mathcal{K}$, the demand for the public option, monopolist's profit, the monopolist's profit if everyone buys directly from her, and expected consumer surplus across the public and the private option are respectively:

$$D(q, \underline{u}; g) = \int_{\underline{\theta}}^{\overline{\theta}} \mathbf{1}_{\{U(\theta) < \theta g\}} \, dF(\theta), \tag{9}$$

$$\Pi(q, \underline{u}; k, g) = \int_{\underline{\theta}}^{\overline{\theta}} \big[1 - p(k, D(q, \underline{u}; g)) \mathbf{1}_{\{U(\theta) < \theta g\}}\big] \pi(\theta) \, dF(\theta), \tag{10}$$

$$\Pi^{\text{mon}}(q, \underline{u}) = \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta) \, dF(\theta), \tag{11}$$

$$\mathrm{S}(\theta; q, \underline{u}, k, g) = U(\theta) + p(k, D(q, \underline{u}; g))[\theta g - U(\theta)]_+. \tag{12}$$

Here $[x]_+ = \max\{x, 0\}$ and $\mathbf{1}_A$ is the indicator function of A. The transaction probability in (10) is one for direct buyers and $1 - p(k, D)$ for those who prefer the public option. By Lemma 1, the set if types which are indifferent between the public and the private option is a null set. So, all the above expressions are independent of routing for these types. The profit from the optimal mechanism, if it exists is:

$$\Pi^*(k, g) := \sup_{(q, \underline{u}) \in \mathcal{K}} \Pi(q, \underline{u}; k, g). \tag{13}$$

I now show that the supremum is attained. Before that, I prove two useful lemmas.

Lemma 8. *The metric space $(\mathcal{K}, d)$ is nonempty and compact, where d is defined in (7). For*

any $(q, \underline{u}), (\widetilde{q}, \widetilde{\underline{u}}) \in \mathcal{K}$*, we have:*

$$\sup_{\theta \in [\underline{\theta}, \overline{\theta}]} |U(\theta) - \widetilde{U}(\theta)| \leqslant |\underline{u} - \widetilde{\underline{u}}| + \sqrt{\frac{\overline{\theta} - \underline{\theta}}{\underline{f}}} \left(\int_{\underline{\theta}}^{\overline{\theta}} |q - \widetilde{q}|^2 \, dF \right)^{1/2}. \tag{14}$$

Proof. I first show 14. The envelope formula gives, for every θ,

$$\begin{aligned}
|U(\theta) - \widetilde{U}(\theta)| &\leqslant |\underline{u} - \widetilde{\underline{u}}| + \int_{\underline{\theta}}^{\overline{\theta}} |q(s) - \widetilde{q}(s)| \, ds \\
&= |\underline{u} - \widetilde{\underline{u}}| + \int_{\underline{\theta}}^{\overline{\theta}} \left(|q - \widetilde{q}| \sqrt{f} \right) \frac{1}{\sqrt{f}} \, ds \\
&\leqslant |\underline{u} - \widetilde{\underline{u}}| + \left(\int_{\underline{\theta}}^{\overline{\theta}} |q - \widetilde{q}|^2 \, dF \right)^{1/2} \left(\int_{\underline{\theta}}^{\overline{\theta}} \frac{ds}{f(s)} \right)^{1/2} \\
&\leqslant |\underline{u} - \widetilde{\underline{u}}| + \sqrt{\frac{\overline{\theta} - \underline{\theta}}{\underline{f}}} \left(\int_{\underline{\theta}}^{\overline{\theta}} |q - \widetilde{q}|^2 \, dF \right)^{1/2}.
\end{aligned}$$

The third inequality is by Cauchy Schwarz. Both integrals are finite because the allocations are square integrable and $f \geqslant \underline{f} > 0$. The last inequality uses $1/f \leqslant 1/\underline{f}$. Taking supremum proves 14.

I now show that for every sequence $(q_n, \underline{u}_n) \in \mathcal{K}$, there is a subsequence n_j and a pair $(q, \underline{u}) \in \mathcal{K}$ such that $\underline{u}_{n_j} \longrightarrow \underline{u}$, and $\int_{\underline{\theta}}^{\overline{\theta}} |q_{n_j} - q|^2 \, dF \longrightarrow 0$. This would prove compactness. Since $(0, 0)$ belongs to $\mathcal{K}$, therefore $\mathcal{K}$ is nonempty. Take an arbitrary sequence $(q_n, \underline{u}_n) \in \mathcal{K}$. By 6, $0 \leqslant \underline{u}_n \leqslant \underline{\theta}^2/2$. By Bolzano Weierstrass, every bounded real sequence has a convergent subsequence, therefore, we have a subsequence of $(q_n, \underline{u}_n)$ for which $\underline{u}_n \longrightarrow \underline{u} \in [0, \underline{\theta}^2/2]$. Here, we have relabeled the subsequence by n. By Helly's selection theorem, we have a further subsequence (again relabeled n) and a non-negative, non-decreasing q with $q \leqslant 2\overline{\theta}$ such that $q_n \to q$, F-a.e. Moreover, since $|q_n(\theta) - q(\theta)|^2 \leqslant 4\overline{\theta}^2$, therefore the dominated convergence theorem gives us $\int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2 \, dF \longrightarrow 0$. Let $U(\theta) := \underline{u} + \int_{\underline{\theta}}^{\theta} q(s) \, ds$. Since $\int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2 \, dF \longrightarrow 0$, therefore by 14, we have

$$\sup_{\theta} |U_n(\theta) - U(\theta)| \leqslant |\underline{u}_n - \underline{u}| + \sqrt{\frac{\overline{\theta} - \underline{\theta}}{\underline{f}}} \left(\int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2 \, dF \right)^{1/2} \longrightarrow 0. \tag{15}$$

Thus, by 15, we have that $\theta q(\theta) - U(\theta) - \frac{q(\theta)^2}{2} = \lim_n \left[\theta q_n(\theta) - U_n(\theta) - \frac{q_n(\theta)^2}{2} \right] \geqslant 0$. We already have $0 \leqslant q(\theta) = \lim_n q_n(\theta) \leqslant 2\theta$, so the limit $(q, \underline{u})$ lies in the class $\mathcal{K}$. This proves compactness. □

I now prove a second useful lemma.

Lemma 9. *The monopolist's profit from $(q, \underline{u}) \in \mathcal{K}$, if all types buy from her, Π^{mon}, as defined in 11, is continuous on $\mathcal{K}$. Moreover, let $\{(q_n, \underline{u}_n)\}, (q, \underline{u}) \in \mathcal{K}$ and suppose*

$$\underline{u}_n \to \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2 \, dF \to 0, \qquad (k_n, g_n) \to (k, g) \in (0, 1] \times (0, \infty), \tag{16}$$

where every $(k_n, g_n) \in 0, 1] \times (0, \infty)$. Then, the following holds:

$$D(q_n, \underline{u}_n; g_n) \to D(q, \underline{u}; g),$$
$$\Pi(q_n, \underline{u}_n; k_n, g_n) \to \Pi(q, \underline{u}; k, g).$$

Proof. Fix a sequence $\{(q_n, \underline{u}_n)\}_{n \in \mathbb{N}}$ in $\mathcal{K}$, and an element $(q, \underline{u}) \in \mathcal{K}$ such that 16 holds.

Step 1: I first show that $D(q_n, \underline{u}_n; g_n) \to D(q, \underline{u}; g)$, where $D(.)$ is as defined in 9. Let $U_n := \underline{u}_n + \int_{\underline{\theta}}^{\theta} q_n(s) \, ds$ and $\pi_n(\theta) := \theta q_n(\theta) - U_n(\theta) - \frac{q_n(\theta)^2}{2}$, and let $U(.)$ and $\pi(.)$ denote the corresponding functions for $(q, \underline{u})$. From Proposition 1, we know that $F\{\theta : U(\theta) = \theta g\} = 0$. Thus, almost everywhere, we have that:

$$U_n(\theta) - \theta g_n \longrightarrow U(\theta) - \theta g \neq 0.$$

Since weak inequalities are preserved under limits, we have:

$$\mathbf{1}_{\{U_n(\theta) < \theta g_n\}} \longrightarrow \mathbf{1}_{\{U(\theta) < \theta g\}} \quad F\text{-a.e.} \tag{17}$$

Dominated convergence theorem applied to 9 gives

$$D(q_n, \underline{u}_n; g_n) = \int \mathbf{1}_{\{U_n < \theta g_n\}} \, dF \longrightarrow \int \mathbf{1}_{\{U < \theta g\}} \, dF = D(q, \underline{u}; g). \tag{18}$$

This completes the proof for $D(q_n, \underline{u}_n; g_n) \to D(q, \underline{u}; g)$. I prove another useful thing for later. By (8), (18), and $k > 0$ we have:

$$p(k_n, D(q_n, \underline{u}_n; g_n)) = \frac{k_n}{\max\{k_n, D(q_n, \underline{u}_n; g_n)\}} \longrightarrow \frac{k}{\max\{k, D(q, \underline{u}; g)\}} = p(k, D(q, \underline{u}; g)). \tag{19}$$

Thus the probability of getting the public option also converges.

Step 2: I now show a couple of useful things before showing that $\Pi^{\text{mon}}(q_n, \underline{u}_n) \to \Pi^{\text{mon}}(q, \underline{u})$. By 6, $0 \leqslant q_n, q \leqslant 2\overline{\theta}$ a.e. so $\frac{1}{2}|q_n^2 - q^2| = \frac{1}{2}|q_n - q|(q_n + q) \leqslant 2\overline{\theta}|q_n - q|$. For almost every θ,

we have:

$$\begin{aligned}|\pi_n(\theta) - \pi(\theta)| &= \left|\theta(q_n(\theta) - q(\theta)) - (U_n(\theta) - U(\theta)) - \tfrac{1}{2}(q_n(\theta)^2 - q(\theta)^2)\right| \\ &\leqslant 3\overline{\theta}|q_n(\theta) - q(\theta)| + |U_n(\theta) - U(\theta)|.\end{aligned}$$

Integrating, and applying Cauchy Schwarz,

$$\begin{aligned}\int_{\underline{\theta}}^{\overline{\theta}} |\pi_n - \pi|\, dF &\leqslant 3\overline{\theta} \int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|\, dF + \sup_{\theta} |U_n(\theta) - U(\theta)| \\ &\leqslant 3\overline{\theta} \left(\int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2\, dF \right)^{1/2} + \sup_{\theta} |U_n(\theta) - U(\theta)| \longrightarrow 0. \end{aligned} \tag{20}$$

The first terms tends to zero by allocation convergence, and the second by (14). Thus, (11) gives us:

$$|\Pi^{\text{mon}}(q_n, \underline{u}_n) - \Pi^{\text{mon}}(q, \underline{u})| = \left| \int (\pi_n - \pi)\, dF \right| \leqslant \int |\pi_n - \pi|\, dF \longrightarrow 0.$$

This completes the proof for **Step** 2.

Step 3: I now show that $\Pi(q_n, \underline{u}_n; k_n, g_n) \to \Pi(q, \underline{u}; k, g)$, where $\Pi(.)$ is as defined in 10. Firstly,

$$\begin{aligned}&\Pi(q_n, \underline{u}_n; k_n, g_n) - \Pi(q, \underline{u}; k, g) \\ &= \int \left[1 - p(k_n, D(q_n, \underline{u}_n; g_n))\mathbf{1}_{\{U_n < \theta g_n\}}\right](\pi_n - \pi)\, dF \\ &\quad + \int \left[p(k, D(q, \underline{u}; g))\mathbf{1}_{\{U < \theta g\}} - p(k_n, D(q_n, \underline{u}_n; g_n))\mathbf{1}_{\{U_n < \theta g_n\}}\right]\pi\, dF.\end{aligned}$$

From the triangle inequality, we have:

$$\begin{aligned}&|\Pi(q_n, \underline{u}_n; k_n, g_n) - \Pi(q, \underline{u}; k, g)| \\ &\leqslant \int |\pi_n - \pi|\, dF \\ &\quad + \int \left|p(k_n, D(q_n, \underline{u}_n; g_n))\mathbf{1}_{\{U_n < \theta g_n\}} - p(k, D(q, \underline{u}; g))\mathbf{1}_{\{U < \theta g\}}\right| \pi\, dF.\end{aligned} \tag{21}$$

The first integral tends to zero by 20. In the second integral, the absolute difference goes to zero a.e. by 17 and 19. So:

$$0 \leqslant \left|p(k_n, D(q_n, \underline{u}_n; g_n))\mathbf{1}_{\{U_n < \theta g_n\}} - p(k, D(q, \underline{u}; g))\mathbf{1}_{\{U < \theta g\}}\right| \pi \leqslant \pi \leqslant \overline{\theta}^2/2 \quad F\text{-a.e.}$$

The second integral goes to zero by dominated convergence. Thus, $\Pi(q_n, \underline{u}_n; k_n, g_n) \to \Pi(q, \underline{u}; k, g)$. This completes the proof of Lemma 9. □

I now finally show the existence of the optimal mechanism in my model.

Theorem 4. *For every public option $(k, g) \in (0, 1] \times (0, \infty)$, there exists $(q^*, \underline{u}^*) \in \mathcal{K}$ such that $\Pi(q^*, \underline{u}^*; k, g) = \Pi^*(k, g)$.*

Proof. Firstly, to see that the supremum is finite, observe that for any $(q, \underline{u}) \in \mathcal{K}$, for any θ, $\pi(\theta) \leqslant \frac{\theta^2}{2}$, where $\pi(\theta) = \theta q(\theta) - \underline{u} - \int_{\underline{\theta}}^{\theta} q(s)ds - \frac{q(\theta)^2}{2}$. So, by 10, we have $0 \leqslant \Pi(q, \underline{u}; k, g) \leqslant \int \pi \, dF \leqslant \frac{\overline{\theta}^2}{2}$. Now, I choose allocations whose profits approach the supremum: $(q_n, \underline{u}_n) \in \mathcal{K}$ with

$$\Pi^*(k, g) - \frac{1}{n} \leqslant \Pi(q_n, \underline{u}_n; k, g) \leqslant \Pi^*(k, g).$$

In particular, $\Pi(q_n, \underline{u}_n; k, g) \to \Pi^*(k, g)$. Lemma 8 gives a subsequence n_j and $(q^*, \underline{u}^*) \in \mathcal{K}$ with

$$\underline{u}_{n_j} \longrightarrow \underline{u}^*, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_{n_j} - q^*|^2 \, dF \longrightarrow 0.$$

These satisfy 16 with constant policy sequence (k, g). (21) from Lemma 9 therefore gives us:

$$\Pi(q^*, \underline{u}^*; k, g) = \lim_{j \to \infty} \Pi(q_{n_j}, \underline{u}_{n_j}; k, g) = \Pi^*(k, g).$$

The first equality is profit continuity. The second follows from the allocations we chose, since $1/n_j \to 0$. This completes the existence proof. □

6.3 Basic Ironing Theorem

In this section, I state the standard ironing theorem that we will invoke later. It follows directly from the separable case of Toikka (2011), which builds on Myerson (1981), and therefore we do not provide a proof here. Let us first define some notation. Let $Z > 0$, and consider a weight function $w : [0, Z] \to \mathbb{R}$, bounded and measurable: there exists $L > 0$ such that $|w| \leqslant L$. Define the cumulative weight $W(z) := \int_0^z w(x)dx$. Let $G : [0, Z] \to \mathbb{R}$ be the greatest convex minorant of $W(.)$: G is the largest convex function such that $G(x) \leqslant W(x)$ for all $x \in [0, Z]$. Let $\overline{g} : [0, Z) \to \mathbb{R}$ denote the right derivative of $G(.)$, which exists as G is convex, and is non-decreasing and right continuous.

Theorem 5. *Consider the following problem:*

$$\begin{aligned} \max_{q:[0,Z]\to\mathbb{R}_+} \quad & \int_0^Z (w(z)q(z) - \frac{q(z)^2}{2})dz \\ \textit{subject to} \quad & q \textit{ non-decreasing} \end{aligned} \tag{P1}$$

The solution to P1 is $q^*(z) = \max\{0, \overline{g}(z)\}$ *up to a set of Lebesgue measure zero.*

We now verify that the hypotheses from Toikka (2011) hold and that Theorem 5 follows from the analysis of his separable case. Choose $B > L$. Replacing q by $\min\{q, B\}$ preserves non-negativity and monotonicity and weakly increases the objective, since

$$\frac{\partial}{\partial x}\left[w(z)x - \frac{x^2}{2}\right] = w(z) - x < 0 \qquad \text{for } x \geqslant B.$$

We may therefore restrict the allocations to $[0, B]$.

Substitute $s = z/Z$. The objective then equals Z times the following integral. Since $Z > 0$, it does not affect the maximizer.

$$\int_0^1 J(q(Zs), s)\, ds, \qquad J(x, s) = xw(Zs) - \frac{x^2}{2}.$$

This integrand is bounded and measurable and has separable form of Toikka (2011), with $a(x) = x$. The cumulative integral of $w(Zs)$ is $W(Zs)/Z$, and the greatest convex minorant of $W(Zs)/Z$ is $G(Zs)/Z$. Thus the ironed objective is

$$\overline{J}(x, s) = x\bar{g}(Zs) - \frac{x^2}{2}.$$

Its unique maximizer on $[0, B]$ is $\max\{0, \bar{g}(Zs)\}$, since the convex minorant slope satisfies $\bar{g} \leqslant L < B$. Corollary 3.9 of Toikka (2011) therefore gives optimizer in Theorem 5. Strict concavity of the objective function gives uniqueness almost everywhere.

6.3.1 Inf Sup Characterization of $\overline{g}(z)$

I now provide a useful characterization of $\overline{g}(z)$, which is the right derivative of G in Theorem 5. This will help establish some comparative statics that we will use later. For any $z \in [0, Z]$, we can characterize $\overline{g}(z)$ in terms of the inf sup of the averages of the raw weight w over intervals $(a, b) \subset [0, Z]$ such that $z \in (a, b)$. Since this characterization is in terms of the original weights, rather than the relatively opaque object that is the right derivative of the greatest convex minorant, it is more useful for proving comparative statics. This characterization is present in Kang & Watt (2024), but our proof is a bit different as we use infsup rather than maxmin to avoid degenerate intervals, so we supply it here for completeness.

Lemma 10. *For any* $(a, b) \subset [0, Z]$ *containing* z*, define* $s(a, b) := \frac{\int_a^b w(x)dx}{b-a} = \frac{W(b)-W(a)}{b-a}$*, the average slope of* $W(.)$ *over* (a, b)*. Then, for any* $z \in (0, Z)$*,*

$$\overline{g}(z) = \inf_{b>z} \sup_{a \leqslant z} s(a,b)$$

Proof. We do the proof in two steps. We first show that for any $b > z$, we have $\sup_{a \leqslant z} s(a,b) \geqslant \overline{g}(z)$. We then show that there is a b, or a sequence of b's, that forces the value of $\inf_{b>z} \sup_{a \leqslant z} s(a,b)$ to $\overline{g}(z)$.

Step 1: Fix z, and fix $b > z$. There are two possibilities: i) $G(z) = W(z)$, or ii) $G(z) < W(z)$. If $G(z) = W(z)$, then, let $a = z$. Then,

$$s(z,b) = \frac{W(b) - W(z)}{b - z} \geqslant \frac{G(b) - G(z)}{b - z} \geqslant \overline{g}(z)$$

Here, the first inequality is because $G(z) = W(z)$ and $G(x) \leqslant W(x)$ for all $x \in [0, Z]$, and the second inequality is by the decreasing slope property of convex functions: $\overline{g}(z) = \lim_{x \downarrow z} \frac{G(x)-G(z)}{x-z}$, so weakly lower than $\frac{G(b)-G(z)}{b-z}$ as $b > z$. So, in case i), for any b, we have that there exists an a such that $s(a,b) \geqslant \overline{g}(z)$. So, we have that $\inf_{b>z} \sup_{a \leqslant z} s(a,b) \geqslant \overline{g}$.

In case ii), when $G(z) < W(z)$, z lies in the interior of an open interval on which $G < w$, and on any such interval, G must be affine, so let (a^*, b^*) be the maximal affine interval containing z. It must be that $G(a^*) = W(a^*)$, and $G(b^*) = W(b^*)$. Now, choose $a = a^*$, and we have:

$$s(a^*, b) = \frac{W(b) - W(a^*)}{b - a^*} \geqslant \frac{G(b) - G(a^*)}{b - a^*}$$

So, by the same argument as earlier, we have that $\inf_{b>z} \sup_{a \leqslant z} s(a,b) \geqslant \overline{g}(z)$.

Step 2: We now show that $\inf_{b>z} \sup_{a \leqslant z} s(a,b) \leqslant \overline{g(z)}$ Again, there are two cases: 1) $G(z) = W(z)$, and ii) $G(z) < W(z)$. Consider ii) first. Again, let (a^*, b^*) be the maximal affine interval containing z. On this interval, the slope of G is constant and is equal to $\overline{g}(z)$. Consider the line $l(x) = G(b^*) + \overline{g}(z)(x - b^*)$ that passes through $(b^*, G(b^*)$ and supports $G(.)$ from below. To see why, suppose $x < b^*$. Then, since $\overline{g}(z)$ is the left derivative of G at b^*, we have $\frac{G(b^*)-G(x)}{b^*-x} \leqslant \overline{g}(z)$. If $x > b$, again, by increasing slopes, we have $\frac{G(x)-G(b^*)}{x-b^*} \geqslant \overline{g}(z)$. So, for any $a \leqslant x$, we have:

$$W(a) \geqslant G(a) \geqslant G(b^*) + \overline{g}(z)(a - b^*) = W(b^*) + \overline{g}(z)(a - b^*)$$

Rearranging, we have, $\frac{W(b^*)-W(a)}{b-a} \leqslant \overline{g}(z)$. This holds for every a, so holds for sup, therefore $\inf_{b>z} \sup_{a \leqslant z} s(a,b) \geqslant \overline{g}(z)$.

Now, consider i), so $G(z) = W(z)$. This implies that there must be an $\varepsilon > 0$, such that for $x \in (z, z + \varepsilon)$, we have $G(x) = W(x)$. This is by continuity of $G - W$: if no such epsilon

exists, this contradicts $G(z) = W(z)$. Pick a sequence $\{b_n\}$ such that $b_n \downarrow z$. For every b_n, consider the line $l_n(x) = G(b_n) + \overline{g}(b_n^-)(x - b_n)$, where $\overline{g}(b_n^-)$ is the left derivative of G at b_n, and by convexity of G, $\overline{g}(b_n^-) \leqslant \overline{g}(b_n)$, the right derivative. By the same arguments at earlier, line l_n, which passes through $(b_n, G(b_n))$, supports G from below. Again, by the same argument for part ii) above, we have $s(a, b_n) = \frac{W(b^*)-W(a)}{b-a} \leqslant \overline{g}(b_n)$. So, for every b, we have:

$$\sup_{a \leqslant z} s(a, b_n) \leqslant \overline{g}(b_n)$$

Since $\overline{g}$ is right continuous, we have that $\overline{g}(b_n) \to \overline{g}(z)$, so $\inf_{b>z} \sup_{a \leqslant z} s(a, b) \geqslant \overline{g}(z)$. Combining with the the inequality in **Step** 1, we have our result, that $\inf_{b>z} \sup_{a \leqslant z} s(a, b) = \overline{g}(z)$

□

We now record some straightforward comparative statics that follow directly from the above characterization. We also prove a useful lemma, where, at a contact point of G and W, we can divide the problem of ironing into two separate subproblems.

Lemma 11. ***Comparative Statics****: i) Consider weight functions $w_1(.), w_2(.)$, and let $s_1(.)$ and $s_2(.)$ be their corresponding slope functions. Suppose, for any $(a, b) \subset [0, Z]$, we have $s_1(a, b) \leqslant s_2(a, b)$. Then, $\overline{g}_1 \leqslant \overline{g}_2$. ii)* ***Domain contraction from left****: Consider the restriction of the cumulative weight function W to $[z', Z]$, where $z' > 0$. Let G' denote the greatest convex minorant of this restricted function, and $\overline{g}'$ its right derivative. Then, for any $z \in [z', Z]$, $\overline{g}'(z) \leqslant \overline{g}(z)$. iii)* ***Domain contraction from right****: Consider the restriction of the cumulative weight function W to $[0, Z']$, where $Z' < Z$. Let G' denote the greatest convex minorant of this restricted function, and $\overline{g}'$ its right derivative. Then, for any $z \in [0, Z']$, $\overline{g}'(z) \geqslant \overline{g}(z)$.*

Proof. i) follows directly from our characterization, ii) and iii) are adapted exactly from Kang and Watt (2024). The idea is that if you truncate the distribution from the left, you are taking the sup over a smaller set, this can only lower $\overline{g}$. For truncation from the right, you are taking inf over a smaller set, this can only increase $\overline{g}$. □

Lemma 12. ***Splitting Lemma****: Let w be the weight function from Lemma 10. Let W be the cumulative weight function, and G is the greatest convex minorant of W. Let z_c be such that $G(z_c) = W(z_c)$. Let $W_{[0,z_c]}$ be the restriction of W to $[0, z_c]$, and let G_L be its greatest convex minorant. Similarly, let $W_{[z_c,Z]}$ be be the restriction of W to $[z_c, Z]m$ and let G_U be greatest convex minorant. Then, $G_L = G_{[0,z_c]}$, and $G_U = G_{[z_c,Z]}$.*

Proof. Consider $W_{[0,z_c]}$. We want to show that $G_L = G_{[0,z_c]}$, where , $G_{[0,z_c]}$ is the restriction of G to $[0, z_c]$. Let $\overline{g}_L : [0, z_c)$ be the right derivative of G_L, and $\overline{g} : [0, z_c)$ be the right

derivative of G. By Lemma 10, we have: $\overline{g}(z) = \inf_{b>z} \sup_{a\leqslant z} s(a,b)$, and $\overline{g}_L(z) = \inf_{z_c\geqslant b>z} \sup_{a\leqslant z} s(a,b)$, for all $z \in (0, z_c)$. We will show:

$$\text{For any } a, b \text{ such that } a < z_c < b, \text{ we have that } s(a, z_c) \leqslant s(z_c, b) \tag{22}$$

We first argue that once we prove 22, we are done proving Lemma 12. To see this, recall that for any a, b in $[0, Z]$, such that $a < b$, $s(a,b) := \frac{W(b)-W(a)}{b-a}$, so, by definition of the function s, we have $s(a,b) = \frac{z_c-a}{b-a}s(a,z_c) + \frac{b-z_c}{b-a}s(z_c,b)$. therefore 22 would imply that $s(a,z_c) \leqslant s(a,b) \leqslant s(z_c,b)$. Since the a, b we fixed were arbitrary, therefore we would have, for any $a < z_c < b$, that $s(a,z_c) \leqslant s(a,b) \leqslant s(z_c,b)$. This implies that, for any $z \in (0, z_c)$, we have that $\inf_{b>z} \sup_{a\leqslant z} s(a,b) = \inf_{z_c\geqslant b>z} \sup_{a\leqslant z} s(a,b)$, or $\overline{g}(z) = \overline{g}_L(z)$, as allowing $b > z$ can only increase $s(a,b)$ for any fixed $a \leqslant z$, and we want to take infimum over $b > z$. Since $G_{[0,z_c]}$, and G_L are convex, therefore the right derivatives being equal implies that by fundamental theorem of calculus, G, and G_L agree almost everywhere on $[0, z_c]$.

We now prove 22. For any $a < z_c < b$, we have:

$$\begin{aligned} s(a,z_c) &= \frac{W(z_c) - W(a)}{z_c - a} \\ &\leqslant \frac{G(z_c) - G(a)}{z_c - a} \\ &\qquad \frac{G(b) - G(z_c)}{b - z_c} \\ &\leqslant \frac{W(b) - W(z_c)}{b - z_c} = s(z_c, b) \end{aligned}$$

where the first inequality holds as $W(z_c) = G(z_c)$, and $G \leqslant W$ on $[0, Z]$. The second inequality holds as G is convex, so it follows by the decreasing slope property of convex functions. Finally, the third inequality is again by $W(z_c) = G(z_c)$, and $G \leqslant W$. This completes the proof of 22, and therefore of Lemma 12.

□

I now state a lemma, without proof, that's a standard corollary of theorem Theorem 5. It modifies P1 to impose upper and lower bounds on the values that the function q can take.

Lemma 13. *Fix constants c_0 and c_1 such that $0 \leqslant c_0 \leqslant c_1 < \infty$. Consider the following problem:*

$$\max_{q:[0,Z]\to[c_0,c_1]} \int_0^Z (w(z)q(z) - \frac{q(z)^2}{2})dz \quad \text{subject to} \quad q \text{ non-decreasing} \tag{P1'}$$

The solution to P1', unique a.e. is $q^*(z) = \min\{c_1, \max\{c_0, \overline{g}(z)\}\}$.

6.4 Solution to Fixed Cutoff Problem

I now solve the fixed cutoff problem P-fixed cutoff in Section 3.4. Fix a cutoff $c \in (\underline{\theta}, \overline{\theta})$, with public demand $F(c) > k$. I want to solve for the optimal mechanism that *induces* this cutoff. Thus, I restrict attention to mechanisms (q, t) such that i) $q(.)$ is non decreasing, ii)$U(\underline{\theta}) \geqslant 0$, and iii) $U(\theta) \geqslant \theta g$ for all $\theta \geqslant c$, and $U(\theta) \leqslant \theta g$ for all $\theta \leqslant c$. Here, i) and ii) imply IC and IR, and iii) implies that when (q, t) is offered, $\theta \geqslant c$ weakly prefer to buy directly, and $\theta < c$ weakly prefer the public option. Combined with the fact that the monopolist can steer indifferent types however she likes, this makes the cutoff c feasible under (q, t).

I denote $U(\underline{\theta})$ by $\underline{u}$ from now on. I first argue that for iii) to hold, it is necessary and sufficient that type $U(c) = cg$, and $\underline{u} \leqslant \underline{\theta} g$, i.e. if the cutoff type is indifferent, and the lowest type weakly prefers the public option, then all $\theta \in (\underline{\theta}, c)$ weakly prefer the public option.

Claim 1. *Fix mechanism* (q, t) *such that* q *is non-decreasing, and* $U(\underline{\theta}) \geqslant 0$. *Then,* $U(\theta) \geqslant \theta g$ *for all* $\theta \geqslant c$, *and* $U(\theta) \leqslant \theta g$ *for all* $\theta \leqslant c$ *holds if and only if i)* $U(c) = cg$, *and ii)* $\underline{u} \leqslant \underline{\theta} g$

Proof. Define $h(\theta) := U(\theta) - \theta g$ for any θ. Thus $U(\theta) \geqslant \theta g$ if and only if $h(\theta) \geqslant 0$. *Necessity*: For $h(\underline{\theta})$, it is obvious. For $h(c) = 0$, since $U(.)$ is convex, so is h(.), therefore $h(.)$ is absolutely continuous and differentiable a.e., therefore, if $h(c) > 0$, there exist $\theta < c$ such that $h(\theta) > 0$. Similarly, if $h(c) < 0$, we cannot have $h(\theta) \geqslant 0$, for all $\theta > c$. *Sufficiency*: By convexity of $h(.)$, $h(c) = 0$, and $h(\underline{\theta}) \leqslant 0$, imply that $h(\theta) \leqslant 0$ for all $\theta < c$. For $\theta > c$, by the increasing slope property of convex functions, we have, for $\theta' < c < \theta$, that $\frac{h(c)-h(\theta')}{c-\theta'} \leqslant \frac{h(\theta)-h(c)}{\theta-c}$. Since $h(c) = 0$, and $h(\theta') \leqslant 0$, this implies that $h(\theta) \geqslant h(c) = 0$. □

Define $\beta_c := 1 - p_c(F(c))$, where recall that $p_c(F(c)) = \frac{k}{F(c)}$. We want to maximize the seller's profit over the feasible set, keeping the pattern that types $\theta < c$ come to the seller only after being rationed and $\theta > c$ come to seller directly, so each $\theta < c$ only transacts with the seller with probability β_c. Our optimization problem therefore is:

$$\begin{aligned} \max_{(q(.),\underline{u})} \quad & \beta_c \int_{\underline{\theta}}^{c} \pi(\theta) dF(\theta) + \int_{c}^{\overline{\theta}} \pi(\theta) dF(\theta) \\ \text{subject to} \quad & q(.) \text{ non decreasing} \\ & \underline{u} \in [0, \underline{\theta} g] \\ & \underline{u} + \int_{\underline{\theta}}^{c} q(\theta) d\theta = cg \end{aligned} \tag{P-fixed cutoff}$$

Here, $\pi(\theta) = t(\theta) - \frac{q(\theta)^2}{2}$, the monopolist's profit from type θ. $\underline{u} \geqslant 0$ is simply the IR constraint, $\underline{u} \leqslant \underline{\theta} g$ is simply $h\underline{u} \leqslant \underline{\theta} g$, and $\underline{u} + \int_{\underline{\theta}}^{c} q(\theta) d\theta = cg$ is $U(c) = cg$. Using $t(\theta) = \theta q(\theta) - U(\theta)$, we arrive at the virtual value representation of the objective function. Define

$$\varphi_c(\theta) := \theta - \frac{F(c) - F(\theta)}{f(\theta)}, \ \theta < c$$

and

$$\varphi(\theta) := \theta - \frac{1 - F(\theta)}{f(\theta)}, \ \theta \geqslant c$$

Recall that by Lemma 3, we can write the objective function in terms of virtual values:

$$J_c(q, \underline{u}) := \beta_c \int_{\underline{\theta}}^{c} (\varphi_c(\theta) q(\theta) - \frac{q(\theta)^2}{2}) dF(\theta) + \int_{\underline{c}}^{\overline{\theta}} (\varphi(\theta) q(\theta) - \frac{q(\theta)^2}{2}) dF(\theta) - \underline{u} \beta_c F(c) - cg(1 - F(c))$$

The above is obtained by using the envelope formula and changing the order of integration. We cannot apply ironing in this optimization problem in its current form, because: i) The quality terms in the upper and lower integrals are not integrated against the same measure (the lower integral has β_c multiplied with it), ii) In addition to the monotonicity constraint, we have the additional constraint $U(c) = cg$, which also means that we cannot separate the optimization problem and maximize over $q(.)$ and $\underline{u}$ separately. In particular, even though the coefficient of $\underline{u}$ is negative, we cannot automatically set it to zero. Of these, ii) is the more problematic, we can deal with i) through a change of variables, as we will see presently.

To deal with ii), we drop the constraint $U(c) = cg$, and use the multiplier $\lambda \in \mathbb{R}$ to reward or penalize its violation. Define:

$$\mathcal{L}_c(q, \underline{u}, \lambda) := J_c(q, \underline{u}) + \lambda(\underline{u} + \int_{\underline{\theta}}^{c} q(\theta) d\theta - cg) \tag{23}$$

Expanding and taking the $\lambda \int_{\underline{\theta}}^{c} f(\theta) d\theta$ term inside the lower integral,

$$\begin{aligned}\mathcal{L}_c(q,\underline{u},\lambda) = &\int_{\underline{\theta}}^{c}[(\varphi_c(\theta)+\frac{\lambda}{\beta_c f(\theta)})q(\theta)-\frac{q(\theta)^2}{2}]\beta_c f(\theta)d\theta \\ &+\int_{c}^{\overline{\theta}}(\varphi(\theta)q(\theta)-\frac{q(\theta)^2}{2})f(\theta)d\theta+\underline{u}(\lambda-\beta_c F(c))-cg(1-F(c)+\lambda)\end{aligned} \tag{24}$$

Consider the following maximization problem:

$$\begin{aligned}\max_{(q(.),\underline{u})} \quad & \mathcal{L}_c(q,\underline{u},\lambda) \\ \text{subject to} \quad & q(.) \text{ non decreasing} \\ & \underline{u}\in[0,\underline{\theta}g]\end{aligned} \tag{P-Lag}$$

Denote the solution to P-Lag by $(q_\lambda, \underline{u}(\lambda))$.

Lemma 14. ***Lagrangian Sufficiency****: Suppose, for multiplier λ, $(q_\lambda,\underline{u}(\lambda))$ solve P-Lag, and $(q_\lambda,\underline{u}(\lambda))$ is feasible for the original optimization problem, so $\underline{u}(\lambda)+\int_{\underline{\theta}}^{c}q_\lambda(\theta)d\theta = cg$. Then, $(q_\lambda,\underline{u}(\lambda))$ is also a solution to P-fixed cutoff, our original problem corresponding to the congested cutoff c.*

Proof. This is the scalar-equality specialization of the Luenberger sufficiency theorem. To see why its true, observe that for any $(q,\underline{u})$ that's feasible in the original optimization problem **??**, including $(q_\lambda,\underline{u}(\lambda))$, we have that the original objective $J_c(q,\underline{u}) = \mathcal{L}_c(q,\underline{u},\lambda)$, since the multiplier term is zero. Consider any $(q',\underline{u}')$ such that q' is non-decreasing, $\underline{u}'\in[0,\underline{\theta}g]$, and $\underline{u}'+\int_{\underline{\theta}}^{c}q'(s)ds = cg$, so $(q',\underline{u}')$ is part of the feasible set for P-fixed cutoff. Then,

$$J_c(q',\underline{u}') = \mathcal{L}_c(q',\underline{u}',\lambda) \leqslant \mathcal{L}_c(q_\lambda,\underline{u}(\lambda),\lambda) = J_c(q_\lambda,\underline{u}(\lambda))$$

where the inequality is by the fact that $(q_\lambda,\underline{u}(\lambda))$ maximizes $\mathcal{L}_c(q,\underline{u},\lambda)$.

□

Therefore, if we find a λ^* such that the solution $(q_{\lambda*},\underline{u}(\lambda^*))$ to the problem P-Lag corresponding to λ^* satisfies $\underline{u}(\lambda^*)+\int_{\underline{\theta}}^{c}q_{\lambda*}(s)ds = cg$, then , then this maximizer also solves the original problem. We solve the problem P-Lag for a fixed multiplier λ, and the show that for every c such that $F(c) > k$, there exists a $\lambda^*(c)$ such that the maximizer of $L_c(q,\underline{u},\lambda^*(c))$ satisfies $\underline{u}+\int_{\underline{\theta}}^{c}q(s)ds = cg$.

The problem P-Lag, of maximizing $L_c(q,\underline{u},\lambda)$ subject to the constraint that $(q,\underline{u})$ is such that q is non-decreasing, and $\underline{u}\in[0,\underline{\theta}g]$, is now separable, since we no longer have the $U(c) = cg$ constraint, and we can solve the $q(.)$ and $\underline{u}$ parts separately.

Lemma 15. *Fix a multiplier $\lambda \in \mathbb{R}$. Then, in the solution $(q_\lambda, \underline{u}(\lambda))$ to P-Lag, we have:*

$$\underline{u}(\lambda) = \begin{cases} 0 & \text{if } \lambda < \beta_c F(c) \\ \underline{\theta} g & \text{if } \lambda > \beta_c F(c) \\ \in [0, \underline{\theta} g] & \text{if } \lambda = \beta_c F(c) \end{cases}$$

Denote the $q(.)$ part of $\mathcal{L}_c(.)$ in 24 by $J_{c,\lambda}(q, \underline{u})$. We want to now maximize $J_{c,\lambda}(q, \underline{u})$, over non-decreasing q.

$$\varphi_\lambda(\theta) := \begin{cases} \varphi_c(\theta) + \frac{\lambda}{\beta_c F(c)} & \text{for } \theta < c \\ \varphi(\theta) & \text{for } \theta > c \end{cases}$$

Our problem is therefore:

$$\begin{aligned} \max_{q(.)} \quad & \beta_c \int_{\underline{\theta}}^{c} (\varphi_\lambda(\theta) q(\theta) - \frac{q(\theta)^2}{2}) f(\theta) d\theta + \int_{\underline{c}}^{\overline{\theta}} (\varphi_\lambda(\theta) q(\theta) - \frac{q(\theta)^2}{2}) f(\theta) d\theta \\ \text{subject to} \quad & q(.) \text{ non decreasing} \end{aligned} \tag{P2}$$

Th solution to P2 is clearly the q_λ in the solution $(q_\lambda, \underline{u}(\lambda))$ to P-Lag. We want to do the standard quantile transformation, followed by the generalized ironing of Toikka (2011). But here, since above and below the cutoff, types transact with the seller with different probabilities, therefore the correct quantile is not the cdf $F(.)$, but rather the cdf of the appropriate cutoff specific distribution of seller transactions. We define it below and do the transformation. First, let

$$\tau_c(\theta) := \begin{cases} \beta_c & \text{for } \theta < c \\ 1 & \text{for } \theta > c \end{cases}$$

So, for any θ, $\tau_c(\theta)$ is the probability with which she transacts with the seller. We now define a finite measure $\mu)c(.)$ on the type space. For any measurable set $A \subseteq \Theta$, $\mu_c(A) := \int_A \tau_c(\theta) dF(\theta)$. So, $\mu_c(A)$ is the expected mass of types in A for which the event of transacting with the seller is realized. We can therefore re-write the objective function in (6) as

$$\int_{\underline{\theta}}^{\overline{\theta}} (\varphi_\lambda(\theta) q(\theta) - \frac{q(\theta)^2}{2}) d\mu_c(\theta)$$

We now define the cdf of $\mu_c(.)$ as

$$z_c(\theta) := \mu_c([\underline{\theta}, \theta]) = \begin{cases} \beta_c F(\theta) & \text{for } \theta < c \\ F(\theta) - F(c) + \beta_c F(c) & \text{for } \theta > c \end{cases}$$

The domain of $z_c(.)$ is $[0, Z]$, where $Z = 1 - k$, which makes sense because $(1 - k)$ is the total mass of types that transacts with the seller. It is easy to see that $z(.)$ is a continuous and strictly increasing bijection, so it has an inverse $\theta : [0, Z] \to \Theta$. Define $w_\lambda(z) := \varphi_\lambda(\theta(z))$, and $\tilde{q}(z) := q(\theta(z))$ for all $z \in [0, Z]$. By definition, $dz(\theta) = d\mu_c(\theta)$, so, doing a change of variables to z, the optimization problem now is:

$$\begin{aligned} \max_{\tilde{q}(.)} \quad & \int_0^Z (w_\lambda(z)\tilde{q}(z) - \frac{\tilde{q}(z)^2}{2})dz \\ \text{subject to} \quad & \tilde{q}(.) \text{ non decreasing} \end{aligned} \tag{P3}$$

This is the standard generalized ironing problem. Let $W_\lambda(z) := \int_0^z w(x)dx$, and let $G_\lambda : [0, Z] \to \mathbb{R}$ be the greatest convex minorant of $W_\lambda(.)$ on $[0, Z]$, and let $\overline{g}_\lambda(z) := G'_\lambda(z^+)$, the right derivative of $G_\lambda(.)$ at z. Then, the following proposition states the solution to our fixed multiplier optimization problem:

Proposition 4. *For any $\lambda \in \mathbb{R}$, the unique solution to problem P2 is $q^*_\lambda(\theta) = [\overline{g}_\lambda(z(\theta))]_+$.*

Now, we come to the final step in determining the optimal mechanism corresponding to the fixed cutoff c. We can show, that there exists a unique multiplier $\lambda^*(c, g)$ such that $\underline{U}^*(\lambda^*(c, g))$ and $q^*_{\lambda^*(c,g)}(.)$ satisfy the indifference condition at c.

Proposition 5. *For every c such that $F(c) > k$, there exists a unique $\lambda^*(c, g)$, such that $\underline{u}(\lambda^*(c, g)) + \int_{\underline{\theta}}^c q_{\lambda^*(c,g)}(\theta)d\theta = cg$. Therefore, the solution to P-fixed cutoff is $(q_{\lambda^*(c,g)}, \underline{u}^*(\lambda^*(c, g)))$.*

Proof. ***Existence***: We first show the existence of a multiplier $\lambda^*(c, g)$ such that when we maximize $\mathcal{L}_c(q, \underline{u}, \lambda^*(c, g))$ over the space of all $q : \Theta \to \mathbb{R}_+$, and $\underline{u}$, such that q is non-decreasing, and $\underline{u} \in [0, \underline{\theta}g]$, then, the solution $(q_{\lambda^*(c,g)}, \underline{u}(\lambda^*(c, g)))$ is a solution to P-fixed cutoff.

For this, for any λ, define $T(\lambda) := \underline{u}(\lambda) + \int_{\underline{\theta}}^c q_\lambda(s)ds$, where $(q_\lambda, \underline{u}(\lambda))$ maximizes the Lagrangian $\mathcal{L}_c(q, \underline{u}, \lambda)$. Further, let $I(\lambda) := \int_{\underline{\theta}}^c q_\lambda(s)ds$ denote the integral part of $T(\lambda)$. Our plan for showing existence of the appropriate multiplier λ^* is the following: i) We first argue that T is non-decreasing in λ, and that I is Lipschitz continuous, and ii) $T(\lambda) \to \infty$ as $\lambda \to \infty$, and $T(\lambda) \to 0$ as $\lambda \to -\infty$. Together, i) and ii) would imply that there exists a λ^* such that $T(\lambda^*) = cg$. We already know from Lemma 14 that if the maximizer for the Lagrangian $\mathcal{L}_c(q, \underline{u}, \lambda)$ satisfies $U^*_\lambda(c) = cg$, where $U^*_\lambda(\theta) = \underline{u}(\lambda) + \int_{\underline{\theta}}^c q_\lambda(\theta)d\theta$ for any θ, then it also solves P-fixed cutoff.

- ***Step*** 1: We first show that T is non-decreasing in λ, i.e. for any λ', λ such that $\lambda' > \lambda$, we have $T(\lambda') \geqslant T(\lambda)$. Recall that $T(\lambda) := \underline{u}(\lambda) + \int_{\underline{\theta}}^{c} q_\lambda^*(s)ds$. From Lemma 15, $\underline{u}(\lambda)$ is non-decreasing in λ. We know that for any θ, $q_\lambda(\theta) = \overline{g}(z_c(\theta))$, where $z_c(\theta)$ is as defined earlier, and from Lemma 10, we have that for any $z \in (0, Z)$,

$$\overline{g}(z) = \inf_{b>z} \sup_{a \leqslant z} s_\lambda(a, b)$$

where $s_\lambda(a, b) := \frac{\int_a^b w_\lambda(x)dx}{b-a}$, and as defined in earlier in this section, $w_\lambda(x) = \varphi_\lambda(\theta(x))$ for any $x \in [0, Z]$. Recall that

$$\varphi_\lambda(\theta) := \begin{cases} \varphi_c(\theta) + \frac{\lambda}{\beta_c f(\theta)} & \text{for } \theta < c \\ \varphi(\theta) & \text{for } \theta > c \end{cases}$$

where $\varphi_c(\theta) = \theta - \frac{F(c)-F(\theta)}{f(\theta)}$. As λ increases, clearly, for every $\theta < c$, $\varphi_\lambda(\theta)$ increases strictly, and therefore so does $w_\lambda(z_c(\theta))$. Thus, by the comparative statics result in Lemma 11, we have that $\overline{g}_\lambda$ is non-decreasing in λ. By Proposition 4, q_λ^* is non-decreasing in λ as well, which completes the proof that T is non-decreasing in λ.

- ***Step*** 2: I now argue that I is Lipschitz continuous. For any λ' and λ, I will show:

$$|I(\lambda') - I(\lambda)| \leqslant \frac{c - \underline{\theta}}{\beta \underline{f}} |\lambda' - \lambda|$$

To show this, I first show that $|w_{\lambda'}(z) - w_\lambda(z)| \leqslant \frac{c-\underline{\theta}}{\beta \underline{f}} |\lambda' - \lambda|$ for any $z \in (0, z^*)$, where $z^* := z_c(c)$. Thus $z < z^*$ if and only if $\theta < c$, since z_c is strictly increasing and continuous in θ. Since $w_\lambda(z) = \varphi_c(\theta(z)) + \frac{\lambda}{\beta_c f(\theta(z))}$ for $z < z^*$, therefore $|w_{\lambda'}(z) - w_\lambda(z)| \leqslant \frac{|\lambda'-\lambda|}{\beta_c \underline{f}}$. By lemma 4, $q_\lambda(\theta) = [\overline{g}(z(\theta))]_+$ for all $\theta < c$ and $\overline{g}_\lambda(z) = \inf_{b>z} \sup_{a \leqslant z} \frac{\int_a^b w_\lambda(x)dx}{b-a}$, therefore $|w_{\lambda'}(z) - w_\lambda(z)| \leqslant \frac{c-\underline{\theta}}{\beta \underline{f}} |\lambda' - \lambda|$ establishes that I is Lipschitz.

- ***Step*** 3: I now show that $T(\lambda) \to \infty$ as $\lambda \to \infty$. By Lemma 15, for λ large enough, $\underline{u}(\lambda) = \underline{\theta} g$. So, to show that $T(\lambda) \to \infty$ as $\lambda \to \infty$, it is sufficient to argue that $I(\lambda) \to \infty$ as $\lambda \to \infty$. To show this, I fix $z_1 \in (0, z^*)$, and consider any $z \in (z_1, z^*)$. Then, using the inf sup characterization, we show that $\overline{g}(z) \geqslant C_0 + \frac{\lambda z_1}{\beta_c \overline{f} Z}$ for any $z \in (z_1, z^*)$, where $C_0 := \underline{\theta} - \frac{1}{\underline{f}}$.

To see this, consider the interval $(0, b)$ containing z, where $b > z$. Since $z > z_1$, thus $(0, b)$ contains the whole of $[0, z_1]$. Since $w_\lambda(x) = \varphi_c(\theta(x)) + \frac{\lambda}{\beta_c f(\theta(x))}$ for any $x < z^*$,

thus, the average of the λ part of w_λ over $(0, b)$ is at least $\frac{\lambda}{b}\int_0^{\min\{b,z^*\}} \frac{dx}{\beta_c f(x)} \geqslant \frac{\lambda z_1}{\beta_c \bar{f} Z}$ since both b and z^* are strictly greater than z_1, and $b \leqslant Z$. This analysis is for a fixed interval (a, b) containing z with $a = 0$ and $b > z_1$. But $\frac{\lambda z_1}{\beta_c \bar{f} Z}$ is a uniform lower bound, so it gets preserved when we take sup over a and inf over b.

This accounts for the λ part of $\overline{g}(z)$. The other part is $\varphi_c(\theta(z))$ which is at least C_0. Thus, for any $\theta > \theta(z_1)$ we have: $q_\lambda(\theta) \geqslant C_0 + \frac{\lambda z_1}{\beta_c \bar{f} Z}$, which goes to infinity as $\lambda \to \infty$. Since I had fixed $z_1 < z^*$, and $\theta(z^*) = c$, this completes the proof for $T(\lambda) \to \infty$ as $\lambda \to \infty$.

- ***Step*** 3: In this step, I show that $T(\lambda) \to 0$ as $\lambda \to -\infty$. Again, by Lemma 15, $\underline{u}(\lambda) = 0$ for λ small enough. Thus, it is enough to show that $I(\lambda) \to 0$ as $\lambda \to -\infty$. This is again a straightforward application of the inf sup characterization. For any $z \in (0, z^*)$ fix the interval (a, z^*) containing z, where $a \leqslant z^*$. Then, the λ part of the average of w_λ over (a, z^*) is at most $\frac{\lambda}{\beta_c \bar{f}}$ for negative λ, and this lower bound persists as we take supremum over $a \leqslant z$. The other part is $\varphi_c(\theta(z))$ which is bounded above by a constant. This gives us that for every $\theta < z$, as $\lambda \to -\infty$, we have that $q_\lambda(\theta) \to 0$. This completes the proof for $I(\lambda) \to 0$ as $\lambda \to -\infty$.

- ***Step*** 4: We are now ready to complete the existence proof. We want to prove the existence of a $\lambda^*(c, g)$ such that $\underline{u}(\lambda^*(c, g)) + I(\lambda^*(c, g)) = cg$. By Steps 1, 2 and 3 that I is Lipschitz continuous, non-decreasing, and that $I(\lambda) \to \infty$ as $\lambda \to \infty$, and $I(\lambda) \to 0$ as $\lambda \to -\infty$.

  Therefore, by the intermediate value theorem, for every $C \in (0, \infty)$, there exists a $\lambda(C)$ such that $I(\lambda(C)) = C$. This, along with the fact that $\underline{u}$ can take all values in $[0, \underline{\theta} g$ when $\lambda = \beta_c F(c)$ ensures that a supporting multiplier exists.

□

This completes the solution to the problem of finding the optimal mechanism for a fixed interior congested cutoff.

6.5 Pure Monopoly Benchmark

We describe the solution to the pure monopoly benchmark here. There is no public option, and the monopolist's problem is given by P-M. From Lemma 1, we also have the characterization of IC and IR mechanisms. The proof of Lemma 1 follows from standard arguments: Myerson (1981) shows that for any allocation rule $q : \Theta \to \mathbb{R}_+$, there exists a payment rule $t : \Theta \to \mathbb{R}$ such that (q, t) is IC if and only if q is non decreasing. From the envelope theorem in Milgrom

and Segal, we have that the payment rule in an IC mechanism is completely determined by the allocation rule up to the constant $U(\underline{\theta})$. In an optimal mechanism in the monopoly case, we must have $U(\underline{\theta}) = 0$, so the problem is to maximize $\int_{\underline{\theta}}^{\overline{\theta}}(\theta q(\theta) - \int_{\underline{\theta}}^{\theta} q(s)ds - \frac{q(\theta)^2}{2})dF(\theta)$ over the space of non decreasing functions $q : \Theta \to \mathbb{R}_+$. The objective function can be written in terms of virtual values. Define $\varphi(\theta) := \theta - \frac{1-F(\theta)}{f(\theta)}$. Then, the seller's problem is:

$$\begin{aligned} \max_{q(.):\Theta\to\mathbb{R}_+} \quad & \int_{\underline{\theta}}^{\overline{\theta}} (\varphi(\theta)q(\theta) - \frac{q(\theta)^2}{2})dF(\theta) \\ \text{subject to} \quad & q(.) \text{ non decreasing} \end{aligned} \tag{P-Mon}$$

Denote the solution to the above problem by q^M. If F is *regular*, i.e., if $\varphi(.)$ is non-decreasing, then there is no problem, we can do point-wise maximization and $q^M = [\varphi(\theta)]_+$, where $[x]_+ := \max\{0, x\}$ for any $x \in \mathbb{R}$. However, φ may not be non-decreasing, and we need to apply ironing. For doing so, we move to the quantile space: for each $x \in [0, 1]$, define $w(x) := \varphi(F^{-1}(x))$, and $W(x) := \int_0^x w(y)dy$. Let $G : [0, 1] \to \mathbb{R}$ be the greatest convex minorant of W: the largest convex function weakly lower than W, and let $g(x) := G'(x)_+$, the right derivative of G at x, which always exists, as G is convex. Define the *ironed virtual value* as $\overline{\varphi}(\theta) := g(F(\theta))$. The following result is due to Mussa Rosen:

Proposition 6. *The optimal mechanism in the pure monopoly case is (q^M, t^M), where $q^M(\theta) = [\overline{\varphi}(\theta)]_+$, and $t^M(\theta) = \theta q^M(\theta) - \int_{\underline{\theta}}^{\theta} q(s)ds$.*

We now state some standard properties of the monopoly optimal mechanism (q^M, t^M). Let $\pi^M := t^M(\theta) - \frac{q^M(\theta)^2}{2}$, the monopolist's profit from selling to θ in the optimal mechanism, and let $\Pi^M := \int_{\underline{\theta}}^{\overline{\theta}} \pi^M(\theta)dF(\theta)$ be her expected profit.

Lemma 16. ***Monopoly Facts**: The mechanism (q^M, t^M) satisfies: i) π^M is non-decreasing, ii) $q^M(\theta) < \theta$ for every $\theta \in (0, \overline{\theta})$ and iii) $\lim_{\theta\uparrow\overline{\theta}} \pi^M(\theta) > \Pi^M > 0$.*

Proof. i) and ii) are standard, and we will not be proving them here. We prove that $\lim_{\theta\uparrow\overline{\theta}} \pi^M(\theta) > \Pi^M$. Since π^M is non-decreasing, $\lim_{\theta\uparrow\overline{\theta}} \pi^M(\theta) =\geqslant \Pi^M > 0$: if this an equality, then π^M must be constant on $[\underline{\theta}, \overline{\theta}]$. We argue that this is not possible. If π^M is constant, then $\Pi^M = \pi^M(\underline{\theta})$. Since $\pi^M(\underline{\theta}) = \underline{\theta}q(\underline{\theta}) - \frac{q(\underline{\theta})^2}{2} \leqslant \frac{\underline{\theta}^2}{2}$, therefore $\Pi^M \leqslant \frac{\underline{\theta}^2}{2}$. But the monopolist can get an expected profit that's strictly higher than $\frac{\underline{\theta}^2}{2}$ by offering the following two contracts: $(\underline{\theta}, \underline{\theta}^2)$, and $(a, a^2 - a(\underline{\theta} - \underline{\theta}^2))$, where $a \in (\underline{\theta}, \overline{\theta})$.

If the monopolist offers these two contracts, type a is indifferent, $\theta > a$ strictly prefer $(a, a^2 - a(\underline{\theta} - \underline{\theta}^2))$, and $\theta < a$ strictly prefer $(\underline{\theta}, \underline{\theta}^2)$. The monopolist therefore makes profit $\pi(\theta) = \frac{\underline{\theta}^2}{2}$ from every $\theta \leqslant a$, and $\pi(\theta) = \frac{\underline{\theta}^2}{2} + \frac{(a-\theta)^2}{2}$ from every $\theta > a$. The expected profit over all types is clearly strictly higher than $\frac{\underline{\theta}^2}{2}$. This completes our proof for $\lim_{\theta\uparrow\overline{\theta}} \pi^M(\theta) =\geqslant \Pi^M > 0$. □

Let $U^M(\theta) = \theta q^M(\theta) - t^M(\theta)$, and let $\overline{q} := \lim_{\theta \uparrow \overline{\theta}} q^M(\theta)$. Since, by Lemma 16, we have that $q^M(\theta) \leqslant \theta$ for θ, therefore $\overline{q} \leqslant \overline{\theta}$. We now prove an additional lemma, that will be particularly useful in later proofs.

Note that since the monopoly profit $\Pi^M > 0$, we must have $q^M > 0$

Lemma 17. *Consider the function $\frac{U^M(\theta)}{\theta)}$ and let $g^* := \frac{U^M(\overline{\theta})}{\overline{\theta}}$. Then: i)$\frac{U^M(\theta)}{\theta)}$ is non-decreasing at every $\theta \in [\underline{\theta}, \overline{\theta}]$, and is strictly increasing at θ such that $q^M(\theta) > 0$, ii) $\overline{q} > g^*$, and iii) $0 < g^* < \frac{\overline{\theta}}{2}$.*

Proof. I first prove i). Since U^M is absolutely continuous, therefore $\frac{U^M(\theta)}{\theta}$ is also absolutely continuous. Thus, it is differentiable a.e., and $(\frac{U^M(\theta)}{\theta})' = \frac{\theta q(\theta) - U(\theta)}{\theta^2}$. But $\theta q(\theta) - U(\theta) = t(\theta)$, which is non-negative. So, $\frac{U^M(\theta)}{\theta}$ is non-decreasing at every $\theta \in [\underline{\theta}, \overline{\theta}]$. Furthermore, if $q^M(\theta) > 0$, then for $\pi^M(\theta) \geqslant 0$, we must have $t^M(\theta) \geqslant \frac{q(\theta)^2}{2} > 0$, which makes $(\frac{U^M(\theta)}{\theta})' > 0$.

I now prove ii). By Lemma 16 that $\Pi^M > 0$, so $q^M > 0$ for a non negative upper interval of θ. Thus, $U^M(\overline{\theta}) > 0$, and therefore $g^* > 0$. Taking limits in the profit formula $\pi(\theta) = \theta q^M(\theta) - U^M(\theta) - \frac{q^M(\theta)^2}{2}$ as $\theta \uparrow \overline{\theta}$, we have $\lim_{\theta \uparrow \overline{\theta}} \pi^M(\theta) = \overline{\theta}\overline{q} - \overline{\theta} g^* - \frac{\overline{\theta}^2}{2}$

$$\begin{aligned} \overline{\theta}(\overline{q} - g^*) &= \lim_{\theta \uparrow \overline{\theta}} \pi^M(\theta) + \frac{\overline{\theta}^2}{2} > 0, \\ \overline{\theta}\left(\frac{\overline{\theta}}{2} - g^*\right) &= \lim_{\theta \uparrow \overline{\theta}} \pi^M(\theta) + \frac{(\overline{\theta} - \overline{q})^2}{2} > 0. \end{aligned} \tag{25}$$

Since $\overline{\theta} > 0$,

$$\overline{q} > g^* > 0, \qquad g^* < \overline{\theta}/2.$$

This finishes the proof of both ii) and iii). □

6.6 Full Capacity Benchmark

In this section, I specify the solution to the full-capacity benchmark, where the public option has capacity $k = 1$, therefore all consumers who prefer the public option get it with probability one, and there is no residual demand that returns to buy from the seller.

This benchmark is a special case of the problem in Jullien (2000), who solves a general monopolistic screening problem with type-dependent outside options. Here, relative to the pure monopoly case, there is no distortion in the quality: those who buy from the seller, get quality q^M. The only change is the fewer consumers buy from the monopolist, and those who buy pay a strictly lower price compared to the pure monopoly case. I state the following result without proof:

Proposition 7. *Let $\overline{\varphi}(\theta)$ denote the ironed virtual value from the pure monopoly case. Let $c_1 := \inf\{\theta|\overline{\varphi}(\theta) \geqslant 2g\}$. Then, the optimal mechanism for the monopolist in the $k = 1$ case, that also maximizes the expected consumer surplus, is as follows: for every θ such that $\overline{\varphi}(\theta) < 2g$, the monopolist recommends the public option, which θ finds it optimal to take. For every θ such that $\overline{\varphi}(\theta) \geqslant 2g$, the monopolist offers $(q_1^*(\theta), t_1^*(\theta))$, where $q^*(\theta) = q^{MR}(\theta)$, and $t^*(\theta) = \theta q^{MR}(\theta) - c_1 g - \int_{c_1}^{\theta} q^{MR}(s)ds$.*

One point is worth emphasizing. In the above mechanism, all $\theta > c_1$ buy from the monopolist, and all $\theta < c_1$, take the public option. This is not the only optimal mechanism however. Another optimal mechanism is the following: let $c_1' := \inf\{\theta|\overline{\varphi}(\theta) > 2g\}$, to all θ such that $\overline{\varphi}(\theta) \leqslant 2g$, the monopolist recommends the outside option, to all θ such that $\overline{\varphi}(\theta) > 2g$, the monopolist offers $(q^{MR}(\theta), \theta q^{MR}(\theta) - c_1' g - \int_{c_1'}^{\theta} q^{MR}(s)ds)$. However, under this mechanism, consumers $\theta > a^*$, the pure monopoly threshold, are strictly worse off if the set $\{\overline{\varphi} = 2g\}$ is not a null set. So, I choose the optimal mechanism that weakly maximizes consumer surplus.

6.7 Proof of Proposition 2

Proposition 8. *Let (q^*, t^*) be an optimal mechanism, and let Π^* be the monopolist's expected profit from it. Then,*

- *If (q^*, t^*) induces **all** types to buy directly from the monopolist, then, $(q^*, t^*) = (q^M, t^M - \underline{\theta} g)$, and $\Pi^* = \Pi^M - \underline{\theta} g$*
- *If (q^*, t^*) is such that **no** type buys directly, then $(q^*, t^*) = (q^M, t^M)$, and $\Pi^* = (1-k)\Pi^M$.*

Proof. I first prove the first part by contradiction. Suppose there is an optimal mechanism (q, t) that induces **all** types to buy directly from the monopolist, and $(q, t) \neq (q^M, t^M - \underline{\theta} g)$. For every $\theta > 0$, we must have $U^*(\theta) \geqslant \theta g$, and $t(\theta) \geqslant \frac{q(\theta)^2}{2}$, because the monopolist's profit from selling to every type must be non-negative. These imply $\theta(q(\theta) - g) \geqslant \frac{1}{2} q(\theta)^2$. Quality cannot be zero because that would imply $\theta g > 0$, which cannot be for $\theta > 0$. So, we have:

$$q(\theta) - g \geqslant \frac{q(\theta)^2}{2\theta} > \frac{g^2}{2\theta} \geqslant \frac{g^2}{2\overline{\theta}}.$$

Therefore,

$$q(\theta) \geqslant g + \frac{g^2}{2\overline{\theta}} \quad \text{a.e.} \tag{26}$$

Since $q(\theta) > g$ for all $\theta > 0$, we have room to perturb it to be closer to the monopoly allocation q^M. Choose $\varepsilon > 0$, and define $q_\varepsilon = (1-\varepsilon)q + \varepsilon q^M$, where ε should be small enough that $q_\varepsilon \geqslant (1-\varepsilon)\left(g + \frac{g^2}{2\theta_h}\right) > g$ for all $\theta > 0$. Keep the utility $\underline{u} := \underline{\theta} q(\underline{\theta}) - t(\underline{\theta})$ for $\underline{\theta}$ same as in (q,t), and specify transfers using the envelope formula:

$$\begin{aligned} U_\varepsilon(\theta) - \theta g &= u_0 + \int_{\underline{\theta}}^{\theta} q_\varepsilon(s)\, ds - \theta g \\ &= (u_0 - \underline{\theta} g) + \int_{\underline{\theta}}^{\theta} [q_\varepsilon(s) - g]\, ds \geqslant 0. \end{aligned} \tag{27}$$

For every $\theta > \underline{\theta}$, the last inequality is strict by 26. So the perturbation ensures all types buy directly (the lowest types $\underline{\theta}$ could be zero but its a null set). I now argue that the profit from this perturbation is strictly higher than from the original mechanism (q,t), which would imply that (q,t) is not optimal. Since both the perturbed and original mechanisms induce everyone to buy directly, therefore, the profits are:

$$\Pi_\varepsilon = \int \left(\varphi q_\varepsilon - \frac{q_\varepsilon^2}{2}\right) dF - \underline{u}, \qquad \Pi = \int \left(\varphi q - \frac{q^2}{2}\right) dF - \underline{u}.$$

Subtracting the two terms, we have:

$$\Pi_\varepsilon - \Pi = \int \left[\varphi(q_\varepsilon - q) - \frac{1}{2}(q_\varepsilon^2 - q^2)\right] dF.$$

Now, after some algebra, we get $q_\varepsilon^2 - q^2 = 2\varepsilon q(q^{MR} - q) + \varepsilon^2 (q^{MR} - q)^2$. Substituting this in the profit difference, and after rearrangements, we get:

$$\Pi_\varepsilon - \Pi = \varepsilon \int \left[\varphi(q^{MR} - q) - \tfrac{1}{2}\left((q^{MR})^2 - q^2\right)\right] dF + \frac{\varepsilon(1-\varepsilon)}{2} \int (q^{MR} - q)^2\, dF \tag{28}$$

The first integral is nonnegative because q^M maximizes the monopoly objective and q is a feasible allocation in the pure monopoly problem. The second integral is strictly positive unless $q = q^M$ a.e. but $q \neq q^M$ on a set of positive measure. This is a contradiction, and proves the first part of Proposition 2.

I now prove the second part. I first argue that by offering (q^M, t^M), the monopolist can always earn a profit of at least $(1-k)\Pi^M$. This is because by Lemma 16, $\pi^M(\theta)$, the monopolist's profit from each type under (q^M, t^M), is non-decreasing in θ. Furthermore, by Lemma 17, the function $\frac{U^M(\theta)}{\theta}$ is non-decreasing at all $\theta > 0$, so when (q^M, t^M) is offered, if $U^M(\theta) \geqslant \theta g$, then $U^M(\theta') \geqslant \theta' g$ for all $\theta' > \theta$. Thus, by offering (q^M, t^M), either all θ buy only after failing to get the public option, or an upper interval buys directly, in which

case, profit is at least $(1-k)\Pi^M$ because higher types have higher profit margins for the monopolist. Moreover, any $(q,t) \neq (q^M, t^M)$ on a set of positive measure, that induces every type to only buy after failing, cannot be optimal, because the conditional distribution of types that buy is the same as in the monopoly case, and (q^M, t^M) is the unique solution to the monopoly problem. Thus, such a mechanism cannot be optimal. This proves the second part as well. □

6.8 Groundwork for Theorems 1 and 2

. I prove some results here that will be useful for proving Theorems 1 and 2. I first show the following reweighting lemma, which I will later use to prove two other useful lemmas.

Lemma 18 (Reweighting)**.** *Let $w : [0,1] \to \mathbb{R}$ be measurable with $|w| \leqslant L$. Fix $m, \beta \in (0,1)$, and let $r(x) = \beta$ for $x < m$, and $r(x) = 1$ for $x \geqslant m$. For $0 \leqslant a < b \leqslant 1$, define:*

$$s(a,b) = \frac{\int_a^b w(t)\,dt}{b-a}, \qquad s_r(a,b) = \frac{\int_a^b r(t)w(t)\,dt}{\int_a^b r(t)\,dt}.$$

Define:

$$\overline{g}(x) := \inf_{x<b\leqslant 1} \sup_{0\leqslant a\leqslant x} s(a,b), \qquad \overline{g}_r(x) := \inf_{x<b\leqslant 1} \sup_{0\leqslant a\leqslant x} s_r(a,b).$$

Then, for every $x \in (0,m)$,

$$\overline{g}_r(x) \leqslant \overline{g}(x), \qquad [\overline{g}_r(x)]_+ \leqslant [\overline{g}(x)]_+. \tag{29}$$

Proof. Fix $x \in (0,m)$, and set $T = \overline{g}(x)$. Choose a sequence $b_n \in (x,1]$ such that:

$$\sup_{0\leqslant a\leqslant x} s(a,b_n) \longrightarrow T. \tag{30}$$

First consider the case where infinitely many $b_n \leqslant m$. Choose the subsequence consisting of these points. Then for every $a \leqslant x$, $s_r(a,b_n) = s(a,b_n)$, since β in the definition of $s_r(a,b_n)$ cancels out, as the interval (a,b) lies entirely below m. Hence,

$$\overline{g}_r(x) \leqslant \sup_{0\leqslant a\leqslant x} s_r(a,b_n) = \sup_{0\leqslant a\leqslant x} s(a,b_n) \longrightarrow T.$$

If only finitely many $b_n \leqslant m$, then pass to a subsequence with $b_n > m$ and $b_n \to b^* \in [m,1]$.

For every $a \in [0, x]$, boundedness of w gives

$$|s(a, b_n) - s(a, b^*)| \leqslant \frac{\left|\int_a^{b_n} w - \int_a^{b^*} w\right|}{b_n - a} + \left|\int_a^{b^*} w\right| \frac{|b_n - b^*|}{(b_n - a)(b^* - a)}$$
$$\leqslant \frac{2L|b_n - b^*|}{m - x} \longrightarrow 0.$$

This bound is uniform over $a \in [0, x]$. Therefore 30 implies:

$$\sup_{0 \leqslant a \leqslant x} s(a, b^*) = T. \tag{31}$$

If $b^* = m$ then β again cancels out, and the desired inequality follows. I now consider the case of $b^* > m$. Since b can take value $m > x$, therefore $T \leqslant \sup_{0 \leqslant a \leqslant x} s(a, m)$. The function $a \mapsto s(a, m)$ is continuous, and $m - a \geqslant m - x > 0$. The extreme value theorem therefore gives a maximizer $a^* \in [0, x]$. By 31, $s(a^*, m) \geqslant T$, and $s(a^*, b^*) \leqslant T$. Therefore,

$$(b^* - m)[s(m, b^*) - T] = (b^* - a^*)[s(a^*, b^*) - T] - (m - a^*)[s(a^*, m) - T] \leqslant 0.$$

Since $b^* > m$, this proves $s(m, b^*) \leqslant T$. For every $a \in [0, x]$ splitting the integral at m gives

$$s_r(a, b^*) - T = \frac{\beta(b^* - a)[s(a, b^*) - T] + (1 - \beta)(b^* - m)[s(m, b^*) - T]}{\beta(m - a) + (b^* - m)} \leqslant 0$$

Both numerator terms are non-positive by 31 and $s(m, b^*) \leqslant T$. thus,

$$\overline{g}_r(x) \leqslant \sup_{0 \leqslant a \leqslant x} s_r(a, b^*) \leqslant T = \overline{g}(x)$$

This completes the proof. □

I now go to the P-fixed cutoff problem corresponding to cutoff c that satisfies $U^M(c) \leqslant cg$, where recall that for any θ, $U^M(\theta)$ is her utility from the mechanism (q^M, t^M). I show that at such a c, the unique supporting multiplier from Lemma 5 has a lower bound.

Lemma 19. *Fix $g > 0$ and $c \in (\underline{\theta}, \overline{\theta})$ with $F(c) > k$. If $U^{MR}(c) \leqslant cg$, then,*

$$\lambda(c, g) \geqslant -\beta_c[1 - F(c)]. \tag{32}$$

In particular, 32 holds whenever $g \geqslant g^*$.

Proof. Define $\lambda_c^0 := -\beta_c[1 - F(c)]$. I first compare the solution to the fixed multiplier problem at λ_c^0 with pure monopoly. I then show that a supporting multiplier smaller that λ_c^0 at c

would contradict uniqueness in Proposition 5. Recall that

$$\varphi_\lambda(\theta) := \begin{cases} \varphi_c(\theta) + \frac{\lambda}{\beta_c f(\theta)} & \text{for } \theta < c \\ \varphi(\theta) & \text{for } \theta > c \end{cases}$$

Thus, at λ_c^0, the raw virtual values are same as pure monopoly, as $1 - F(c) + \frac{\lambda_c^0}{\beta_c} = 0$, so $\varphi_{\lambda_c^0}(\theta) = \varphi(\theta)$ on both sides of c. Only the transaction probability below c is β_c. Apply Lemma 18 with $w(x) = \varphi(F^{-1}(x))$, $m = F(c)$ and $\beta = \beta_c$. This w is bounded because the θ is from a bounded type interval, and $f \geqslant \underline{f} > 0$. Proposition 4 and 29 give:

$$q_{\lambda_c^0}(\theta) \leqslant q^{MR}(\theta) \quad \text{for almost every } \theta < c, \qquad \int_{\underline{\theta}}^{c} q_{\lambda_c^0}(\theta)\, d\theta \leqslant U^{MR}(c). \tag{33}$$

Moreover, $\lambda_c^0 < 0 < \beta_c F(c)$, so Lemma 15 gives $\underline{u}(\lambda_c^0) = 0$. Now suppose the supporting multiplier at (c, g) is $\lambda(c, g) < \lambda_c^0$. Lemma 15 gives $\underline{u}(\lambda(c, g)) = 0$. Increasing the multiplier increases the score below c, and leaves it unchanged above c. Lemma 11(i) and Proposition 4 therefore imply $q_{\lambda^*(c,g)}(\theta) \leqslant q_{\lambda_c^0}(\theta)$ a.e.The constraint $U(c) = cg$, 33, and $\lambda(c, g) < \lambda_c^0$ give:

$$cg = \int_{\underline{\theta}}^{c} q_{\lambda(c,g)}(\theta)\, d\theta \leqslant \int_{\underline{\theta}}^{c} q_{\lambda_c^0}(\theta)\, d\theta \leqslant U^{MR}(c) \leqslant cg. \tag{34}$$

Every inequality in (34) is an equality. Hence $(q_{\lambda_c^0}, 0)$ both maximizes its Lagrangian and satisfies $U(c) = cg$. Lemma 14 makes it a fixed-cutoff optimum supported by λ_c^0. Uniqueness in Proposition 5 then gives $\lambda_c^0 = \lambda^*(c, g)$ which is a contradiction. Finally, by definition of g^*, we have $U^M(c) \leqslant cg^* \leqslant cg$ when $g \geqslant g^*$. □

The following lemma proves that there is a strict decline in the maximum profit attainable from a fixed cutoff c as g increases.

Lemma 20. *Fix $c \in (\underline{\theta}, \overline{\theta})$ such that $F(c) > k$. For $g_2 > g_1 \geqslant g^*$, we have:*

$$V(c, g_2) \leqslant V(c, g_1) - \frac{ck[1 - F(c)]}{F(c)}(g_2 - g_1) < V(c, g_1). \tag{35}$$

Proof. Here, $V(c, g)$ refers to the monopolist's profit from the solution to P-fixed cutoff, the fixed-cutoff problem corresponding to c. Let $\lambda_1 = \lambda(c, g_1)$. I solve the fixed multiplier problem with this λ at both g_1 and g_2. I show that its maximized Lagrangian equals the fixed-cutoff value $V(c, g_1)$ at g_1, and acts as an upper bound to $V(c, g_2)$. Equation (24) and

Lemma 15 imply

$$\begin{aligned}
&\max_{\substack{q\geqslant 0 \text{ non-decreasing} \\ 0\leqslant \underline{u}\leqslant \underline{\theta} g}} \mathcal{L}_c(q,\underline{u},\lambda_1;g) \\
&\quad = \max_{q\geqslant 0 \text{ non-decreasing}} \left\{ \beta_c \int_{\underline{\theta}}^{c} \left[\varphi_{\lambda_1}(\theta) q(\theta) - \frac{q(\theta)^2}{2} \right] dF(\theta) + \int_{c}^{\overline{\theta}} \left[\varphi(\theta) q(\theta) - \frac{q(\theta)^2}{2} \right] dF(\theta) \right\} \\
&\qquad + \underline{\theta} g [\lambda_1 - \beta_c F(c)]_+ - cg[1 - F(c) + \lambda_1]. \qquad (36)
\end{aligned}$$

The q-part of the maximization is identical at g_1, and g_1. Only the last two terms depend on g. For any $q, \underline{u}$ satisfying $U(c) = cg$, the multiplier term in the multiplier term in (23) is zero. Thus,

$$\begin{aligned}
V(c,g_2) &\leqslant \max_{\substack{q\geqslant 0 \text{ non-decreasing} \\ 0\leqslant \underline{u}\leqslant \underline{\theta} g_2}} \mathcal{L}_c(q,\underline{u},\lambda_1;g_2) \\
&= \max_{\substack{q\geqslant 0 \text{ non-decreasing} \\ 0\leqslant \underline{u}\leqslant \underline{\theta} g_1}} \mathcal{L}_c(q,\underline{u},\lambda_1;g_1) \\
&\quad + (g_2 - g_1) \left\{ \underline{\theta} [\lambda_1 - \beta_c F(c)]_+ - c[1 - F(c) + \lambda_1] \right\} \\
&= V(c,g_1) + (g_2 - g_1) \left\{ \underline{\theta} [\lambda_1 - \beta_c F(c)]_+ - c[1 - F(c) + \lambda_1] \right\}. \qquad (37)
\end{aligned}$$

The first equality is (36). The second equality uses the supporting multiplier from Proposition 5 and Lemma 14.

Now, let $\lambda_c^0 = -\beta_c[1 - F(c)]$. Lemma 19 gives $\lambda_1 \geqslant \lambda_c^0$. Since $\lambda_c^0 < \beta_c F(c)$, therefore $[\lambda_1 - \beta_c F(c)]_+ \leqslant \lambda_1 - \lambda_c^0$. Finally,

$$\begin{aligned}
\underline{\theta}[\lambda_1 - \beta_c F(c)]_+ - c[1 - F(c) + \lambda_1] &\leqslant -c[1 - F(c) + \lambda_c^0] - (c - \underline{\theta})(\lambda_1 - \lambda_c^0) \\
&\leqslant -c[1 - F(c)](1 - \beta_c) = -\frac{ck[1 - F(c)]}{F(c)}.
\end{aligned}$$

Substitution into (37) proves (35). The last coefficient is strictly negative because $c > 0$, $F(c) < 1$, and $k > 0$. □

I now come to the final lemma for this groundwork section. Consider again the fixed cutoff problem corresponding to cutoff c such that $F(c) > k$, and $U^M(c) \leqslant cg$. I show that the solution to the fixed cutoff problem satisfies $U(\theta) \geqslant U^M(\theta)$ for all $\theta < c$.

Lemma 21. *Fix $c \in (\underline{\theta}, \overline{\theta})$, and $g > 0$ such that $U^M(c) \leqslant cg$. Suppose $(q, \underline{u})$ solved the fixed-cutoff problem P-fixed cutoff at c, and let $U(\theta) = \underline{u} + \int_{\underline{\theta}}^{\theta} q(s)\, ds$. Then,*

$$U(\theta) \geqslant U^M(\theta) \qquad \text{for every } \theta \in [\underline{\theta}, c]. \qquad (38)$$

In particular, this holds whenever $g \geqslant g^*$.

Proof. Denote $\lambda(c, g)$ by λ for now. By Lemma 19, $1 - F(c) + \frac{\lambda}{\beta_c} \geqslant 0$. Thus by definition of φ_λ, we have:

$$\varphi_\lambda(\theta) \geqslant \varphi(\theta) \qquad (\theta < c). \tag{39}$$

I first compare qualities on any lower interval ending at a contact point. If ironing results in pooling across the cutoff c, then this comparison gives utility at one endpoint. The constraint $U(c) = cg$ gives the other endpoint. Convexity of U^M then proves the comparison between them.

Let $(a \in (\underline{\theta}, c]$ satisfy $G_\lambda(z_c(a)) = W_\lambda(z_c(a))$. By Lemma 12, we can iron $[\underline{\theta}, a]$ separately. On this interval, $z_c(\theta) = \beta_c F(\theta)$. Rescaling to $x = F(\theta)$ cancels β_c in the average over any interval. For $x \in (0, F(a))$, (39) and Lemma 11(i), and (iii) give: Lemma 11(i), (iii) give explicitly

$$\begin{aligned}
&\inf_{x<b\leqslant F(a)} \sup_{0\leqslant d\leqslant x} \frac{\int_d^b \varphi_\lambda(F^{-1}(t))\,dt}{b-d} \\
&\quad\geqslant \inf_{x<b\leqslant F(a)} \sup_{0\leqslant d\leqslant x} \frac{\int_d^b \varphi(F^{-1}(t))\,dt}{b-d} \\
&\quad\geqslant \inf_{x<b\leqslant 1} \sup_{0\leqslant d\leqslant x} \frac{\int_d^b \varphi(F^{-1}(t))\,dt}{b-d}.
\end{aligned}$$

The first inequality is due to the higher raw score; the second is because infimum over a smaller set is larger. Taking positive parts, Proposition 4 and the monopoly characterization imply

$$q(\theta) \geqslant q^M(\theta) \text{ a.e. on } [\underline{\theta}, a], \qquad U(\theta) - U^M(\theta) = \underline{u} + \int_{\underline{\theta}}^{\theta} (q - q^{MR})\,ds \geqslant 0 \quad (\theta \leqslant a).$$

If $z_c(c)$ is a contact point, take $a = c$; this proves the lemma. If not, i.e. $G_\lambda(z_c(c)) < W_\lambda(z_c(c))$, then by the fact that for a continuous function W, the greatest convex minorant G is affine on each component of $G_\lambda < W_\lambda$ with contact points at the endpoints. Let $z_c(a) < z_c(c)$ be the left endpoint of this component. On $[a, c]$, the allocation is constant, denote it by $Q > 0$. We then have $U(a) \geqslant U^{MR}(a)$; if $a = \underline{\theta}$, this instead follows directly from $\underline{u} \geqslant 0 = U^{MR}(\underline{\theta})$. Indifference at cutoff gives $U(c) = cg \geqslant U^{MR}(c)$. Since U is affine on $[a, c]$, and U^M is convex, theerfore $U - U^M$ is concave, and for every $\theta \in [a, c]$,

$$\begin{aligned}
U(\theta) - U^M(\theta) \geqslant{}& \frac{c-\theta}{c-a}[U(a) - U^M(a)] \\
&+ \frac{\theta - a}{c-a}[cg - U^M(c)] \geqslant 0.
\end{aligned}$$

This proves (38).

□

6.9 Proof of Theorem 1

I now prove Theorem 1 by dividing it into two parts. Let $g^* := \frac{U^M(\overline{\theta})}{\overline{\theta}}$, where , for any θ, $U^M(\theta)$ is utility of θ from (q^M, t^M), the solution to the pure monopoly problem P-M. I now prove the first part of Theorem 1.

Claim 2. *For every $k \in (0,1)$, there exists a public quality $\hat{g}(k) \in (g^*, \overline{\theta}/2)$, such that i) for every $g \geqslant \hat{g}(k)$, there exists an optimal mechanism whose outcome is zero direct demand: every type demands the public option and buys from the monopolist only upon failing to get it, ii) for every $g > \hat{g}(k)$, a mechanism with zero direct demand is uniquely optimal, iii) for every $g < \hat{g}(k)$, there is no optimal mechanism with zero direct demand.*

Proof. I do the proof in steps:

Step 1. For any $k \in (0,1)$, let $P(k)$ denote the set of public qualities g such that at (g,k), there exists an optimal mechanism that induces direct demand zero. I first show that $P(k)$ is non-empty. Then, the desired $\hat{g}(k)$, if it exists, must be $\inf P(k)$. To see why $P(k)$ is non-empty, I show that for any $g > \overline{\theta}/2$, there is no optimal mechanism where a positive mass of consumers buy directly. To see this, suppose (q,t) is optimal at $g > \overline{\theta}/2$. For θ to buy directly, $U(\theta) \geqslant \theta g$. Moreover, if θ buys directly, the monopolist makes positive profit from selling to θ, so $\pi(\theta) = t(\theta) - \frac{q(\theta)^2}{2} \geqslant 0$, which implies that $U(\theta) = \theta q(\theta) - t(\theta) \leqslant \frac{\theta^2}{2}$. Together, these imply $\theta g \leqslant U(\theta) \leqslant \theta^2/2$, which implies that $\theta \geqslant 2g$. So, if, in an optimal mechanism, θ buys directly, then $\theta \geqslant 2g$. At $g > \overline{\theta}/2$, there is no such type. So, in any optimal mechanism, direct demand must be zero.

Step 2. Let $\hat{g}(k) := \inf P(k)$. I will now argue that it has all the properties mentioned in Theorem 1. Fix k. I begin by showing that $\hat{g}(k) \geqslant g^*$. I first show the weak inequality. By Proposition 2, if an optimal mechanism induces zero direct demand, it must be (q^M, t^M). A corollary of Proposition 2 is that full public demand can *only* be optimal at g if (q^M, t^M) induces it, i.e. if $U^M(\theta) \leqslant \theta g$ for all θ, where $U^M(\theta) = \theta q^M(\theta) - t^M(\theta)$. At $g < g^*$, there exists $a < \overline{\theta}$ such that $U^M(\theta) > \theta g$ for all $\theta \in [a, \overline{\theta}]$. This is because the function $\frac{U^M(\theta)}{\theta}$ is non-decreasing on $[\underline{\theta}, \overline{\theta}]$ (see Lemma 17 in the Appendix). Since $g^* = \frac{U^M(\overline{\theta})}{\overline{\theta}}$, therefore at any $g < g^*$, $U(\overline{\theta}) > \overline{\theta} g$. By continuity of U^M, this strict inequality extends to a non empty interval below $\overline{\theta}$. Therefore, at any $g < g^*$, (q^M, t^M) does not induce full public demand. This implies that the outcome of an optimal mechanism cannot be that every type demands the public option.

I now prove the strict inequality, i.e. $\hat{g}(k) > g^*$. I begin with the following claim:

Claim 3. *If the monopolist offers (q^M, t^M) at g^*, the $U^M(\overline{\theta}) = \overline{\theta} g^*$, and $U^M(\theta) < \theta g^*$ for all $\theta \in (\underline{\theta}, \overline{\theta})$.*

Proof. $U^M(\overline{\theta}) = \overline{\theta} g^*$ follows from the definition of g^*. For $\theta \in (\underline{\theta}, \overline{\theta})$, from Lemma 17, $\frac{U^M(\theta)}{\theta}$ is non-decreasing at all θ, and strictly increasing at θ for which $q^M(\theta) > 0$. By Lemma 17, we also have that $q^M(\theta) > 0$ for θ close enough to $\overline{\theta}$. This gives $U^M(\theta) < \theta g^*$ for all $\theta < \overline{\theta}$. □

The proof plan for showing $\hat{g}(k) > g^*$ is as follows: by Claim 3, offering the monopoly menu (q^M, t^M) at g^* induces all types to prefer the public option, but $\overline{\theta}$ is indifferent, which means that if the monopolist "improves" the mechanism (q^M, t^M) a little, she can attract an upper interval of types $[c, \overline{\theta}]$ directly. We show that there exists an improvement such that if the monopolist offers it at g^*, a positive measure of types buy directly, and her profit is strictly higher than $(1-k)\pi^M$, which is what she gets by offering (q^M, t^M). Moreover, such an improvement, by continuity, can be constructed for $g > g^*$ close enough to g^*. So, there exists an $\varepsilon > 0$, such that for $g \in [g^*, g^* + \varepsilon)$, the mechanism (q^M, t^M) cannot be optimal. This would prove that the public demand one threshold, $\hat{g}(k) > g^*$, since by Proposition 2, if an optimal mechanism induces all types to prefer the public option, then it must be (q^M, t^M). I prove $\hat{g}(k) > g^*$ in steps:

- *Step* 1: *Scale up the menu* (q^M, t^M). For any $c \in (\underline{\theta}, \overline{\theta})$, consider scaling factor $\alpha_c := \frac{cg^*}{U^M(c)}$. I first argue that for c large enough, $\alpha_c > 1$. To see this, $\alpha_c - 1 = \frac{cg^* - U^M(c)}{U^M(\theta)}$. Since $q^M(c) > 0$ for c large enough, therefore $\frac{U^M(\theta)}{\theta}$ is strictly increasing on $(c, \overline{\theta})$, and therefore $\frac{U^M(c)}{c} < g^*$, which implies $\alpha_c > 1$.

  I now argue that for any $c \in (\underline{\theta}, \overline{\theta})$ that's close enough to $\overline{\theta}$, the scaled-up menu $(\alpha_c q^M, \alpha_c t^M)$ induces cutoff c. For any θ, let $U_c(\theta) := \theta \alpha_c q^M(\theta) - \alpha_c t^M(\theta) = \alpha_c U^M(\theta)$. To see why $(\alpha_c q^M, \alpha_c t^M)$ induces cutoff c, observe that $U_c(c) = cg^*$ by definition of α_c. For $\theta < c$, we have $U_c(\theta) \leqslant \theta g^*$ by the fact that $\frac{U^M(\theta)}{\theta}$ is non-decreasing, and $U_c(c) = \alpha_c U^M(c) = cg^*$. So, all types $\theta < c$ weakly prefer the public option. For $\theta > c$,

$$U_c(\theta) - \theta g^* = \int_c^\theta \alpha_c q^M(s) ds + U_c(c) - \theta g^*$$

$$= \int_c^\theta (\alpha_c q^M(s) - g^*) ds$$

  For c close enough to $\overline{\theta}$, we have $q^M(s) > g^*$ for all $s > c$. This is because from Lemma 17, we have that $\lim_{\theta \uparrow \overline{\theta}} q^M(\theta) > g^*$. Moreover, for c large enough, $\alpha_c > 1$, so $\int_c^\theta (\alpha_c q^M(s) - g^*) ds > 0$, therefore $U_c(\theta) > \theta g^*$ for all $\theta > c$.

- *Step* 2: In this step, I establish two useful facts.

  Fact 1: This fact describes the loss from scaling the monopoly optimal menu if every type buys from the monopolist. For a scalar α, let $\Pi^{\text{mon}}(\alpha)$ be the profit from the scaled menu $(\alpha q^M, \alpha t^M)$ if every type bought from the monopolist. Thus, $\Pi^{\text{mon}}(\alpha) = \alpha \int_{\underline{\theta}}^{\overline{\theta}} \varphi(\theta) q^M(\theta) dF(\theta) - \frac{\alpha^2}{2} \int_{\underline{\theta}}^{\overline{\theta}} (q^M(\theta))^2 dF(\theta)$. Here, $\varphi = \theta - \frac{(1-F(\theta))}{f(\theta)}$, the monopoly virtual value. Every $(\alpha q^M, \alpha t^M)$ is feasible under the monopoly problem, and $\alpha = 1$ maximizes the monopoly profit. So, the first order condition gives us:

$$0 = \frac{d}{d\alpha} \Pi^{\text{mon}}(\alpha) \Big|_{\alpha=1} = \int \phi q^{MR} \, dF - \int (q^{MR})^2 \, dF.$$

  It follows that:

$$\Pi^M = \int \varphi q^M \, dF - \frac{1}{2} \int (q^M)^2 \, dF = \frac{1}{2} \int (q^M)^2 \, dF.$$

  Substituting these equalities into the scaled profit $\Pi^{\text{mon}}(\alpha)$:

$$\begin{aligned} \Pi^{\text{mon}}(\alpha) &= 2\alpha \Pi^M - \alpha^2 \Pi^M \\ &= [\, 2\alpha - \alpha^2 \,] \Pi^M \\ &= [\, 1 - (\alpha - 1)^2 \,] \Pi^M. \end{aligned}$$

  Thus, the difference between the loss from scaling, relative to the monopoly menu is:

$$\Pi^M - \Pi^{\text{mon}}(\alpha) = (\alpha - 1)^2 \Pi^M \tag{40}$$

  Fact 2. This fact shows that if we use scaling factor α_c from ***Step*** 1, that induces cutoff c, then, the profit the monopolist gets from each $\theta > c$, i.e. the types that she retains by scaling up the monopoly menu, is strictly higher than her expected monopoly profit Π^M. Let $\pi_c(\theta) := \alpha_c t^M(\theta) - \frac{\alpha_c^2}{2} (q^M(\theta))^2$. Then $\pi_c(\theta)$ is the monopolist's profit from selling to θ under menu $(\alpha_c q^M, \alpha_c t^M)$. I now show that $\pi_c(\theta) > \Pi^M$ for all $\theta > c$.

$$\begin{aligned} \pi_c(\theta) &= \alpha_c t^M(\theta) - \frac{\alpha_c^2}{2} (q^M(\theta))^2 \\ &= \alpha_c \pi^M(\theta) - \frac{\alpha_c(\alpha_c - 1)}{2} (q^M(\theta))^2. \end{aligned}$$

For $\theta > c$, by Lemma 16, π^M is non-decreasing and $q^M(\theta) \leqslant \theta \leqslant \overline{\theta}$. So,

$$\begin{aligned}\pi_c(\theta) - \Pi^M &\geqslant \pi^M(\theta) - \Pi^M - \frac{\alpha_c(\alpha_c - 1)}{2}\overline{\theta}^2 \\ &\geqslant \pi^M(c) - \Pi^M - \frac{\alpha_c(\alpha_c - 1)}{2}\overline{\theta}^2\end{aligned}$$

Let $\pi_H := \lim_{\theta \uparrow \overline{\theta}} \pi^M(\theta)$.As $c \uparrow \theta_h$,

$$\pi^M(c) \longrightarrow \pi_H, \qquad \alpha_c \longrightarrow 1, \qquad \alpha_c - 1 \longrightarrow 0.$$

Therefore, we have:

$$\pi^M(c) - \Pi^M - \frac{\alpha_c(\alpha_c - 1)}{2}\overline{\theta}^2 \longrightarrow \pi_H - \Pi^M > 0.$$

where the last inequality is because $\pi_H > \Pi^M$ by Lemma 16. Thus, for c large enough, from the scaled menu $(\alpha_c q^M, \alpha_c t^M)$, the preceding analysis gives us a uniform lower bound for c sufficiently close to $\overline{\theta}$.

$$\pi_c(\theta) - \Pi^M \geqslant \frac{\pi_H - \Pi^M}{2} > 0 \ \textit{ for all } \ \theta > c \tag{41}$$

Thus, a small scaling change preserves the direct types' strictly above monopoly margin.

- *Step* 3: I now write the actual expected profit from the scaled menu $(\alpha_c q^M, \alpha_c t^M)$, and argue that for c sufficiently close to $\overline{\theta}$, it is strictly higher than $(1-k)\Pi^M$, the profit from full capitulation by Proposition 2.

$$\begin{aligned}\Pi_c &= \beta_c \int_{\underline{\theta}}^{c} \pi_c(\theta)\, dF(\theta) + \int_{c}^{\overline{\theta}} \pi_c(\theta)\, dF(\theta) \\ &= \beta_c \int_{\underline{\theta}}^{\overline{\theta}} \pi_c(\theta)\, dF(\theta) + (1 - \beta_c) \int_{c}^{\overline{\theta}} \pi_c(\theta)\, dF(\theta) \\ &= \beta_c \Pi^{\text{mon}}(\alpha_c) + \frac{k}{m} \int_{c}^{\overline{\theta}} \pi_c(\theta)\, dF(\theta).\end{aligned}$$

Substituting 40 and subtracting $(1-k)\Pi^M$:

$$\begin{aligned}\Pi_c - (1-k)\Pi^M &= \beta_c[1-(\alpha_c-1)^2]\Pi^M + \frac{k}{F(c)}\int_c^{\overline{\theta}} \pi_c \, dF - (1-k)\Pi^M \\ &= -\beta_C(\alpha_c-1)^2\Pi^M - \frac{k(1-F(c))}{F}(c)\Pi^M + \frac{k}{F(c)}\int_c^{\overline{\theta}} \pi_c \, dF \\ &= -\beta_c(\alpha_c-1)^2\Pi^M + \frac{k}{F(c)}\int_c^{\overline{\theta}} [\pi_c(\theta) - \Pi^M] \, dF(\theta).\end{aligned}$$

The first term is the cost of altering the monopoly mechanism. The second is the gain from retaining an higher than monopoly profit interval as direct buyers. To compare the loss and gain, divide both by $\theta_h - c > 0$. By $\alpha_c \to 1$ as $c \to \overline{\theta}$ and $0 < \beta_c < 1$,

$$\begin{aligned}0 \leqslant \frac{\beta_c(\alpha_c-1)^2\Pi^M}{\overline{\theta}-c} &\leqslant \frac{\Pi^M}{\overline{\theta}-c}\left[\frac{2(\overline{\theta}c)}{g^*}\right]^2 \\ &= \frac{4\Pi^M}{g^{*2}}(\overline{\theta}-c) \longrightarrow 0.\end{aligned}$$

The profit loss therefore vanishes. In contrast, 41 gives:

$$\begin{aligned}\frac{k}{F(c)(\overline{\theta}-c)}\int_c^{\overline{\theta}} [\pi_c(\theta) - \Pi^M] \, dF(\theta) &\geqslant \frac{k}{F(c)}\frac{1-F(c)}{\overline{\theta}-c}\frac{\pi_H - \Pi^M}{2} \\ &\geqslant \frac{k\underline{f}(\pi_H - \Pi^M)}{2} > 0.\end{aligned}$$

This is because $1/F(c) \geqslant 1$ and

$$1 - F(c) = \int_c^{\overline{\theta}} f(\theta) \, d\theta \geqslant \underline{f}(\overline{\theta} - c).$$

The gain is therefore bounded below by a strictly positive number independent of c. Therefore for c large enough, gain is strictly higher than loss. So, at g^*, I have constructed a menu that does strictly better than (q^M, t^M). So, zero direct demand cannot be optimal at g^*.

- *Step* 4: I now argue that the gain from scaling up a menu persists at $g > g^*$ as long as g is close enough to g^*. Keep the cutoff that worked for g^* fixed. We only need to change the scaling. For any $g > 0$, use

$$\alpha_c(g) = \frac{cg}{U^M(c)}, \qquad q_{c,g} = \alpha_c(g)q^M, \qquad U_{c,g} = \alpha_c(g)U^M.$$

Then $U_{c,g}(c) = cg$. Below c, similar arguments as for g^* give us $U_{c,g}(\theta) \leqslant \theta g$. For $\theta > c$, the earlier choices give $cq^M(\theta) > cg^* > U^M(c)$. Hence, for $\theta > c$,

$$\begin{aligned} U_{c,g}(\theta) - \theta g &= \int_c^\theta \left[\alpha_c(g) q^M(s) - g\right] ds \\ &= g \int_c^\theta \left[\frac{cq^M(s)}{U^M(c)} - 1\right] ds > 0. \end{aligned}$$

The profit from this scaling is:

$$\begin{aligned} \Pi_c(g) = \alpha_c(g) &\left[\beta_c \int_{\underline{\theta}}^c t^M \, dF + \int_c^{\overline{\theta}} t^M \, dF\right] \\ &- \frac{\alpha_c(g)^2}{2} \left[\beta_c \int_{\underline{\theta}}^c (q^M)^2 \, dF + \int_c^{\overline{\theta}} (q^M)^2 \, dF\right]. \end{aligned}$$

$\Pi_c(g)$ is continuous in g and we already proved $\Pi_c(g^*) > (1-k)\Pi^M$. By continuity, there exists $\varepsilon > 0$ such that

$$\Pi_c(g) > (1-k)\Pi^M \qquad (g_{\text{cap}} \leqslant g < g^* + \varepsilon).$$

This completes the proof that $\hat{g}(k) \geqslant g^* + \varepsilon > g^*$.

Step 3: I now argue that the optimality of zero direct demand is absorbing:

Lemma 22. *Fix $k \in (0,1)$. If there is an optimal mechanism at g that induces zero direct demand, then, at every $g' > g$, the unique optimal mechanism is (q^M, t^M), and it induces zero public demand.*

Proof. Suppose at g, there exists an optimal mechanism where direct demand is zero, and all types demand the public option. Then, by proposition 2, this mechanism must be (q^M, t^M), and its profit must be $(1-k)\Pi^M$. As I argued earlier, since zero public demand is optimal, it must be that $g \geqslant g^*$. Moreover, at any $g > g^*$, the profit from offering (q^M, t^M) is $(1-k)\Pi^M$, since $\frac{U^M(\theta)}{\theta}$ is non-decreasing, and at $g > g^*$, we have $U^M(\overline{\theta}) < \overline{\theta} g$. Thus, the profit from mechanism (q^M, t^M) remains the same at every $g' \geqslant g^*$.

I now prove the following: If, at $g \geqslant g^*$, there is no mechanism with positive direct demand that is strictly better than (q^M, t^M), then such a mechanism cannot be weakly better than (q^M, t^M) at $g' > g$. To show this, I show the following: fix any $c < \overline{\theta}$, and g, g' such that $g^* \leqslant g < g'$. Suppose there is an optimal mechanism (q', t') at g' such that it induces cutoff c, so $\theta > c$ buy directly, and its profit is Π'_c. Then, at g, there is a mechanism (q, t) that induces cutoff c, and has expected profit $\Pi_c > \Pi'_c$. We need to consider three cases:

- Suppose $c = \underline{\theta}$, so at g', there is an optimal mechanism (q', t') such that all types buy directly from the monopolist. Then, by Proposition 2, $(q', t') = (q^M, t^M - \underline{\theta} g')$. Since this mechanism is optimal at g', therefore the profit from this mechanism, $\Pi'_{\underline{\theta}} = \Pi^M - \underline{\theta} g \geqslant (1-k)\Pi^M$. Now consider mechanism $q^M, t^M - \underline{\theta} g)$ at g. This mechanism induces all types to buy directly at g, and has expected profit $\Pi_{\underline{\theta}} = \Pi^M - \underline{\theta} g > (1-k)\Pi^M$, contradicting the optimality of zero direct demand at g.

- If the optimal mechanism (q', t') at g' induces $c \in (\underline{\theta}, \overline{\theta})$ such that public demand $F(c') < k$, then it must be that c is indifference, i.e. $U(\theta) = cq'(c) - t'(c) = cg$. If we modify (q', t') such that we keep q same, and increase every transfer by $(g' - g)c$, then, at g, this modified mechanism again gives something strictly higher than $(1 - k)\Pi^M$. If c is such that $F(c) > k$, then Lemma 20 applies.

□

Step 4: I now prove that $\hat{g}(k) < \frac{\overline{\theta}}{2}$ for any $k \in (0, 1)$. Fix k. In the main text, I already argued that $\hat{g}(k) \leqslant \frac{\overline{\theta}}{2}$. I now argue that $\hat{g}(k) \neq \overline{\theta}/2$. I will prove this by contradiction. The idea is that if $\hat{g}(k) = \overline{\theta}/2$, then we can construct a sequence g_n, which approaches $\overline{\theta}/2$ from below. Since the optimality of zero direct demand is absorbing, therefore, at any $g_n < \overline{\theta}/2$, in any optimal mechanism, a positive mass of consumers must buy directly. However, as g_n approaches $\overline{\theta}/2$, the mass of direct buyers must go to zero, because for any buyer that buys directly, i) $U(\theta) \geqslant \theta g$, and ii) $\pi(\theta) \geqslant 0 \implies t(\theta) \geqslant \frac{q(\theta)^2}{2}$. Together, i) and ii) imply that $\theta \geqslant 2g$, so only a subset of $[2g_n, \overline{\theta}]$ can buy directly. Therefore, the profit earned from direct buyers goes to zero, but the loss from attracting them directly is bounded away from zero, as lower types also benefit from it. This results in a strictly lower profit than $(1 - k)\Pi$, which can be achieved by offering (q^M, t^M). This achieves the desired contradiction.

I now do the proof by contradiction. Suppose $\hat{g}(k) = \overline{\theta}/2$. Then, we can choose the following sequence:

$$g_n := \frac{\overline{\theta}}{2} - \frac{\overline{\theta} - \underline{\theta}}{4n}, \qquad n = 1, 2, \ldots.$$

For every n, we have $\underline{\theta}/2 < g_n < \overline{\theta}/2$, and $g_n \uparrow \overline{\theta}/2$. As discussed above, zero direct demand cannot be optimal at any g_n. Every consumer buying directly cannot be optimal either: as I argued in the preceding paragraph, only types in $[2g_n, \overline{\theta}]$ can buy directly in an optimal mechanism. Existence of an optimal mechanism implies that for each g_n, there is an interior

cutoff $c_n \in [2g_n, \overline{\theta}/2)$ corresponding to an optimal mechanism.[15] Thus, we have :

$$0 < \theta_h - c_n \leqslant \theta_h - 2g_n = \frac{\theta_h - \theta_l}{2n} \longrightarrow 0.$$

Thus, $c_n \longrightarrow \theta_h$, and $F(c_n) \longrightarrow F(\overline{\theta}) = 1$, as F is continuous. Because $k < 1$, for all sufficiently large n we have $F(c_n) > k$: so, at all these optimal mechanisms, the demand for the public option exceeds capacity. Define $\beta_n := 1 - \frac{k}{F(c_n)} > 0$, the rejection probability at my chosen optimal mechanism at g_n, with cutoff c_n. As $F(c_n) \longrightarrow 1$, therefore $\beta_n \longrightarrow 1 - k > 0$.

Let $(q_n, \underline{u}_n)$, be the optimal mechanism that induces c_n at g_n, where $\underline{u}_n$ denote the utility of the lowest type. For any θ,

$$U_n(\theta) = \underline{u}_n + \int_{\underline{\theta}}^{\theta} q_n(s)\, ds, \qquad \underline{u}_n \geqslant 0.$$

Let the profit from selling to type θ, and the profit if every type buys from the monopolist directly under $(q_n, \underline{u}_n)$ respectively be denoted by:

$$\pi_n(\theta) := \theta q_n(\theta) - U_n(\theta) - \frac{q_n(\theta)^2}{2}, \qquad \Pi_n^{\text{mon}} := \int \pi_n(\theta)\, dF(\theta).$$

Since (q^M, t^M) is the unique maximizer to the monopoly problem, therefore, $\Pi_n^{\text{mon}} \leqslant \Pi^M$. Since every type faces the monopolist with positive probability here, therefore, $\pi_n(\theta) \geqslant 0$ for all θ. Moreover,

$$\pi_n(\theta) = \frac{\theta^2}{2} - \frac{(q_n(\theta) - \theta)^2}{2} - U_n(\theta) \leqslant \frac{\theta^2}{2} \leqslant \frac{\overline{\theta}^2}{2}.$$

$\pi_n \geqslant 0$, and $U_n \geqslant 0$ also imply:

$$\frac{q_n(\theta)^2}{2} \leqslant \theta q_n(\theta), \qquad 0 \leqslant q_n(\theta) \leqslant 2\theta \leqslant 2\overline{\theta} \quad \text{a.e.}$$

So, these qualities are bounded. The monopolist's profit at g_n is

$$\begin{aligned} \Pi_n &= \beta_n \int_{\theta_l}^{c_n} \pi_n\, dF + \int_{c_n}^{\overline{\theta}} \pi_n\, dF \\ &= \beta_n \Pi_n^{\text{mon}} + (1 - \beta_n) \int_{c_n}^{\theta_h} \pi_n\, dF \\ &= \beta_n \Pi_n^{\text{mon}} + \frac{k}{F(c_n)} \int_{c_n}^{\overline{\theta}} \pi_n\, dF. \end{aligned} \tag{42}$$

[15] At any g_n, every optimal mechanism must correspond to an interior cutoff $c_n \in [g_n, \overline{\theta}/2)$. I simply choose one of these cutoffs.

By offering (q^M, t^M), the monopolist can get a profit of at least $(1-k)\Pi^M$. This is because from 16, we have that π^M is non-decreasing, and when (q^M, t^M) is offered, either every type only buys after failing to get the public option, in which case, the profit is exactly $(1-k)\Pi^M$, or an upper interval buys directly, which is a more favorable composition of buyers for the monopolist, and the profit is at least $(1-k)\Pi^M$. Therefore, $\Pi_n \geqslant (1-k)\Pi^M$. Also, $\beta_n \leqslant 1-k$ and $\Pi^M \geqslant 0$. Combining all these with 42, we have:

$$\begin{aligned} 0 \leqslant \beta_n(\Pi^M - \Pi_n^{\text{mon}}) &\leqslant (1-k)\Pi^M - \beta_n \Pi_n^{\text{mon}} \\ &\leqslant \Pi_n - \beta_n \Pi_n^{\text{mon}} \\ &= \frac{k}{F(c_n)} \int_{c_n}^{\overline{\theta}} \pi_n \, dF \\ &\leqslant \frac{k\overline{\theta}^2}{2F(c_n)}(1 - F(c_n)). \end{aligned}$$

Divide by $\beta_n = (F(c_n) - k)/F(c_n) > 0$:

$$0 \leqslant \Pi^M - \Pi_n^{\text{mon}} \leqslant \frac{k\overline{\theta}^2(1 - F(c_n))}{2(F(c_n) - k)} \longrightarrow 0. \tag{43}$$

The limit follows from $F(c_n) \to 1$ and $1 - k > 0$. Thus if the group of direct buyers is vanishing, then Π_n^{mon}, the hypothetical profit from the optimal mechanism $(q_n, \underline{u}_n$ at g_n, if every type bought from the monopolist (as if this were the pure monopoly case), must approach Π^M. This is because if Π_n^{mon} is strictly lower than Π^M, then this loss relative to monopoly cannot be made up by the vanishing group of direct buyers. I now show utility convergence from the profit convergence in 43. By the virtual value representation of the pure monopoly profits, we have:

$$\Pi^{\text{mon}}(q, \underline{u}) = \int \left[\varphi q - \frac{q^2}{2}\right] dF - \underline{u}, \qquad \varphi = \theta - \frac{1-F}{f} \tag{44}$$

For $0 \leqslant t \leqslant 1$, the quality $q^M + t(q_n - q^M)$ is nonnegative and non-decreasing. It is therefore feasible for the pure monopoly problem $\underline{u}$, the utility of the lowest type, being zero. Expanding 44, optimality of q^M for the pure monopoly case gives:

$$0 \geqslant t \int (\varphi - q^M)(q_n - q^M)\, dF - \frac{t^2}{2} \int (q_n - q^M)^2 \, dF.$$

For $t > 0$, dividing by t and taking $t \downarrow 0$, we get $\int(\varphi - q^M)(q_n - q^M)\, dF \leqslant 0$. Now, I once again write the difference between the hypothetical pure monopoly profit at $(q_n, \underline{u}_n)$, the

optimal mechanism at g_n, and Π^M, the profit from the optimal mechanism (q^M, t^M) at the pure monopoly benchmark:

$$\begin{aligned}\Pi^M - \Pi_n^{\text{mon}} &= \underline{u}_n - \int(\varphi - q^{MR})(q_n - q^M)\,dF + \frac{1}{2}\int(q_n - q^{MR})^2\,dF \\ &\geqslant \underline{u}_n + \frac{1}{2}\int(q_n - q^M)^2\,dF.\end{aligned} \tag{45}$$

Since q_n and q^M are bounded, bounded density makes φ bounded. By 43, and 45, we have:

$$\underline{u}_n \longrightarrow 0, \qquad \int(q_n - q^M)^2\,dF \longrightarrow 0.$$

Now, from the envelope formulas, we get:

$$\begin{aligned}|U_n(\overline{\theta}) - U^M(\overline{\theta})| &\leqslant \underline{u}_n + \left|\int_{\underline{\theta}}^{\overline{\theta}}(q_n - q^M)\,dt\right| \\ &\leqslant \underline{u}_n + \left(\int(q_n - q^M)^2\,dF\right)^{1/2}\left(\int_{\theta_l}^{\theta_h}\frac{dt}{f(t)}\right)^{1/2} \longrightarrow 0.\end{aligned}$$

The second inequality follows from Cauchy Schwarz. Thus, I have shown:

$$U_n(\overline{\theta}) \longrightarrow U^M(\overline{\theta}) \tag{46}$$

I now show the final step to obtain the desired contradiction. Recall that we started out with supposing that $\hat{g}(k) = \overline{\theta}/2$, which allowed us to choose the sequence $g_n \uparrow \hat{g}(k)$. I then showed that given the choice of g_n, any optimal mechanism corresponding to g_n must induce a cutoff $c_n \in [2g_n, \overline{\theta}/2)$, so $c_n \longrightarrow \overline{\theta}$, and $F(c_n)$, the public demand in this optimal mechanism, goes to one. I then use this to show 46. I now derive the contradiction using 46, by showing that the top type $\overline{\theta}$ requires a utility of strictly more than this limit $U^M(\overline{\theta})$ in every optimal mechanism $(q_n\underline{u}_n)$. We know that at every g_n, types above c_n but directly under $(q_n\underline{u}_n)$, so $U_n(\theta) \geqslant \theta g_n$ for every $\theta > c_n$. Since U_n is continuous, taking direct types approaching $\overline{\theta}$ gives $U_n(\overline{\theta}) \geqslant \overline{\theta} g_n$. Combining with 46, we have:

$$U^M(\overline{\theta}) = \lim_{n\to\infty} U_n(\overline{\theta}) \geqslant \lim_{n\to\infty} \overline{\theta} g_n = \frac{\overline{\theta}^2}{2}.$$

But the pure monopoly benchmark satisfies $U^M(\overline{\theta}) = \overline{\theta} g^* < \frac{\overline{\theta}^2}{2}$. This generates our desired contradiction, and shows that $\hat{g}(k) < \overline{\theta}/2$.

Step 5: To complete the proof of the first part of the theorem. I show that at $\hat{g}(k) = \inf P(k)$,

the mechanism (q^M, t^M) is optimal. We already showed that $\hat{g}(k) > g^*$, so when (q^M, t^M) is offered at $\hat{g}(k)$, it induces zero direct demand. I prove that (q^M, t^M) is optimal at $\hat{g}(k)$ by contradiction. Suppose not, i.e. at $\hat{g}(k)$, there is an optimal mechanism (q^*, t^*) with a positive direct mass, whose expected profit is strictly higher than $(1-k)\Pi^M$. As I argued earlier, we can choose this optimal mechanism such that $\pi^*(\theta) \geqslant 0$ for every θ, even those that never transact with the seller, and this implies that $\{\theta | U^*(\theta) = \theta\hat{g}(k)\}$, the set of indifferent types, is a null set.

Hold this mechanism (q^*, t^*) fixed, and consider $g_n \downarrow \hat{g}(k)$, so $g_n > \hat{g}(k)$ for every n. Let $D_n := \{\theta | U^*(\theta) < \theta g_n\}$, where $U^*(\theta) = \theta q^* \theta - t^*(\theta)$ denote the set of types that strictly prefer the public option at g_n, when (q^*, t^*) is offered. Note that in the definition of D_n, only θg_n changes. Since the monopolist can route the indifferent types however she likes, consider the routing such that only $\theta \in D_n$ demand the public option when (q^*, t^*) is offered at g_n. The monopolist's profit from (q^*, t^*) is therefore at least the profit she gets from this particular routing.

Let D, the set of types that demands the public option at $\hat{g}(k)$ in the optimal mechanism (q^*, t^*) at $\hat{g}(k)$. Since the set of indifferent types at $\hat{g}(k)$ is a null set, therefore, for θ a.e. we have $1_{D_n}(\theta) \longrightarrow 1_D(\theta)$. Recall that for any set measurable subset A of Θ, $\beta(F(A)) := \max\{0, 1 - \frac{k}{F(A)}\}$, so $\beta(A)$ is the probability of rejection from the public option when the set A demands the public option. Since the function β is continuous on $[0,1]$, therefore the almost everywhere convergence of indicator functions, and the dominated convergence theorem gives:

$$F(D_n) \longrightarrow F(D), \qquad \int_{D_n} \pi^* dF \longrightarrow \int_D \pi^* dF, \qquad \int_{D_n^c} \pi^* dF \longrightarrow \int_{D^c} \pi^* dF$$

The profit at g_n, from (q^*, t^*), and my selected routing rule is $\Pi_n := \int_{D_n} \pi^* dF + \beta(F(D_n)) \int_{D_n^c} \pi^* dF$. Since we chose an optimal mechanism (q^*, t^*) such that $\pi^* \geqslant 0$, this implies, by Lemma 7, that q^*, and t^* are bounded, therefore π^* is integrable. So, we have that $\Pi_n \longrightarrow \int_D \pi^* dF + \beta(F(D)) \int_{D^c} \pi^* dF$, which is the monopolist's profit at $\hat{g}(k)$, and by our hypothesis, this is strictly higher than $(1-k)\Pi^M$, the profit from (q^m, t^M) at $\hat{g}(k)$. So, for n large enough, the profit from (q^*, t^*), and my selected routing rule at $g_n > \hat{g}(k)$ is strictly higher than $(1-k)\Pi^M$, which is a contradiction. □

I now prove the next part of Theorem 1, i.e. the continuity, strictly increasing and limit parts. Before that, I lay down some groundwork.

We are going to prove the following.

Lemma 23. *Since we can, in our search for the optimal mechanism, restrict attention to class $\mathcal{K}$ defined in 6, the following holds:*

1. *Optimal profit is jointly continuous:*

$$(k_n, g_n) \to (k, g) \in (0, 1] \times (0, \infty) \quad \Longrightarrow \quad \Pi^*(k_n, g_n) \to \Pi^*(k, g).$$

2. *Let $(k_n, g_n) \to (k, g)$, and, for every n, choose* ***any*** *$(q_n, \underline{u}_n) \in \mathcal{K}$ that's optimal at (k_n, g_n). If $q \in L^2(F)$ and $\underline{u} \in \mathbb{R}$ satisfy*

$$\underline{u}_n \to \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_n(\theta) - q(\theta)|^2 \, dF(\theta) \to 0,$$

then $(q, \underline{u}) \in \mathcal{K}$ and $\Pi(q, \underline{u}; k, g) = \Pi^*(k, g)$.

Proof. Let $(k_n, g_n) \to (k, g)$, with all policies and the limit in $(0, 1] \times (0, \infty)$. Suppose $(q_n, \underline{u}_n) \in \mathcal{K}$ is optimal at (k_n, g_n) for each n, and suppose that:

$$\underline{u}_n \to \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2 \, dF \to 0, \qquad q \in L^2(F)$$

By Lemma 8, $\mathcal{K}$ is compact. A compact subset of a metric space is closed, so $(q, \underline{u}) \in \mathcal{K}$. By Theorem 4, optimal mechanism exists, so choose $(q^*, \underline{u}^*) \in \mathcal{K}$ optimal at (k, g). The class 6 does not depend (k, g). Thus $(q^*, \underline{u}^*)$ is also feasible at each (k_n, g_n), and optimality of $(q_n, \underline{u}_n)$ at g_n implies:

$$\Pi(q_n, \underline{u}_n; k_n, g_n) = \Pi^*(k_n, g_n) \geqslant \Pi(q^*, \underline{u}^*; k_n, g_n) \quad \text{for every } n. \tag{47}$$

From Lemma 9, by changing allocations on the left, and keeping the allocation fixed at $(q^*, \underline{u}^*)$ on the right, and taking limits in the inequality gives $\Pi(q, \underline{u}; k, g) \geqslant \Pi(q^*, \underline{u}^*; k, g) = \Pi^*(k, g)$. The reverse inequality follows from $(q, \underline{u}) \in \mathcal{K}$ and 13. Therefore, we have:

$$\Pi(q, \underline{u}; k, g) = \Pi^*(k, g), \qquad \Pi^*(k_n, g_n) = \Pi(q_n, \underline{u}_n; k_n, g_n) \longrightarrow \Pi^*(k, g). \tag{48}$$

The first equality proves the optimality of the limiting allocation, and the second proves part 2 of Lemma 23.

I now show that $\Pi^*(k_n, g_n) \longrightarrow \Pi^*(k, g)$. Suppose not. Then, there exists $\varepsilon > 0$ and indices n_j such that

$$\left|\Pi^*(k_{n_j}, g_{n_j}) - \Pi^*(k, g)\right| \geqslant \varepsilon \quad \text{for every } j$$

Again, by existence of optimal mechanism, choose $(q_{n_j}, \underline{u}_{n_j}) \in \mathcal{K}$ optimal at each (k_{n_j}, g_{n_j}). By compactness of $\mathcal{K}$, there is a further subsequence, which I continue to denote by n_j, such

that

$$\underline{u}_{n_j} \to \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_{n_j} - q|^2 \, dF \to 0$$

48 applies to this subsequence and gives us:

$$\left|\Pi^*(k_{n_j}, g_{n_j}) - \Pi^*(k, g)\right| \longrightarrow 0,$$

contradicting its lower bound ε. Thus $\Pi^*(k_n, g_n) \to \Pi^*(k, g)$. This concludes the proof. □

I now prove the second part of Theorem 1.

Claim 4. *$\hat{g} : (0,1) \to \mathbb{R}_+$ is a strictly increasing and continuous function. Moreover, $\lim_{k\downarrow 0} \hat{g}(k) = g^*$, and $\lim_{k\uparrow 1} \hat{g}(k) = \frac{\overline{\theta}}{2}$.*

Proof. Recall the definitions in 10, 11, and 9. I will use these in the proof. I begin the proof of Claim 4 in steps.

Step 1: In this step, I show that $\hat{g}$ is strictly increasing in k. I prove this by showing that for any $k \in (0,1)$, at $(k, \hat{g}(k))$, there exists a competing optimal mechanism for (q^M, t^M), where a positive mass of consumers buy directly. As k increases, the profit from this competitor becomes strictly higher than the profit from (q^M, t^M), which would prove that (q^M, t^M) cannot remain optimal at a higher k. This can only happen if $\hat{g}$ increases strictly with k.

Lemma 24. *For every $k \in (0,1)$, there is an optimal mechanism $(q, \underline{u})$ in $\mathcal{K}$ at $(k, \hat{g}(k))$ which induces a positive mass of consumers to buy directly. It satisfies*

$$\Pi(q, \underline{u}; k, \hat{g}(k))) = (1-k)\Pi^{\mathrm{M}}, \tag{49}$$

$$U(\overline{\theta}) \geqslant \overline{\theta}\hat{g}(k) > U^{\mathrm{M}}(\overline{\theta}), \tag{50}$$

$$\Pi^{\mathrm{mon}}(q, \underline{u}) < \Pi^{\mathrm{M}}. \tag{51}$$

Proof. Fix $k \in (0,1)$. I first construct a sequence of public qualities g_n below $\hat{g}(k)$, and choose optimal mechanisms at these qualities. Then, I show that the the limiting mechanism is optimal at $\hat{g}(k)$, and induces a positive mass of direct consumers. For $n = 1, 2, \ldots$, let

$$g_n = \hat{g}(k) - \frac{\hat{g}(k) - g^*}{n+1}$$

By the first part of Theorem 1, $\hat{g}(k) > g^*$, so $g^* < g_n < \hat{g}(k)$, and $g_n \uparrow \hat{g}(k)$. By Theorem 4, which shows existence of the optimal mechanism, at every (k, g_n), there is an optimal

mechanism $(q_n, \underline{u}_n) \in \mathcal{K}$. Thus,

$$\Pi(q_n, \underline{u}_n; k, g_n) = \Pi^*(k, g_n)$$

Let $U_n(\theta) = \underline{u}_n + \int_{\underline{\theta}}^{\theta} q_n(s)\, ds$ for the utility of type θ from $(q_n, \underline{u}_n)$. Since $g_n < \hat{g}(k)$, therefore again, from the first part of Theorem 1, at $(q_n, \underline{u}_n)$, a positive mass of consumers must buy directly. Since the consumers who buy directly in an optimal mechanism form an upper interval, therefore $\overline{\theta}$ buys directly, so $U_n(\overline{\theta}) \geqslant \overline{\theta} g_n$.

I now take a limit. By Lemma 8, there exists a subsequence, relabeled n, and a mechanism $(q, \underline{u}) \in \mathcal{K}$ that satisfies:

$$\underline{u}_n \longrightarrow \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2\, dF \longrightarrow 0.$$

From 14, we have:

$$\sup_{\theta} |U_n(\theta) - U(\theta)| \leqslant |\underline{u}_n - \underline{u}| + \sqrt{\frac{\overline{\theta} - \underline{\theta}}{\underline{f}}} \left(\int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2\, dF \right)^{1/2} \longrightarrow 0.$$

To establish optimality of this limit, I apply Lemma 9 to the converging allocations and $(k, g_n) \to (k, g_F(k))$, and Lemma 23 to the optimal values. This gives us:

$$\begin{aligned} \Pi(q, \underline{u}; k, \hat{g}(k)) &= \lim_n \Pi(q_n, \underline{u}_n; k, g_n) \\ &= \lim_n \Pi^*(k, g_n) \\ &= \Pi^*(k, \hat{g}(k))). \end{aligned}$$

The first equality is joint profit continuity from Lemma 9; the second is from optimality of each selected menu; the third is optimal profit continuity from Lemma 23. Thus $(q, \underline{u})$ is optimal at the threshold. Uniform utility convergence gives us:

$$U(\overline{\theta}) = \lim_n U_n(\overline{\theta}) \geqslant \lim_n \overline{\theta} g_n = \overline{\theta} \hat{g}(k) > \overline{\theta} g^* = U^{\mathrm{M}}(\overline{\theta}).$$

The strict inequality is by $\hat{g}(k) > g^*$, and the last equality is by the definition of g^*. So, the limiting mechanism $(q, \underline{u})$ cannot be obe that induces zero direct demand, otherwise by Proposition 2, $(q, \underline{u}) = (q^M, t^M)$, which gives $U(\overline{\theta}) = U^{\mathrm{M}}(\overline{\theta})$, and contradicts the strict inequality obtained above. Hence, it has a positive mass of direct buyers. At $(k, \hat{g}(k))$, the mechanism (q^M, t^M) inducing zero direct demand is also optimal, by the first part of Theorem

1, and its profit is $(1-k)\Pi^M$. So, we get,

$$\Pi(q, \underline{u}; k, \hat{g}(k)) = \Pi^*(k, g_F(k)) = (1-k)\Pi^{\mathrm{MR}}.$$

This proves 49 and 50.

Finally, $\Pi^{\mathrm{mon}}(q, \underline{u}) \leqslant \Pi^{\mathrm{M}}$, since (q^M, t^M) is the unique optimal mechanism for the pure monopoly benchmark. If $\Pi^{\mathrm{mon}}(q, \underline{u}) = \Pi^{\mathrm{M}}$, then, we must have $\underline{u} = 0$ and $q = q^{\mathrm{M}}$ F-almost everywhere. This would imply $U(\overline{\theta}) = U^M(\overline{\theta})$, contradicting 50. Thus $\Pi^{\mathrm{mon}}(q, \underline{u}) < \Pi^{\mathrm{M}}$, proving 51. This finishes the proof of Lemma 24. □

I now complete ***Step*** 1 by showing that $\hat{g}(k)$ is strictly increasing in k. Fix $0 < k_0 < k_1 < 1$. Keep public quality fixed at $\hat{g}(k_0)$, and consider the competing optimal mechanism $(q, \underline{u})$ from Lemma 24. I will holds the mechanism $(q, \underline{u}$ fixed, and show that at k_1, it strictly beats (q^M, t^M). Thus, the actual optimal mechanism at k_1 can only do better than $(q, \underline{u})$, so at k_1, the mechanism (q^M, t^M), and therefore zero direct demand cannot be optimal. This would prove that $\hat{g}(k_1)$,the zero direct demand threshold at k_1, is strictly higher than $\hat{g}(k_0)$. Let D be the set of types that demand the public option when $(q, \underline{u}$ is offered at public quality $\hat{g}(k_0)$. We do not have to worry about indifferent types, because by choice, the optimal mechanism is from the set $\mathcal{K}$, so the set of indifferent types is a null set. For $k_1 > k_0 > 0$,

$$\frac{p(k_1, D)}{k_1} = \frac{1}{\max\{k_1, D\}} \leqslant \frac{1}{\max\{k_0, D\}} = \frac{p(k_0, D)}{k_0}. \tag{52}$$

The integral of π over D is nonnegative. Since $\Pi(q, \underline{u}; k, g) = \Pi^{\mathrm{mon}}(q, \underline{u}) - p(k, D) \int_{\{U(\theta) < \theta g\}} \pi(\theta)\, dF(\theta)$, therefore, combined with 52, we have:

$$\begin{aligned}
\Pi(q, \underline{u}; k_1, \hat{g}(k_0)) &\geqslant \Pi^{\mathrm{mon}}(q, \underline{u}) - \frac{k_1}{k_0} p(k_0, D) \int_{\{U(\theta) < \theta g_F(k_0)\}} \pi(\theta)\, dF(\theta) \\
&= \frac{k_1}{k_0} \Pi(q, \underline{u}; k_0, \hat{g}(k_0)) - \left(\frac{k_1}{k_0} - 1\right) \Pi^{\mathrm{mon}}(q, \underline{u}).
\end{aligned}$$

Substituting 49 and subtracting the profit from (q^M, t^M) at k_1, we have,

$$\begin{aligned}
&\Pi(q, \underline{u}; k_1, \hat{g}(k_0)) - (1-k_1)\Pi^{\mathrm{M}} \\
&\quad \geqslant \left[\frac{k_1}{k_0}(1-k_0) - (1-k_1)\right] \Pi^{\mathrm{M}} - \left(\frac{k_1}{k_0} - 1\right) \Pi^{\mathrm{mon}}(q, \underline{u}) \\
&\quad = \frac{k_1 - k_0}{k_0} \left[\Pi^{\mathrm{M}} - \Pi^{\mathrm{mon}}(q, \underline{u})\right] > 0.
\end{aligned} \tag{53}$$

The last inequality is by 51. Thus, we have that the optimal profit at $(\hat{g}(k_0), k_1)$, which is

$\Pi^*(k_1, g_F(k_0))$, strictly exceeds $(1-k_1)\Pi^M$, the profit from zero direct demand. Thus, zero direct demand cannot be optimal, and therefore $\hat{g}(k_1) > \hat{g}(k_0)$. This completes the proof that $\hat{g}$ is strictly increasing in k.

Step 2. I now show that $\hat{g}(k)$ is continuous in k. Take a sequence $k_n \in (0,1)$ with $k_n \to k \in (0,1)$. I want to show that $\hat{g}(k_n) \to \hat{g}(k)$. The bounds on $\hat{g}$ from the first part of Theorem 1 means that the threshold sequence $\hat{g}(k_n)$ is bounded. By Bolzano Weierstrass, there is a convergent subsequence which has limit $g_F(k)$. Taking such a subsequence, and relabeling it n, we have $\hat{g}(k_n) \longrightarrow g \in [g^*, \overline{\theta}/2]$.For each n, Lemma 24 gives a competing optimum $(q_n, \underline{u}_n)$ at $(k_n, g_F(k_n))$, with

$$\Pi(q_n, \underline{u}_n; k_n, \hat{g}(k_n)) = (1-k_n)\Pi^{\mathrm{M}}, \qquad U_n(\overline{\theta}) \geqslant \overline{\theta}\hat{g}(k_n) \tag{54}$$

By Theorem 4, we also have a further subsequence and $(q, \underline{u}) \in \mathcal{K}$ satisfying

$$\underline{u}_n \to \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2 \, dF \to 0, \qquad \sup_{\theta \in [\underline{\theta}, \overline{\theta}]} |U_n(\theta) - U(\theta)| \to 0 \tag{55}$$

Lemma 23 gives the optimality of the limiting menu, and Lemma 9 gives the joint continuity of the realized profit. Thus,

$$\begin{aligned} \Pi^*(k, g) &= \Pi(q, \underline{u}; k, g) \\ &= \lim_n \Pi(q_n, \underline{u}_n; k_n, g_F(k_n)) \\ &= \lim_n (1-k_n)\Pi^{\mathrm{M}} = (1-k)\Pi^{\mathrm{M}}. \end{aligned} \tag{56}$$

The first equality is by Lemma 23, the second is by Lemma 9, and the third is by substituting 54. The uniform utility convergence in 14, together with the utility inequality in (54) gives us

$$U(\overline{\theta}) = \lim_n U_n(\overline{\theta}) \geqslant \lim_n \overline{\theta} g_F(k_n) = \overline{\theta} g. \tag{57}$$

I now use these two conclusions to identify g where recall that $\hat{g}(k_n) \longrightarrow g$. First, $g \geqslant g^*$, so if (q^M, t^M) is offered at g, it induces zero direct demand, and its profit is $(1-k)\Pi^M$. Hence, zero direct demand is optimal, and thus, $g \geqslant \hat{g}(k)$. If $g > \hat{g}(k)$, then (q^M, t^M) is the unique optimal mechanism at (k, g), which means that the utility function is given by U^M. This contradicts (57), because

$$U^{\mathrm{M}}(\overline{\theta}) = U(\overline{\theta}) \geqslant \overline{\theta} g > \overline{\theta} g^* = U^{\mathrm{M}}(\overline{\theta}).$$

Thus, we must have $g = \hat{g}(k)$. This finishes the continuity proof.

Step 3: I now show that $\hat{g}(k) \longrightarrow g^*$ as $k \to 0$. Fix $g > g^*$. I first show that $\hat{g}(k) \leqslant g$ for all sufficiently small k. The monopoly optimal menu (q^M, t^M) induces zero direct demand at g, and at any k, its profit is $(1-k)\Pi^M$, which tends to Π^M as $k \to 0$. A competing optimal mechanism would have to approach the monopoly profit Π^M while continuing to give top type at least $\overline{\theta}g$. Suppose that $\hat{g}(k) \leqslant g$ for all sufficiently small k is false. So, we can choose $k_n \in (0,1)$ such that $k_n \longrightarrow 0$, and $g < g_F(k_n)$ for every n. At each (k_n, g), existence of optimal mechanism gives an optimum $(q_n, \underline{u}_n) \in \mathcal{K}$. Since $g < \hat{g}(k_n)$, the optimal mechanism $(q_n, \underline{u}_n)$ cannot induce zero direct demand, thus, the top type buys directly, and we have $U_n(\overline{\theta}) \geqslant \overline{\theta}g$. Because $g > g^*$, thus zero direct demand is feasible, and that, combined with optimality gives the first inequality below. The second inequality follows from $\Pi(q_n, \underline{u}_n; k, g) = \Pi^{\text{mon}}(q_n, \underline{u}_n) - p(k, D_n) \int_{\{U_n(\theta) < \theta g\}} \pi_n(\theta)\, dF(\theta)$, and the third from the fact that (q^M, t^M) is the unique maximizer for the monopoly problem.

$$(1-k_n)\Pi^{\text{M}} \leqslant \Pi(q_n, \underline{u}_n; k_n, g) \leqslant \Pi^{\text{mon}}(q_n, \underline{u}_n) \leqslant \Pi^{\text{M}}. \tag{58}$$

In particular,

$$0 \leqslant \Pi^{\text{MR}} - \Pi^{\text{mon}}(q_n, \underline{u}_n) \leqslant k_n \Pi^{\text{MR}} \longrightarrow 0.$$

From compactness of $\mathcal{K}$, we have a subsequence converging to $(q, \underline{u}) \in \mathcal{K}$. Lemma 23 gives:

$$\Pi^{\text{mon}}(q, \underline{u}) = \lim_n \Pi^{\text{mon}}(q_n, \underline{u}_n) = \Pi^{\text{M}}.$$

Thus, we must have $\underline{u} = 0$, and $q = q^M$ almost everywhere. Thus, from the envelope formula, we have that the limiting utility is U^M, and uniform utility convergence from]46, and $U(\overline{\theta}) \geqslant \overline{\theta}g$ gives us the following contradiction:

$$U^{\text{M}}(\overline{\theta}) = \lim_n U_n(\overline{\theta}) \geqslant \overline{\theta}g > \overline{\theta}g_{\text{cap}} = U^{\text{M}}(\overline{\theta}).$$

Thus, I have proved that for small enough k we must have $\hat{g}(k) \leqslant g$, for every $g > g^*$. For $\varepsilon > 0$, apply this result at $g = g_{\text{cap}} + \varepsilon$. By the first part of Theorem 1, in particular the bounds on $\hat{g}$, we get:

$$g^* < \hat{g}(k) \leqslant g^* + \varepsilon \quad \text{for all sufficiently small } k > 0.$$

This proves that $\hat{g}(k) \longrightarrow g^*$ as $k \to 0$.

Step 4. I now show the final part, i.e. $\hat{g}(k) \longrightarrow \overline{\theta}/2$ as $k \to 1$. Fix $0 < g < \overline{\theta}/2$. I will

give a feasible mechanism with profit bounded away from zero, irrespective of k, so this mechanism rules out an optimal mechanism with zero direct demand at g. I now construct the mechanism. The menu corresponding to this mechanism consists of the null contract $(0,0)$, and one other contract satisfying

$$q = \overline{\theta}, \qquad \frac{\overline{\theta}^2}{2} < t < \overline{\theta}(\overline{\theta} - g).$$

Such a transfer exists because $\overline{\theta}(\overline{\theta} - g) - \frac{\overline{\theta}^2}{2} = \overline{\theta}\left(\frac{\overline{\theta}}{2} - g\right) > 0$. The profit margin for the monopolist from this contract is $t - \overline{\theta}^2/2 > 0$. Every $\theta > \frac{t}{\overline{\theta} - g}$ satisfies $\theta\overline{\theta} - t > \theta g \geqslant 0$ and therefore buys directly. Since $t/(\overline{\theta} - g) < \overline{\theta}$, the direct consumers have positive mass. Thus, for every $k \in (0,1)$, we have:

$$\Pi^*(k,g) \geqslant \left(t - \frac{\overline{\theta}^2}{2}\right)\left[1 - F\left(\max\left\{\underline{\theta}, \frac{t}{\overline{\theta} - g}\right\}\right)\right] > 0. \tag{59}$$

The positive bound in (59) is independent of k, whereas $(1-k)\Pi^M \to 0$ as $k \to 1$. Thus, for k sufficiently close to one, we have $\Pi^*(k,g) > (1-k)\Pi^M$. Thus, at (k,g), where k is sufficiently close to one, zero direct demand cannot be optimal, so $g < g_F(k)$. For $0 < \varepsilon < \overline{\theta}/2$, apply this argument at $g = \overline{\theta}/2 - \varepsilon$. Thus, we have:

$$\frac{\overline{\theta}}{2} - \varepsilon < \hat{g}(k) < \frac{\overline{\theta}}{2} \quad \text{for all } k < 1 \text{ sufficiently close to one.}$$

This completes the proof for the second limit. □

This completes the proof of Theorem 1.

6.10 Proof of Theorem 2

I now prove the surplus drop result. I first state and prove a useful result. We know from Theorem 1 that for any$g < \hat{g}(k)$, the optimal mechanism must induce a cutoff strictly lower than $\overline{\theta}$, i.e. a positive mass of consumers must buy directly. The following lemma shows that for any k, as g approaches $\hat{g}(k)$ from below, the cutoff remains bounded away from $\overline{\theta}$.

Lemma 25. ***Top Separation Lemma****: There exist an $\eta_0 > 0$, and an $\varepsilon_0 > 0$ such that for every $g \in (\hat{g}(k) - \eta_0, \hat{g}(k))$, and for any optimal mechanism at g with an interior cutoff c, it must be that $c \leqslant \overline{\theta} - \varepsilon_0$.*

Proof. I prove the above lemma by contradiction. The idea is that if there exists a sequence of optimal cutoffs that approach $\overline{\theta}$, then, the contribution of the direct buyers to the profit

vanishes. The limiting mechanism must then solve the pure monopoly problem, because offering the menu (q^M, t^M) guarantees a profit of $(1-k)\Pi^M$. So, the utility function associated with the limiting mechanism must be U^M, but indifference at the optimal cutoff, which is strictly lower than $\overline{\theta}$, would give $\overline{\theta}$ utility $\overline{\theta}\hat{g}(k) > U^M(\overline{\theta})$, a contradiction. I do the proof in steps:

Step 1: Suppose the statement in the Lemma is false. Then, for each n, choose

$$\hat{g}(k) - \frac{1}{n} < g_n < \hat{g}(k), \qquad \overline{\theta} - \frac{1}{n} < c_n < \overline{\theta},$$

and an optimal mechanism at g_n with interior cutoff c_n. Thus $g_n \to \hat{g}(k)$ and $c_n \to \overline{\theta}$. Since $\hat{g}(k) > g^*$, $F(c_n) \to 1$, and $k < 1$, for n large enough, we have $g_n > g^*$ and $F(c_n) > k$. From now on, I only consider such n. Let the mechanisms I have chosen for each g_n be denoted by $(q_n, \underline{u}_n)$, and set

$$U_n(\theta) = \underline{u}_n + \int_{\underline{\theta}}^{\theta} q_n(s)\, ds, \qquad \pi_n(\theta) = \theta q_n(\theta) - U_n(\theta) - \frac{q_n(\theta)^2}{2}.$$

Because $F(c_n) > k$, every type faces the monopolist with positive probability. Thus, we must have $\pi_n(\theta) \geqslant 0$ for all θ, and therefore, Lemma 7 and the IC, IR characterization in Lemma 1 give us that $(q_n, \underline{u}_n) \in \mathcal{K}$, where $\mathcal{K}$ is defined in 6. By Lemma 8, there exists a subsequence, relabeled by n, and a $(q, \underline{u}) \in \mathcal{K}$ such that

$$\underline{u}_n \longrightarrow \underline{u}, \qquad \int_{\underline{\theta}}^{\overline{\theta}} |q_n - q|^2\, dF \longrightarrow 0. \tag{60}$$

Define $U(\theta) = \underline{u} + \int_{\underline{\theta}}^{\theta} q(s)\, ds$. Equation 14 then gives

$$\sup_{\theta \in [\underline{\theta}, \overline{\theta}]} |U_n(\theta) - U(\theta)| \longrightarrow 0. \tag{61}$$

Step 2: Applying Lemma 9 to the allocation sequence in (60) and the sequence $(k, g_n) \to (k, \hat{g}(k))$, we get $F(c_n) = D(q_n, \underline{u}_n; g_n) \longrightarrow D(q, \underline{u}; \hat{g}(k)$. Since $F(c_n) \to 1$, it follows that $D(q, \underline{u}; \hat{g}(k)) = 1$, so at the limiting mechanism, public demand is one. By the profit continuity in Lemma 9, we get $\Pi(q_n, \underline{u}_n; k, g_n) \longrightarrow \Pi(q, \underline{u}; k, \hat{g}(k)$. Equations (8), (10), and (11) imply

$$\Pi(q, \underline{u}; k, \hat{g}(k)) = (1-k) \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta q(\theta) - U(\theta) - \frac{q(\theta)^2}{2} \right] dF(\theta) = (1-k)\Pi^{\text{mon}}(q, \underline{u}).$$

Combining thesse two, we have

$$\Pi(q_n, \underline{u}_n; k, g_n) \longrightarrow (1-k)\Pi^{\text{mon}}(q, \underline{u}). \tag{62}$$

For every positive θ, Lemma 17(i) gives $U^M(\theta) \leqslant \theta g^* < \theta g_n$. Thus the monopoly menu at g_n induces full public demand and earns profit $(1-k)\Pi^M$. Optimality of the chosen mechanism gives $\Pi(q_n, \underline{u}_n; k, g_n) \geqslant (1-k)\Pi^M$. Taking limits by 62, and using $1-k>0$, gives $\Pi^{\text{mon}}(q, \underline{u}) \geqslant \Pi^M$. Conversely, $(q, \underline{u})$ is IC and IR and is therefore feasible for the ordinary monopoly problem. Thus $\Pi^{\text{mon}}(q, \underline{u}) = \Pi^M$. I now apply the inequality in 45, and applying to $(q, \underline{u})$ and gives

$$0 = \Pi^M - \Pi^{\text{mon}}(q, \underline{u}) \geqslant \underline{u} + \frac{1}{2}\int_{\underline{\theta}}^{\overline{\theta}} |q - q^M|^2 \, dF \geqslant 0.$$

Thus $\underline{u} = 0$ and $q = q^M$ almost everywhere. By (14), $U = U^M$ at every type.

Step 3: For every g_n, by the indifference of the optimal cutoff c_n, we have $U_n(c_n) = c_n g_n$.By (61), $U = U^M$, and continuity of U^M,

$$|U_n(c_n) - U^M(\overline{\theta})| \leqslant \sup_{\theta} |U_n(\theta) - U^M(\theta)| + |U^M(c_n) - U^M(\overline{\theta})| \longrightarrow 0.$$

On the other hand, $c_n g_n \to \overline{\theta}\hat{g}(k)$. It follows that

$$\overline{\theta} g^* = U^M(\overline{\theta}) = \lim_{n\to\infty} U_n(c_n) = \lim_{n\to\infty} c_n g_n = \overline{\theta}\hat{g}(k),$$

contradicting $\hat{g}(k) > g^*$ and $\overline{\theta} > 0$. This proves the lemma. If necessary, reduce ε_0 so that $0 < \varepsilon_0 < \overline{\theta} - \underline{\theta}$; the conclusion remains valid. □