Public Sector Rationing and Private Sector Selection

Simona Grassi
Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)
España
sgrassi@eco.uc3m.es

Ching-to Albert Ma
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
USA
ma@bu.edu

Incomplete, preliminary. Do not quote, cite, or circulate.

March 2008

Acknowledgement: We thank seminar participants at Boston University and at the Health Economic Policy conference in Crans Montana, Switzerland. The first author thanks the Italian Fulbright Foundation for financial support.
Abstract

We consider the interaction between a public sector and a private sector. The public sector has a limited budget to provide a good to some consumers. A private firm may supply the same good to those consumers who do not receive the good at the public system. Consumers differ in two dimensions. They have different wealth levels, and the costs of supplying this good to them may also differ. The wealth heterogeneity dimension is self-explanatory. The cost heterogeneity dimension arises because consumer characteristics may determine how much it costs to supply the good. In the health market, patients’ illness severity affects how much medical resources a course of treatment requires.

The public regulator observes consumers’ wealth information, but not the cost information. The private firm observes consumers’ cost information, but not the wealth information. The public regulator aims to maximize total consumer utility subject to the budget constraint, while the private firm aims to maximize profit. The public regulator sets a rationing scheme that assigns the good to consumers according to wealth information. The private firm sets a price to sell to consumers according to cost information. In the simultaneous-move game, the public regulator chooses the rationing scheme and the firm chooses the pricing scheme simultaneously. In the sequential-move game, the public regulator first commits to the rationing scheme, and then the firm chooses the pricing scheme.

We characterize equilibria in the simultaneous-move game. In each equilibrium, the regulator rations all consumers with high wealth, while the firm will stop reducing its price even as cost falls. There is a continuum of equilibria, but the characterization fits each one of them. The best equilibrium for the regulator is one where the entire budget is spent on consumers with low wealth. The equilibrium pricing rule in the private sector exhibits selection. The private sector only sells to consumers who have high wealth levels and therefore high willingness to pay. Cournot competition in the private sector does not alter these results qualitatively. Under perfect competition, the rationing scheme that allocates the entire budget to consumers with low wealth is the unique equilibrium.

Because the private sector has no access to poor consumers due to the public supply, prices in the private sector tend to be high, and will stop falling even when cost decreases. In an equilibrium of the sequential-
move game, the public regulator uses a rationing scheme to implement pricing schemes that are strictly increasing in costs. We show that there are rationing rules that implement a uniform price reduction in the private sector from the monopoly pricing rule. The key is to ration some consumers who are poor and who have low willingness to pay. As cost decreases, the private firm will find it optimal to reduce the price to sell to these poor consumers.

We construct a binary example (where consumers’ wealth is either high or low) to show that the equilibrium rationing rule will ration the poor even when the budget is sufficient to cover all poor consumers. Leaving poor consumers in the private market implements a price reduction that benefits the wealthy consumers, and raises total consumer welfare.
1 Introduction

Many governments and public organizations provide or subsidize the consumption of goods and services such as education and health care. Public provision often coexists with a private market. While firms seek to maximize profits, governments are concerned with goals such as consumer surplus, universal access, equity and redistribution.

We study the interaction between a public regulator and private firms when they provide the same indivisible good to heterogeneous consumers. Consumers differ in two dimensions: wealth and cost. The cost heterogeneity dimension arises because consumer characteristics may determine how much it costs to supply the good. The regulator rations the access to the consumption of the good to some but not all consumers, and private firms can price discriminate.

A consumer’s utility depends on wealth and on the benefit derived from the consumption of the good, not directly on his cost characteristic. On the other hand, the cost dimension directly influences producers’ decisions. For example, in the market for health care, a treatment can turn out to be more expensive for some patients than others, because patients are of different ages and have different pre-existing health status. A profit maximizing health care provider would make use of the cost information if it is available. We assume that all consumers derive an equal level of benefit (in utility unit) from the consumption of one unit of the good. We choose this assumption in order to focus on two dimensions of heterogeneity, wealth and cost.

The regulator observes consumers’ wealth levels but not their costs, while private firms have access to cost but not wealth information. We think that this is a natural assumption to make. The regulator has information on consumers’ tax returns; moreover, even if the regulator observes the cost level, it is unlikely that it may use this information, because price discrimination may be infeasible due to equity consideration.

Cost heterogeneity and mixed (public and private) provision generate problems of consumer selection and crowding out. So far many works, especially in the field of health care, have focused on the incentive for firms (and insurance companies) to select low costs consumers and to leave high cost consumers in the public sector. Other studies have analyzed the crowding out effect of public provision on private supply. Public provision crowds out private provision when an increase in the public supply is matched with a decrease in
the demand for the good and supply by the private firms.

In this paper, we set up a model that allows us to consider selection and crowding out simultaneously; we depart from the literature by introducing heterogeneity in both cost and wealth dimensions, while the literature has mostly focused on cost heterogeneity. We are able to identify a *wealth effect* that is ignored by other models. Fundamentally, if the market price of the good depends only on the cost level, among consumers with an equal cost level, the wealthier consumer gets a higher surplus than the less wealthy if he purchases the good. This simple intuition drives the main results of the paper.

We consider a public regulator with a limited budget that is insufficient to supply the good for free to all consumers. The public regulator uses a rationing policy based on wealth \( w \) to allocate the good to consumers; a monetary subsidy or tax policy is infeasible, and may have already been implemented earlier or by a different administrative body. We abstract from redistribution and taxation issues, and interpret the consumer’s wealth as net of taxes and subsidies. Further changes in the income distribution of consumers served by the public sector is not considered here so that we can focus on pricing and selection issues.

The public regulator’s rationing rule consists of a function \( \theta : [\underline{w}, \overline{w}] \rightarrow [0, 1] \). Wealth is a random variable distributed according to a distribution function \( F \). For any given \( w \), the public regulator supplies consumers with wealth below \( w \) a total of \( \int_{\underline{w}}^{w} (1 - \theta(x)) f(x) \, dx \) units of the good at zero cost to them, but not the remaining consumers \( \int_{w}^{\overline{w}} \theta(x) f(x) \, dx \). The rationing rule \( \theta \) modifies the density \( f \) so that at \( w \), \( [1 - \theta(w)] f(w) \) of consumers are provided by the public sector at zero price. Because the rationing rule is based on wealth, the cost \( c \) among rationed consumers is distributed according to an exogenous distribution function \( G \). Cost and wealth are independently distributed among the population. Our rationing rule corresponds to random rationing; if there are discrete wealth classes of consumers, for each wealth class, the rationing rule will supply a fraction of consumers.

We consider different kinds of private markets: a monopoly, an oligopoly and a perfect competitive market. We focus on the monopoly case and prove that the main results can be extended to different kind of markets. The monopolistic firm observes consumers’ cost levels and charges prices that increase with cost. A high-cost consumer can afford to buy the good only if he is wealthy enough. Low-cost consumers pay a
lower price; less wealthy consumers can afford to buy the good only if their cost is low. Selection occurs in
two dimensions: wealth and cost.

The public regulator’s rationing rule affects the demand available to the monopolist. A crowding out
effect arises: those consumers who are supplied by the public sector exit the private market, and the price
function \( p(c) \) in the private market changes as a result of the rationing policy \( \theta(w) \). A policy based on wealth
influences the selection in the private market through its effects on the cost-based price function.

The public regulator chooses the rationing policy to maximize total consumers’ surplus, given a budget
constraint. We consider two kinds of interaction between the public regulator and the monopolist. If the
regulator cannot commit to a rationing policy, the rationing policy and the pricing strategy are chosen
simultaneously by the public regulator and the firm, respectively. If the announced rationing policy is
credible, the game between the public regulator and the firm is sequential.

The equilibrium rationing policies in the simultaneous and sequential games are different but they share
a common key feature: for a given cost level, and corresponding price in the market, the surplus generated
by leaving a wealthy consumer in the market is greater than that created by leaving a less wealthy consumer,
provided that they pay the same price.

In the simultaneous-move game, we derive a continuum of equilibria. In the equilibrium associated
with the highest consumer welfare, the rationing policy rations wealthy consumers from the public sector.
Precisely, consumers with wealth below a threshold level \( \tilde{w}^* \), get the good for free from the public regulator,
while those above the threshold are rationed from the public sector and can buy the good in the private
market. In other words, \( \theta(w) = 1 \) for \( w > \tilde{w}^* \) and \( \theta(w) = 0 \) for \( w < \tilde{w}^* \), where \( \tilde{w}^* \) is the highest wealth level
among poor consumers who exhaust the budget.

The corresponding equilibrium pricing strategy is a weakly increasing function of cost. Consumers with
wealth above \( \tilde{w}^* \) buy the good at a constant price if their cost is below an endogenously determined cost
threshold \( \tilde{c}^* \); the price rises with cost for cost levels above \( \tilde{c}^* \). Wealthy consumers will have to pay more
as their costs rise; some wealthy consumers may not buy at all. In terms of selection, among the high-cost
consumers, only those with very high wealth can afford the good in the private market, while consumers
with average wealth are too rich to be supplied by the public sector but not wealthy enough to buy the good from the monopolist.

This result is surprising: our regulator is not concerned with equity or redistribution, and in the absence of a private sector it would be completely indifferent between any budget equivalent wealth-based rationing policies. In the presence of a private market where the price is based on cost, the equilibrium rationing policy favors poorer consumers. The result is driven by the *wealth effect* and the fact that the regulator cannot commit to a policy. For each cost level, total surplus is higher when wealthier consumers trade in the market, because they gain more surplus than poorer consumers by buying at a fixed price $p(c)$. Were the market competitive or oligopolistic with firms competing *a la* Cournot, we would observe the same kind of equilibrium rationing policy.

In the sequential game, the public regulator anticipates the firm’s reaction to a given rationing policy. It remains true that when the market price is based on cost, a wealthier consumer gains more surplus in the market than a less wealthy consumer. Nevertheless, the regulator can now commit to a policy and exploits its first-mover advantage by choosing a policy that maximizes the surplus generated by trades in the market. If the regulator commits to make poorer consumers unavailable to the private sector, the monopolist does not lower the prices even when costs are low. This cannot be an equilibrium rationing strategy in the sequential-move game. We prove that consumer surplus can be increased by leaving some poorer consumers in the market. By leaving some poorer consumers in the market, the regulator induces the monopolist to reduce the price at each cost level. The total surplus must be increased: those consumers who were buying the good at the old higher price, can now buy at a lower price. Rationing does lead to selection, but prices are lower at each cost level. We construct a binary example (where consumers' wealth is either high or low) to show that the equilibrium rationing rule will still ration the poor, even when the budget is sufficient to cover all poor consumers.

Many works in health economics (for example Barros and Olivella [2], Ellis [8]), study selection and crowding out. They focus on cream skimming, which is the selection of low cost (low severity) patients. Barros and Olivella focus on waiting lists and patient selection. They show that despite private physicians
usually select the lowest cost cases from those on the waiting lists for the public hospital, this is not necessarily true for very strict or very loose rationing policy (respectively, when only the most severe cases are admitted in the public hospital or also very mild cases are admitted). Ellis studies the effect of different reimbursement schemes on cream skimming (and dumping.) Wealth heterogeneity is acknowledged, but it is not made operative in these models. We prove that a public regulator with wealth information does not eliminate cost selection, provided that those consumers who are most disadvantaged by selection in terms of wealth are supplied by the public sector or only some of them are left in the market to induce the monopolist to lower prices.

We introduce the main building blocks of the model in the subsections of Section 2. In Section 3, we describe the two possible extensive forms of the game played by the public regulator and the private firm. Section 4 is about the price best responses in the private market. In Sections 5 and 6, we derive equilibrium rationing and pricing schemes in the simultaneous-move and sequential-move games, respectively. The last Section draws some conclusions.

2 The Model

We begin with the description of consumers in the model. Next, we introduce a private sector and then a public sector. We complete the model description by the extensive-form games between the consumers, public, and private sectors.

2.1 Consumers and willingness to pay

There is a set of consumers. Each consumer may consume at most one unit of an indivisible good. We let there be a continuum of these consumers, with total mass normalized to 1. Each consumer is indexed by two parameters, \(w\) and \(c\). The variable \(w\) denotes the consumer’s wealth. The variable \(c\) denotes the cost of supplying this good to the consumer. The cost \(c\) of provision is identical whether the good is supplied by the public or private sectors; we do not consider any productive comparative advantage between the private and public sectors in order to focus on information problems. We often use the term consumer \((w, c)\) to refer to one who has wealth \(w\) and cost parameter \(c\).
The variables $w$ and $c$ are assumed to be independently distributed. Let $F : [w, \overline{w}] \rightarrow [0, 1]$ be the distribution function of $w$. We further assume that $F$ is differentiable, and let the corresponding density be $f$. Similarly, let $G : [c, \overline{c}] \rightarrow [0, 1]$ be the distribution function of $c$. We also assume that $G$ is differentiable, and let the corresponding density be $g$. The domains of these two distributions are strictly positive and bounded. We will also assume that the densities $f$ and $g$ are both strictly positive over their domains. This implies that the distribution functions $F$ and $G$ are both strictly increasing.

A consumer derives a utility from the good. In this study we concentrate on problems arising from missing information about wealth or cost, so we suppose that the consumption of the good gives a fixed utility to each consumer, irrespective of his wealth or cost. We further normalize the increment in utility from consuming one unit of the good to 1. We work with a convenient preference specification: if the consumer with wealth $w$ pays a price $p$ to consume the good, a type $(w, c)$ consumer’s utility is $U(w - p) + 1$. The consumer’s utility is $U(w)$ if he does not consume the good.\(^1\) The utility function $U$ is strictly increasing, and strictly concave.

Consider a consumer with wealth $w$. If a consumer with wealth $w$ is indifferent between paying a price $\tau$ to consume the good and the status quo, we have:

$$U(w - \tau) + 1 = U(w). \tag{1}$$

This equation implicitly defines a willingness-to-pay function $\tau : [w, \overline{w}] \rightarrow \mathbb{R}_+$ for consumers with various wealth levels. Since $U$ is concave and hence almost everywhere differentiable, the willingness to pay function is differentiable. From total differentiation of (1), we have

$$\frac{d\tau}{dw} = 1 - \frac{U'(w)}{U'(w - \tau)} > 0. \tag{2}$$

A consumer’s willingness to pay for the good is strictly increasing in wealth due to the strict concavity of $U$.

\(^1\)We can use a general utility function where the utilities from consuming the good at price $p$ and from not consuming the good are $U(w - p, 1)$ and $U(w, 0)$, respectively. This will just require more notation (since we will have to carry the 0 around). Our analysis only requires the monotonicity that consumers with higher wealth are more willing to pay for the good.
2.2 The private sector and profit-maximizing prices

We now describe a private sector. We let there be a single firm monopoly in the private sector; in a later section, we will consider a Cournot private sector as well as a perfectly competitive private sector. The selection issue we focus on is this. Facing any consumer, the private firm may observe the cost of providing a unit of the good to the consumer, but the private firm does not get to observe a consumer’s wealth level. For consumer \((w, c)\), the private firm gets to observe only \(c\), but not \(w\).

The monopolist aims to maximize profit by setting a price to sell to consumers, given the cost \(c\). For now, assume that the monopolist may have access to the entire mass of consumers. Not knowing the consumer’s wealth, the monopolist does not know the willingness to pay. By setting a price (that is above cost \(c\)), the monopolist may sell to those consumers with willingness to pay higher than the price. Obviously, the monopolist will not set a price outside the range of willingness to pay \(\tau\). Setting a price \(p\) is equivalent to selecting the wealth level of the marginal consumer \(w\), where \(p = \tau(w)\). By the strictly monotonicity of \(\tau\), consumers with \(w' > w\) has \(\tau(w') > \tau(w)\), and hence are willing to pay \(p\) to purchase the good. The function \(\tau\) is like a demand function; we simply restate the common principle that a monopolist may choose equivalently between a price and a quantity while respecting the demand function.

Suppose that the monopolist knows the cost \(c\) of providing the good to a consumer. If it intends to sell to consumers with wealth \(w\) or higher, it sets a price \(\tau(w)\), and its profit is

\[
\pi(w; c) \equiv \int_{w}^{\bar{w}} f(x)dx \ [\tau(w) - c] = [1 - F(w)] [\tau(w) - c].
\] (3)

Let the profit-maximizing choice of the marginal consumer be \(\hat{w}^m : [\underline{w}, \overline{w}] \rightarrow [w, \overline{w}]\),

\[
\hat{w}^m(c) \equiv \arg\max_{w} [1 - F(w)] [\tau(w) - c].
\] (4)

The marginal consumer, one who pays the price equal to his willingness to pay, has wealth \(\hat{w}^m(c)\), and the profit-maximizing quantity is \(1 - F(\hat{w}^m(c))\). Although \(\hat{w}^m(c)\) denotes the marginal consumer, we also call \(\hat{w}^m(c)\) a quantity function when there is no possibility of confusion.

We assume that \(\hat{w}^m(c)\) is single-valued. By the Maximum Theorem \(\hat{w}^m(c)\) is continuous. We further assume that as \(c\) varies over \([\underline{c}, \overline{c}]\), the marginal consumers vary over a proper subset of \([\underline{w}, \overline{w}]\) with \(w < \overline{w} < \overline{w} \).
\[ \hat{w}(\epsilon) < \hat{w}(\tau) < \overline{w} \]. This requires that the variation in wealth is sufficiently large relative to the variation in costs.

**Lemma 1** The quantity function \( \hat{w}^m(c) \) is strictly increasing. If the monopolist has access to all consumers, the monopolist raises its price and sells to less consumers as cost increases.

Proof of Lemma 1: The proof of Lemma 2 below shows that the quantity function \( \hat{w}^m(c) \) must be increasing.\(^2\) Here we show that it is strictly increasing. The profit function (3) is differentiable in \( w \). The first-order derivative of \( \pi(w, c) \) with respect to \( w \) is
\[
- f(w)[\tau(w) - c] + \tau'(w)[1 - F(w)],
\]
and this vanishes at \( w = \hat{w}^m(c) \) since \( f(w) > 0 \) by assumption. Hence, for \( \epsilon > 0 \),
\[
- f(w)[\tau(w) - (c - \epsilon)] + \tau'(w)[1 - F(w)]
\]
is strictly negative at \( w = \hat{w}^m(c) \). Again because \( f(w) > 0 \), lowering \( w \) from \( \hat{w}^m(c) \) must strictly increase profit at cost \( c - \epsilon \). In other words, \( \hat{w}^m(c) \) does not maximize \( \pi(w, c - \epsilon) \). We conclude that \( \hat{w}^m(c) \) must be strictly increasing. \( \blacksquare \)

Lemma 1 does make use of the assumption \( f(w) > 0 \) or equivalently \( F(w) \) is strictly increasing. Indeed, when all consumers are available, the private firm may always sell to more consumers by lowering its price. Hence when cost falls, it must sell to more consumers. Later we will see that if some consumers are supplied by the public sector, there may not be consumers around to accept a price reduction. Price may stop falling even when cost decreases.

**2.3 The public sector and rationing**

There is a public sector which can supply the good to consumers. The public sector has a fixed budget \( B \) but the budget is insufficient to supply to all consumers. Consumers’ wealth information is available to the public regulator, and nonprice rationing will be used to allocate the budget for providing the good to consumers. It is assumed that either the public regulator is unable to observe the cost of provision \( c \) to a consumer, or is disallowed to use any policy that is based on cost.

The public regulator’s rationing rule is defined by the function \( \theta : [\underline{w}, \overline{w}] \rightarrow [0, 1] \). For any given \( w \in [\underline{w}, \overline{w}] \), the public regulator supplies consumers with wealth below \( w \) a total of \( \int_{\underline{w}}^{w} (1 - \theta(x)) f(x) \, dx \) units of the good.

\(^2\)Simply set \( \theta(w) \) to 1 for all \( w \) in Lemma 2.
at zero cost to them, but not the remaining consumers \( \int_w^\infty \theta(x)f(x)\, dx \). The rationing rule \( \theta \) modifies the density \( f \) so that at \( w \) \([1 - \theta(w)]f(w)\) of consumers are provided by the public sector at zero price. Because the rationing rule is based on wealth, the cost \( c \) among rationed consumers is distributed according to \( G \).

Our rationing rule corresponds to random rationing; if there are discrete wealth classes of consumers, for each wealth class, the rationing rule will supply a fraction of consumers.\(^3\)

The rationing scheme can be implemented by waiting times. We can add to the consumer preference specification a new parameter, say \( \delta \), a random variable that is independently distributed according to some distribution, say \( H \). The utility of a consumer is now \( U(w) + 1 - \delta t \) if he gets the good after a delay of \( t \) units of time. The parameter \( \delta \) describes the consumer’s marginal waiting cost. An impatient consumer (one with a high value of \( \delta \)) may decide against the public system if he expects a long delay. By setting the delay \( t \), the regulator determines the fraction of consumers within a wealth group who choose to get the good.\(^4\)

For now suppose that the public sector is the sole provider of the good. Let \( \gamma \equiv \int c\, dG \) denote the expected cost. We let the regulator adopt a utilitarian welfare index equal to total consumer utility. For a given rationing rule \( \theta \), total consumer benefit from consuming the good is \( \int_w^\infty (1 - \theta(x))f(x)\, dx \) as each unit of consumption increases a consumer’s utility by one unit. Expected cost is \( \gamma \int_w^\infty (1 - \theta(x))f(x)\, dx \).

Therefore, the utilitarian welfare index is

\[
V(\theta) \equiv \int_w^\infty U(w)\, dF + \int_w^\infty [1 - \theta(x)]f(x)\, dx. \tag{5}
\]

The rationing rule must satisfy the budget constraint

\[
\gamma \int_w^\infty (1 - \theta(x))f(x)\, dx \leq B, \tag{6}
\]

which says that the expected cost must not exceed the available budget.

When the public sector is the sole supplier, the determination of a rationing rule to maximize (5) subject to (6) is rather trivial. Any rationing rule that exhausts the budget is optimal. There is no strict preference

\(^3\)We consider rationing rules such that the function \( \theta(w)f(w) \) remains integrable, and \( \int_w^\infty \theta(x)f(x)\, dx \) is well-defined for \( w \in [w, \overline{w}] \).

\(^4\)We can restrict the regulator to supply to either all or none of the consumers within a wealth class. A rationing scheme is then a function that maps \([w, \overline{w}]\) to \([0, 1]\). The general rationing function \( \theta : [w, \overline{w}] \to [0, 1] \) is a mixed strategy.
to assign the good to consumers with one wealth level over another.

Again, we have abstracted from redistribution issues. Because the utility function $U$ is strictly concave, under a utilitarian welfare function the regulator might prefer to implement a tax and subsidy scheme to equalize consumers’ wealth (and hence their marginal utility of wealth). Nevertheless, we have adopted the interpretation that the wealth level is measured net of taxes and subsidies. The regulator may not use the allocation of the good to change the wealth distribution. This seems to be a reasonable setup, because general taxation and the public provision of specific goods are often decided separately.

2.4 Interaction between the public and private sectors

We now use the above components to set up interactions between the public and private sectors. Our theory emphasizes that each sector observes a component of the information of consumer $(w, c)$. The rationing scheme in the public sector is a function of wealth $w$, while the pricing scheme in the private sector is a function of cost $c$.

We consider two extensive forms. They differ in the sequence of moves between the public regulator and the private firm. In the simultaneous-move game, there are three stages:

**Stage 1:** Nature determines $(w, c)$; the public sector regulator observes $w$, and the private firm observes $c$.

**Stage 2:** The public sector regulator chooses rationing function $\theta$, and the private firm chooses quantity function $\hat{w}$

**Stage 3:** Consumers supplied by the public sector get the good for free, and consumers not supplied by the public may purchase from private firm at prices set in Stage 2.

In the sequential-move game, there are four stages

**Stage 1:** Nature determines $(w, c)$; the public sector regulator observes $w$, and the private firm observes $c$

**Stage 2:** The public sector regulator chooses rationing function $\theta$

**Stage 3:** The private firm chooses quantity function $\hat{w}$
Stage 4: Consumers supplied by the public sector get the good for free, and consumers not supplied by the public may purchase from private firm at prices set in Stage 3.

In the sequential-move game the public regulator may commit to a rationing scheme, but may not do so in the simultaneous-move game. In both extensive forms, the information structure remains the same.

Whereas in other models, selection means only the private sector serving lower cost consumers, here the allocation of consumers is two dimensional. Even when a consumer’s cost is high, if he is sufficiently wealthy, the private sector may find it profit-maximizing to serve him.

3 Prices in the private market

For a given rationing scheme \( \theta \), the private firm’s profit from selling to consumers with wealth higher than \( w \) is

\[
\pi(w; c, \theta) \equiv \int_w^{w_0} \theta(x)f(x) \, dx \ [\tau(w) - c],
\]

which differs from the expression in (3) in that at \( w \) only a fraction \( \theta(w) \) of consumers with wealth \( w \) would consider buying from the private firm. Let \( \hat{w}(c) \) be the optimal quantities, and \( \hat{\pi}(c) \) the maximum profit:

\[
\hat{w}(c) = \arg \max_w \pi(w; c, \theta),
\]

\[
\hat{\pi}(c) = \pi(w'; c, \theta), w' \in \hat{w}(c).
\]

According to the Maximum Theorem, the correspondence \( \hat{w}(c) \) is upper-semi continuous. An equilibrium (in either extensive form) is a selection from such a correspondence. We present some monotonicity results on the firm’s profit-maximization problem.

Lemma 2 The maximum profit is strictly decreasing in \( c \). Any selection from the profit-maximizing prices, \( \hat{w}(c) = \arg \max_w \pi(w; c, \theta) \), is increasing in \( c \); that is, if \( c_1 < c_2 \), then \( w_1 \leq w_2 \) where \( w_1 \in \hat{w}(c_1) \) and \( w_2 \in \hat{w}(c_2) \).

Proof of Lemma 2: For \( c_1 < c_2 \), let \( w_1 \in \hat{w}(c_1) \) and \( w_2 \in \hat{w}(c_2) \). Because the profit function \( \pi(w; c, \theta) \) in (7) is strictly decreasing in \( c \), we have \( \hat{\pi}(c_1) \geq \pi(w_2; c_1, \theta) > \pi(w_2; c_2, \theta) = \hat{\pi}(c_2) \). Hence, the maximum profit function \( \hat{\pi}(c) \) is strictly decreasing in \( c \).
Next, by the definitions of $w_1$ and $w_2$, we have

\[
\int_{w_1}^{w_2} \theta(w)f(w) \, dw \ [\tau(w_1) - c_1] \geq \int_{w_2}^{w_2} \theta(w)f(w) \, dw \ [\tau(w_2) - c_1]
\]

\[
\int_{w_2}^{w_2} \theta(w)f(w) \, dw \ [\tau(w_2) - c_2] \geq \int_{w_1}^{w_1} \theta(w)f(w) \, dw \ [\tau(w_1) - c_2].
\]

Adding these two inequalities yields

\[
\int_{w_1}^{w_2} \theta(w)f(w) \, dw \ [c_2 - c_1] \geq 0,
\]

which says that $w_2$ must be at least $w_1$ since $\theta(w) \geq 0$. ■

The profit-maximizing price may not be strictly increasing in cost $c$, although the maximum profit is strictly decreasing. The reason is that for some rationing rules, the profit-maximizing prices at two different cost levels may be identical. For example, suppose that the rationing rule specifies that $\theta(w) = 0$ for $w < \bar{w}$, and $\theta(w) = 1$ for $w > \bar{w}$. This scheme supplies those consumers (at zero price) if and only if their wealth is below a threshold $\bar{w}$. The firm may not reduce its price as cost falls below a certain level. At a low cost, the firm already may sell to all available consumers, setting the price at the willingness to pay $\tau(\bar{w})$. The optimal price will not reduce further even when cost falls. Figure 1 illustrates this. There, two quantity functions are graphed. First, the quantity function $\hat{w}^m(c)$ is the profit maximizing quantity function when the firm may sell to all consumers, while the quantity function $\hat{w}(c)$ maximizes profit when consumers with wealth less than $\bar{w}$ are supplied by the public sector (at zero price). The quantity function $\hat{w}(c)$ coincides with $\hat{w}^m(c)$ for cost levels above a threshold, and it becomes the horizontal dotted line when the cost falls below that threshold.

For some rationing rules, there may be multiple points that maximize profit. For example, suppose that $\theta(w) = 0$ for $w \in [w_1, w_2]$ where $w < w_1 < w_2 < \bar{w}$, and $\theta(w) = 1$ otherwise. The public sector supplies only to consumers with wealth levels in a medium range. Figure 2 illustrates the density of consumers available to the private firm. The profit maximizing quantity function is illustrated in Figure 3. For $c < c_1$ or $c > c_2$, the profit-maximizing quantity is unique. For $c \in (c_1, c_2)$, the price remains constant. As the cost falls below $c_2$, the firm does not lower its price because all consumers with wealth in $[w_1, w_2]$ are supplied by the public
sector. At cost $c_1$, the firm makes equal amounts of profit whether it charges a price equal to $\tau(w_2)$ or a price $\tau(w_0)$. The profit from selling to consumers with $w$ between $w_0$ and $w_1$ and those with $w > w_2$ at a lower price $\tau(w_0)$ is exactly the same as selling only to those with wealth above $w_2$ at the higher price $\tau(w_2)$. Notice that the value of $w_0$ must be strictly below $w_1$.

A selection from the correspondence $\hat{w}(c)$ need not be continuous. Nevertheless, because it must be increasing, any point of discontinuity of $\hat{w}(c)$ must be an upward jump, as in Figure 3. An equilibrium is a selection of the profit-maximizing quantity correspondence. By a slight abuse of notation, we denote
such a selection by \( \hat{w} : [c, \bar{c}] \to [w, w'] \). From Lemma 2, we only need to consider those quantity functions \( \hat{w} : [c, \bar{c}] \to [w, w'] \) that are increasing.

For a given quantity function \( \hat{w}(c) \) and the corresponding price function \( \tau(\hat{w}(c)) \), consumer \((w, c)\) buys from the private firm if and only if \( w \geq \hat{w}(c) \). Let this set of consumers be denoted by \( \Omega \equiv \{(w, c) : w \geq \hat{w}(c)\} \). In Figure 3, this is the set above the graph of \( \hat{w}(c) \). If we integrate the utilities of consumers in \( \Omega \), we obtain the total consumer benefit.

It is more convenient to view the set \( \Omega \) as one that is indexed by a function \( \hat{c} : [w, w'] \to [c, \bar{c}] \) that is like an “inverse” of \( \hat{w} \). Define \( \hat{c}(w) = \sup \{c : w \geq \hat{w}(c)\} \); if there is no \( c \in [c, \bar{c}] \) such that \( w \geq \hat{w}(c) \), set \( \hat{c}(w) = \bar{c} \). While the function \( \hat{w} \) gives the wealth of the marginal consumer in terms of his cost, the function \( \hat{c} \) gives the threshold cost level below which a consumer with wealth \( w \) will buy from the firm at price \( \tau(\hat{w}(c)) \).

Whenever \( \hat{w} \) is strictly increasing and continuous, the function \( \hat{c} \) is its inverse. When \( \hat{w} \) is constant, then \( \hat{c} \) exhibits a discontinuity. When \( \hat{w} \) is discontinuous, then \( \hat{c} \) becomes constant. Clearly \( \hat{c}(w) \) is increasing in \( w \). The set \( \Omega' \equiv \{(w, c) : c \leq \hat{c}(w)\} \) differs from \( \Omega \) at most for a set of measure zero. Functions \( \hat{w} \) and \( \hat{c} \) are two equivalent ways to keep track of consumer types who purchase from the private firm.\(^5\)

---

\(^5\)If we begin with the function \( \hat{w} \), then the double integral for calculating the total benefit for consumers in \( \Omega \) is with respect to \( w \) first, and \( c \) second. If we use the equivalent function \( \hat{c} \), we just reverse the order of the integration.
Given a quantity function \( \hat{w} \) (and its equivalent \( \hat{c} \)), if the public regulator chooses a rationing scheme \( \theta \), the welfare index \( V(\theta) \) is

\[
\int \theta(w)f(w) \left[ \int_{\mathcal{E}(w)} \{U(w - \tau(\hat{w}(c))) + 1\} g(c) \, dc + \int_{\mathcal{F}(w)} U(w) g(c) \, dc \right] \, dw + \int \theta(w)[U(w) + 1] \, dw
\]

(10)

In this expression, the first term inside the square brackets is the utility of consumers who buy from the private firm while the other terms refer to utilities of consumers who are either given the good for free or refuse to buy from the private firm after having been rationed out. Consumer \( (w, c) \) pays the price \( \tau(\hat{w}(c)) \) when he buys from the private firm, and this price is always lower than his willingness to pay \( \tau(w) \). The welfare index can be simplified to

\[
V(\theta) = \int \theta(w)f(w) \left[ \int_{\mathcal{E}(w)} \{U(w - \tau(\hat{w}(c))) + 1 - U(w)\} g(c) \, dc + \int_{\mathcal{F}(w)} U(w) + (1 - \theta(w)) \, f(w) \, dw, \right.
\]

(11)

where the first term is the expected inframarginal surplus consumers enjoy when they purchase from the private market, and the second term is the consumer surplus from the public supply.

### 4 Equilibrium rationing and prices in the simultaneous-move game

An equilibrium is a pair of rationing and quantity schemes \( (\theta, \hat{w}) \) such that \( \theta \) maximizes the welfare index (10) subject to the budget constraint (6) given a quantity scheme \( \hat{w} \), and \( \hat{w} \) maximizes profit (7) for every \( c \) given \( \theta \); that is, the rationing and pricing schemes are mutual best responses against each other.

For the analysis, we let the regulator pick the net density of consumers that will be made available to the private firm \( \theta f \), and impose the requirement that \( 0 \leq \theta f \leq f \). The welfare index (11) is linear in \( \theta f \), and for each \( w \) its first-order derivative with respect to \( \theta f \) is

\[
\frac{\partial V}{\partial \theta f} = \int_{\mathcal{E}(w)} \{U(w - \tau(\hat{w}(c))) + 1\} g(c) \, dc + \int_{\mathcal{F}(w)} U(w) g(c) \, dc - [U(w) + 1]
\]

\[
= \int_{\mathcal{E}(w)} \{U(w - \tau(\hat{w}(c))) + 1 - U(w)\} g(c) \, dc - 1.
\]

(12)

This expression measures the welfare tradeoff between one unit of public provision and letting one unit of consumer to the private market. In the private market, consumer \( (w, c) \) faces the price \( \tau(\hat{w}(c)) \). He will buy from the private market at the price \( \tau(\hat{w}(c)) \) if his wealth is above \( \hat{w}(c) \) (or equivalently if his cost is below
\( \hat{c}(w) \). The term inside the integral in (12) is the expected incremental surplus from such transactions for a consumer with wealth \( w \). Against this, the welfare index is reduced by 1, the incremental utility of the good if it is supplied by the public sector at zero charge. The following key lemma establishes a monotonicity in the regulator’s preferences.

**Lemma 3** The first-order derivative \( \frac{\partial V}{\partial \theta f} \) (12) is increasing in \( w \). It is strictly increasing in \( w \in [w_1, w_2] \) unless \( \hat{c}(w) = \zeta \) for each such \( w \).

Proof of Lemma 3: Consider \( w_1 \) and \( w_2 \) with \( w_1 < w_2 \). Evaluating (12) at \( w_1 \) and \( w_2 \) and then taking the difference, we have

\[
\frac{\partial V}{\partial \theta f} \bigg|_{w=w_2} - \frac{\partial V}{\partial \theta f} \bigg|_{w=w_1} = \int_{\zeta} \{ [U(w_2 - \tau(\hat{w}(c))) - U(w_1 - \tau(\hat{w}(c)))] - [U(w_2) - U(w_1)] \} g(c) \, dc \\
+ \int_{\hat{c}(w_1)} \{ U(w_2 - \tau(\hat{w}(c))) + 1 - U(w_2) \} g(c) \, dc \geq 0 \tag{13}
\]

The inequality in (13) follows from the concavity of \( U \), and \( \hat{c} \) being increasing. Finally, (13) is zero if and only if \( \hat{c}(w_1) = \hat{c}(w_2) = \zeta \). ■

Lemma 3 says that the regulator’s tradeoff between supplying the good to a consumer and rationing him favors rationing as the consumer’s wealth level increases. This is the fundamental principle in our model. The price in the private sector depends only on cost \( c \). Consumer \((w, c)\) gets more surplus from a trade at price \( \tau(\hat{w}(c)) \) as \( w \) increases: \( U(w - \tau(\hat{w}(c))) + 1 - U(w) \) is increasing in \( w \).

The last part of Lemma 3 says that the derivative (13) at \( w \) remains at a constant if and only if consumers with wealth lower than \( w \) do not get to purchase from the private market. Again, if consumer \((w, c)\) gets to purchase from the private market for some levels of cost, then as \( w \) increases, the incremental surplus increases. The derivative (13) then must be strictly positive. Figure 4 illustrates a situation where the quantity function \( \hat{w} \) becomes a constant as cost falls below \( \hat{c} \), and the price remains at \( \tau(\hat{w}) \). Consumers with wealth below \( \hat{w} \) do not buy, and the integral in (13) vanishes.

---

6This preference is not due to an equity concern because the regulator is unable to use any taxes and subsidies to redistribute income to reduce variations in the marginal utilities of wealth.
Against a quantity function \( \hat{w}(c) \) (and the corresponding \( \hat{c}(w) \)), the public regulator chooses \( \theta_f \) to maximize (10) subject to the budget constraint (6). Using pointwise optimization, we consider the Lagrangean

\[
\theta(w)f(w) \left[ \int_{\hat{c}(w)}^{\hat{c}(w)} \{U(w - \tau(\hat{w}(c))) + 1\} g(c) \, dc \right] + \left[ 1 - \theta(w) \right] f(w) [U(w) + 1] - \lambda [1 - \theta(x)] f(x) - B]
\]

where \( \lambda \) is the multiplier. The first-order derivative of the Lagrangean with respect to \( \theta_f \) is

\[
\frac{\partial V}{\partial \theta_f} + \lambda \gamma = \int_{\hat{c}(w)}^{\hat{c}(w)} \{U(w - \tau(\hat{w}(c))) + 1 - U(w)\} g(c) \, dc - 1 + \lambda \gamma.
\]

(14)

From Lemma 3, the first-order derivative of the Lagrangean is increasing in \( w \), and increasing strictly whenever some consumers with wealth less than \( w \) purchase from the private market.

**Lemma 4** In an equilibrium, the public sector rations consumers with wealth above a threshold \( \bar{w} \). That is, in an equilibrium there is \( \bar{w} < \bar{w} \) such that \( \theta(w) = 1 \) for \( w > \bar{w} \).

Proof of Lemma 4: Because of the limited budget, the public regulator must leave some consumers to the private sector, so that \( \theta(w) > 0 \) for some \( w \). Because the rationing rule depends only on wealth, some consumers must purchase from the private market. Let \( \bar{w} = \inf\{w : \theta(w) > 0\} \), and for some \( w > \bar{w}, \hat{c}(w) \) must be higher than \( \hat{c} \). By Lemma 3, the first-order derivative of the Lagrangean with respect to \( \theta_f \) must be strictly increasing in \( w \) for \( w > \bar{w} \). If \( \theta(w) > 0 \), then the first-order derivative (14) must be nonnegative,
and for any $w' > w$, the value of (14) must be strictly positive, and $\theta(w') = 1$. ■

Lemma 4 follows from the monotonicity of the regulator’s preferences. If it is optimal for the regulator to ration a consumer at some wealth level, then it must also be optimal to ration all consumers with higher wealth. In any equilibrium, there must exist $\tilde{w}$ such that $\theta(w) = 1$ if $w > \tilde{w}$. Lemma 4 does not assert that there is a unique equilibrium. Nor does it say that in an equilibrium, $\theta(w) < 1$ for $w < \tilde{w}$.

**Lemma 5** In an equilibrium, the private firm sets a constant price for costs below a threshold $\tilde{c}$. That is, $\hat{w}(c)$ is constant for $c < \tilde{c}$.

Proof of Lemma 5: If $\hat{w}(c)$ is an equilibrium quantity function, it is increasing. Suppose that the Lemma is false. That is, suppose that for some $\tilde{c} > \underline{c}$, $\hat{w}(c)$ is strictly increasing for all $c < \tilde{c}$. Then $\underline{w} < \hat{w}(c) < \overline{w}$, for $c < \tilde{c}$, which implies that $\tilde{c}(w) > \underline{c}$ for all $w$. By Lemma 3, the first-order derivative (14) must be strictly increasing at any $w$. This means that the set of $w$ at which the first-order derivative (14) vanishes is a single point. Hence, in this equilibrium, the public supplies the good to consumers with wealth below a certain threshold, and $\theta(w) = 0$ for $w < \tilde{w}$. Against such a rationing scheme, the quantity function $\hat{w}(c)$ that is strictly increasing for $c < \tilde{c}$ cannot be optimal, because switching $\hat{w}(c) < \tilde{w}$ to $\hat{w}(c) = \tilde{w}$ yields strictly higher profit. ■

We already know that an equilibrium quantity function is increasing. Lemma 5 says that it cannot be strictly increasing as cost decreases. The reason is that such a strictly increasing quantity function will make the public regulator supply to consumers with low wealth. But a strictly increasing quantity function when costs are low is not a best response against that; because there are no consumers with low willingness to pay to accept the price reduction, the private firm might as well leave the price unchanged even as cost falls. Again, Lemma 5 does not assert that there is a unique equilibrium cost threshold.

The last two lemmas establish the form of an equilibrium. The public sector must ration consumers with high wealth, and the private firm must not sell to consumers with wealth below a threshold. The basic economic principle is the following. The regulator’s preferences are linear in the rationing rule. It must be a “corner” solution. Because of the limited budget, and the higher surplus for wealthy consumers in the private market, rationing of wealthy consumers must happen in an equilibrium. The public sector supplying
the less wealthy consumers with the goods makes these consumers unavailable to the private firm. As cost decreases, the private firm also has a “corner” solution in its profit-maximizing choice of prices. Prices will become constant even as costs drop further since there may not so few or even zero consumers with lower wealth and willingness to pay to take any price reduction.

We now present an equilibrium.

**Proposition 1** The following is an equilibrium in the simultaneous-move game. The public regulator rations all consumers whose wealth is above a threshold \( \bar{w}^s \) and supplies all consumers with wealth below \( \bar{w}^s \): \( \theta(w) = 1, w > \bar{w}^s \) and \( \theta(w) = 0, w < \bar{w}^s \). The value of \( \bar{w}^s \) exhausts the budget and is given by \( F(\bar{w}^s)\gamma = B \). The private firm sets a price equal to \( \tau(\hat{w}^m(c)) \) for \( c > \bar{c}^s \) where \( \hat{w}^m(\bar{c}^s) = \bar{w}^s \), and a price equal to \( \tau(\bar{w}^s) \) for \( c < \bar{c}^s \).

Proof of Proposition 1: We verify that the strategies in the proposition form an equilibrium. Given the public regulator’s rationing scheme, it is obvious that it is optimal for the firm to set prices at \( \tau(\hat{w}^m(c)) \) when \( c > \bar{c}^s \), and at \( \tau(\bar{w}^s) \) when \( c < \bar{c}^s \). Given this quantity function \( \hat{w} \), we have the equivalent function \( \bar{c} \) where \( \bar{c}(w) = c \) for \( w < \bar{w}^s \), and \( \bar{c}(w) > c \) for \( w > \bar{w}^s \). For (14) we set the multiplier \( \lambda \) to \( 1/\gamma \). Then the first-order derivative (14) is zero for \( w < \bar{w}^s \) and strictly positive for \( w > \bar{w}^s \). Moreover, the budget constraint holds as an equality. Hence the rationing scheme defined in the proposition is optimal. The strategies form an equilibrium. □

The equilibrium in Proposition 1 is illustrated in Figure 4. The private firm sets its prices like it is the monopoly in the market except that it has no access to consumers with wealth below \( \bar{w}^s \), so the prices in the monopoly quantity schedule \( \hat{w}^m(\bar{c}) \) will stop falling at \( \tau(\bar{w}^s) \) even as cost falls below \( \bar{c} \). The regulator attempts to maximize consumer surplus given this quantity schedule. The regulator’s preferences always favor rationing the wealthy consumers.

Surprisingly, the equilibrium in Proposition 1 is not the only equilibrium in the simultaneous-move game. Lemmas 4 and 5 do not pin down the regulator’s or the firm’s strategies. A pricing schedule that becomes constant for low costs may still be consistent with the public regulator rationing some consumers with very low wealth levels. The following is another equilibrium in the simultaneous-move game. Let \( \epsilon > 0 \) and \( \delta > 0 \).
be both small numbers. The regulator’s strategies is the following rationing scheme:

\[ \theta(w) = 1 \text{ for } w < w < w + \epsilon \]
\[ \theta(w) = 0 \text{ for } w + \epsilon < w < \bar{w}^s + \delta \]
\[ \theta(w) = 1 \text{ for } \bar{w}^s + \delta < w < \bar{w}. \]

In this rationing rule the regulator shifts some resources from those with wealth just above the lowest value \( w \) to those consumers with wealth just above \( \bar{w}^s \), the equilibrium threshold in Proposition 1. The values of \( \epsilon \) and \( \delta \) can be so chosen that the new rationing scheme satisfies the budget. Against this rationing scheme, the private firm sets a quantity function equal to \( \tilde{w}^m(c) \) for \( c > \bar{c}^s + \eta \) and \( \tilde{w}^m(\bar{c}^s + \eta) \) for \( c < \bar{c}^s + \eta \), where \( \bar{c}^s \) is the cost threshold in Proposition 1, and \( \tilde{w}^m(\bar{c}^s + \eta) = \bar{w}^s + \delta \).

In this equilibrium, the regulator rations a little less wealthy consumers than those in Proposition 1. In addition, it rations a very small set of consumers who have the lowest wealth, near \( w \), and make them available to the private market. The private firm’s price still falls from \( \tau(\tilde{w}^m(\bar{c})) \) as cost reduces from \( \bar{c} \), but it will not fall all the way to \( \tau(\bar{w}^s) \). Moreover, it does not sell to those consumers with wealth near \( w \). To do so requires a large price reduction, and because there are so few of these consumers it is not profit-maximizing. Both Lemmas 4 and 5 continue to hold. Figure 4 continues to depict the equilibrium quantity function if we just move \( \bar{w}^s \) to \( \bar{w}^s + \delta \), and the horizontal line there up a bit.

A continuum of equilibria can be constructed in a similar fashion. As long as the private firm does not find it profit-maximizing to reduce price in order to sell to consumers with low willingness to pay, a quantity function like the one in Figure 4 remains a best response. In all these equilibria the public regulator rations some consumers with low wealth, but must ration all consumers with wealth above a threshold.

The equilibrium in Proposition 1 is focal. This is the one that achieves the highest welfare index for the regulator. This is because it has the widest range of price reduction as cost decreases. Any equilibrium different from the one in Proposition 1 would have fewer transactions in the private market, as the next Proposition shows.

**Proposition 2** The equilibrium in Proposition 1 achieves the highest equilibrium welfare for the regulator.
In any other equilibrium, the public regulator sets $\theta(w) = 1$, for $w > \tilde{w}^c$, where $\tilde{w}^c > \tilde{w}^a$ (defined by $F(\tilde{w}^a)\gamma = B$ as in Proposition 1) and the firm sets a price equal to $\tau(\tilde{w}^m(c))$ for $c > \tilde{c}^c$, and a price equal to $\tau(\tilde{w}^c)$ for $c < \tilde{c}^c$ where $\tilde{w}^m(\tilde{c}^c) = \tilde{w}^c$.

Proof of Proposition 2: In any equilibrium, the budget constraint $\gamma \int_{w}^{\tilde{w}^c} (1 - \theta(x)) f(x) dx \leq B$ must hold as an equality. The equilibrium in Proposition 1 supplies those with wealth between $w$ and $\tilde{w}^a$, where $F(\tilde{w}^a)\gamma = B$, so $\theta(w) = 0$ for $w < \tilde{w}^a$ and $\theta(w) = 1$ for all $w > \tilde{w}^a$. Consider any other equilibrium. Here, the regulator must supply to some consumers with wealth below $\tilde{w}^a$ because the budget constraint must hold. Hence, for this equilibrium the threshold $\tilde{w}^c$ at which $\theta(w) = 1$ for all $w > \tilde{w}^c$ must be strictly higher than $\tilde{w}^a$.

Let $\tilde{c}^c$ be defined by $\tilde{w}^m(\tilde{c}^c) = \tilde{w}^c$. Clearly, $\tilde{c}^c > \tilde{c}^a$ because $\tilde{w}^m$ is strictly increasing and $\tilde{w}^c > \tilde{w}^a$. We now show that the firm’s equilibrium price is $\tau(\tilde{w}^m(\tilde{c}^c))$ for $c < \tilde{c}^c$. At $w < \tilde{w}^c$, the first-order derivative of the Lagrangean (14) must be nonnegative; if that derivative was negative, then $\theta(w) = 0$, which would violate the budget constraint. Because $\theta(w) = 1$ for all $w > \tilde{w}^c$, the value of (14) is positive for all $w > \tilde{w}^c$. Therefore, by Lemma 3, for $w < \tilde{w}^c$ the value of (14) must be exactly zero and $\tilde{c}(w) = c$ for $w < \tilde{w}^c$. Because $\theta(w) = 1$ for all $w > \tilde{w}^c$ the equilibrium quantity must be $\tilde{w}^m(c)$ for $c > \tilde{c}^c$, and remains constant at $\tilde{w}^m(\tilde{c}^c)$ for $c < \tilde{c}^c$.

From $\tilde{c}^c > \tilde{c}^c$ and $\tilde{w}^c > \tilde{w}^a$, by comparing the values of (10) across the equilibria in Proposition 1 and the alternative, we conclude that the regulator’s payoff is higher in the equilibrium in Proposition 1. ■

We have assumed a monopolistic private sector. The extension to an imperfectly competitive sector poses no conceptual problem. Because of the homogeneous good assumption, we consider a Cournot model. Let there be $N$ firms in the private sector. Given any rationing scheme $\theta$, let each firm choose a quantity function $\hat{q}_i(c)$, where $i = 1, ..., N$. The total supply is $q(c) = \sum_{i=1}^{N} q_i(c)$. For the market to clear the marginal consumer is $\hat{w}(c)$ where $\int_{\hat{w}(c)}^{\tilde{w}(c)} \theta(w) f(w) dw = q(c)$, and the price in the private sector is $\tau(\hat{w}(c))$. All results derived above continue to hold for any given number of firms in the private sector.

Next, we can extend our model to the case of a perfectly competitive private sector. Here, we simply
let the price in the private sector be marginal cost: \( \tau(c) = c \). Given this pricing function, the corresponding quantity function \( \hat{\omega}(c) \) is implicitly defined by \( U(\hat{\omega} - c) + 1 = U(\hat{\omega}) \). Lemma 3 can be applied on this quantity function. Because the perfectly competitive quantity function is strictly increasing, the derivative (14) is strictly increasing for all values of \( w \). Lemma 4 continues to hold. In sum we have the following result.

**Corollary 1** If the private market is perfectly competitive so that firms there charge marginal costs, the equilibrium in Proposition 1 is the unique equilibrium.

Discussions on commitment.

5 Equilibrium rationing and prices in the sequential-move game

We now turn to the model where the regulator can commit to a rationing scheme before the private market sets its prices. We begin with a model where there are only two wealth classes.

Suppose that consumers’ wealth is either low or high, \( w_1 \) or \( w_2 \), with \( w_1 < w_2 \). Let \( m_i \) be the mass of consumers with wealth \( w_i, i = 1, 2 \). We maintain the assumption that cost \( c \) is distributed according to \( G \). Furthermore, we assume that the hazard rate \( G(c)/g(c) \) is increasing. Let \( \tau_i \) be the willingness to pay: \( U(w_i - \tau_i) + 1 = U(w_i), i = 1, 2 \). A rationing rule is given by \( (\theta_1, \theta_2) \) where the regulator supplies \( (1 - \theta_i)m_i \) consumers with wealth \( w_i \), and the private firm has access to \( \theta_i m_i \) of consumers with wealth \( w_i, i = 1, 2 \).

Given the rationing rule, we write down the firm’s equilibrium price decisions. The firm will set its price to either \( \tau_1 \) or \( \tau_2 \). Clearly at any cost above \( \tau_1 \), the firm must set its price at \( \tau_2 \), selling only to the more wealthy consumers. Suppose that the cost is below \( \tau_1 \). The firm may set a lower price \( \tau_1 \) selling to both classes of consumers or a higher price \( \tau_2 \) selling only to wealthy consumers; these profits are respectively:

\[
\pi(\tau_1; c \leq \tau_1) \equiv (m_1 \theta_1 + m_2 \theta_2) [\tau_1 - c]
\]

\[
\pi(\tau_2; c \leq \tau_1) \equiv m_2 \theta_2 [\tau_2 - c].
\] (15)

The profits in (15) and (16) are linear in \( c \), and (15) decreases in \( c \) at a faster rate than (16). The firm sets the higher price if and only if \( \pi(\tau_2; c \leq \tau_1) \geq \pi(\tau_1; c \leq \tau_1) \).
Given \((\theta_1, \theta_2)\) there may be a cost level \(c_1\) in \([c, \bar{c}]\) such that \(\pi(\tau_1; c_1) = \pi(\tau_2; c_1)\). This is illustrated in Figure 5, and the value of \(c_1\) is given by

\[
c_1 = \tau_1 - \frac{m_2 \theta_2}{m_1 \theta_1} (\tau_2 - \tau_1). \tag{17}
\]

If there are so few consumers with low wealth, the firm may not reduce its price from \(\tau_2\) to \(\tau_1\) even when cost falls to \(c\); in that case, the price will be \(\tau_2\) for all costs. For a given rationing scheme, the value of \(c_1\) in (17) completely defines the firm’s continuation equilibrium pricing strategy. We assume that if the firm has access to all consumers, the firm reduces the price at some interior cost threshold. That is, for \(\theta_1 = \theta_2 = 1\), (17) there is \(c_m\) in the interior of \([c, \bar{c}]\) where

\[
c_m = \tau_1 - \frac{m_2 \theta_2}{m_1 \theta_1} (\tau_2 - \tau_1). \tag{18}
\]

We introduce a new notation \(\beta \equiv B/\gamma\); because \(B\) denotes the budget available to the regulator, the term \(\beta\) is the number of consumers whose total expected cost will exhaust the budget. Adapting from the welfare expression (11) in the previous section, and noting that \(U(w_i - \tau_i) + 1 = U(w_i), i = 1, 2\), we write the regulator’s welfare index as

\[
V(\theta_2, c_1) \equiv [m_1 U(w_1) + m_2 U(w_2) + \beta] + m_2 \theta_2 \int_c^{c_1} [U(w_2 - \tau_1) + 1 - U(w_2)] g(c) dc, \tag{19}
\]

where \(c_1 \geq c\) characterizes the firm’s continuation equilibrium price strategy.
The equilibrium in the sequential-move game is a rationing rule \((\theta_1, \theta_2)\) and the continuation equilibrium price strategy \(c_1\) which maximize (19) subject to the cost threshold (17), the budget constraint

\[
m_1(1-\theta_1) + m_2(1-\theta_2) \leq \beta = \frac{B}{\gamma} (< m_1 + m_2)
\]

and the boundary conditions \(c \leq c_1\), and \(0 \leq \theta_i \leq 1, i = 1, 2\).

**Proposition 3** In the equilibrium of the sequential-move game, the regulator rations consumers in each wealth class: \(\theta_1 > 0\) and \(\theta_2 > 0\), while the firm charges a lower price when the consumer’s cost is below a threshold: \(p(c) = \tau_1\) whenever \(c < c_1^*\), where \(c < c_1^* < \tau_1\). If the budget \(\beta\) is sufficiently large, the value of \(c_1^*\) is given by the unique solution of

\[
\frac{G(c_1^*)}{g(c_1^*)} = \frac{(\tau_1 - c_1^*)(\tau_2 - c_1^*)}{\tau_2 - \tau_1}
\]

and \(\theta_1 < 1\) and \(\theta_2 < 1\); the regulator will supply some consumers in each wealth class. If the budget is small, either \(\theta_1\) or \(\theta_2\) may be equal to 1, and the budget constraint then can be used to solve for the optimal rationing rule; the regulator may ration an entire wealth class.

Proof of Proposition 3: Because all terms in square brackets in the objective function (19) are constant, we alternatively can write the objective function as \(m_2\theta_2 G(c_1)\).

We first claim that the boundary conditions \(c \leq c_1\), and \(0 \leq \theta_i\) do not bind. If \(c_1 = c\) at a solution, then the optimized value is 0 because \(G(c) = 0\). We show that this cannot be optimal. Consider a rationing rule where \(\theta_1 = \theta_2 = k\). This policy satisfies the budget constraint (20) for some \(0 < k < 1\). Moreover, from (17) and (18), we have \(c_1 = c_m > c\) by assumption. Therefore this ration rule \(\theta_1 = \theta_2 = k\) is feasible, and yields a payoff \(m_2kG(c_m) > 0\). This implies that \(c_1 = c\) is not a solution, and \(c_1 \geq c\) does not bind. By the same argument, \(\theta_2 \geq 0\) does not bind either. Because \(c_1 > c\), it follows from (17) that \(\theta_1\) must be bounded away from 0, and hence \(\theta_1 \geq 0\) does not bind.

Now we characterize the solution. For the time being, ignore the remaining boundary conditions \(\theta_i \leq 1\). Rewrite the budget constraint (20) as \(m_1\theta_1 + m_2\theta_2 \geq m_1 + m_2 - \beta \equiv K > 0\); clearly the budget constraint must bind at a solution. From constraint (17), we have \(m_2\theta_2(\tau_2 - \tau_1) = m_1\theta_1(\tau_1 - c_1)\) which yields

\[
m_1\theta_1 = m_2\theta_2 \frac{\tau_2 - \tau_1}{\tau_1 - c_1}.
\]
Substituting this into the modified budget constraint, we can solve for \( m_2 \theta_2 \):

\[
m_2 \theta_2 = K \frac{\tau_1 - c_1}{\tau_2 - c_1}.
\] (22)

We next substitute (22) into the objective function \( m_2 \theta_2 G(c_1) \). The constrained maximization problem (with the boundary conditions \( \theta_1 \leq 1 \) omitted) is the same as the unconstrained maximization problem:

\[
\max_{c_1} K \frac{\tau_1 - c_1}{\tau_2 - c_1} G(c_1).
\]

Ignoring the parameter \( K \), we obtain the first-order derivative, and after simplification it is:

\[
\frac{\tau_1 - c_1}{\tau_2 - c_1} g(c_1) - G(c_1) \frac{\tau_2 - \tau_1}{\tau_2 - c_1} = \frac{\tau_1 - c_1}{\tau_2 - c_1} \left[ (\tau_1 - c_1) - \frac{G(c_1)}{g(c_1)} \frac{\tau_2 - \tau_1}{\tau_2 - c_1} \right].
\]

Setting the first-order derivative to zero, we have

\[
\frac{G(c_1)}{g(c_1)} = \frac{(\tau_1 - c_1)(\tau_2 - c_1)}{\tau_2 - \tau_1}.
\] (23)

The left-hand side of (23) is increasing in \( c_1 \). For \( c_1 \) between \( \underline{c} \) and \( \tau_1 \), the right-hand side is decreasing. At \( c_1 = \underline{c} \), the left-hand side of (23) is zero, while the right-hand side of (23) is strictly positive. At \( c_1 = \tau_1 \), the left-hand side of (23) is strictly positive, while the right-hand of (23) is zero. Therefore, there exists a unique \( c_1^* \) strictly between \( \underline{c} \) and \( \tau_1 \) that satisfies (23).

To recover \( \theta_i m_i \), we use (21) and (23) to get: \( m_2 \theta_2 = m_1 \theta_1 \frac{G(c_1^*)}{g(c_1^*) (\tau_2 - c_1^*)} \), which, together with the budget constraint (20), can be used to solve for the values of \( m_1 \theta_1 \) and \( m_2 \theta_2 \):

\[
m_1 \theta_1 = \frac{m_1 + m_2 - \beta}{1 + \frac{G(c_1^*)}{g(c_1^*)} \frac{1}{\tau_2 - c_1^*}} \quad \text{and} \quad m_2 \theta_2 = \frac{m_1 + m_2 - \beta}{1 + \frac{g(c_1^*)}{G(c_1^*)} (\tau_2 - c_1^*)}.
\] (24)

If \( \beta \) is sufficiently large, the right-hand side values in (24) will be less than \( m_1 \) and \( m_2 \). The omitted boundary conditions \( \theta_i \leq 1 \) are satisfied. Otherwise, if \( \beta \) is small, one of the right-hand side values in (24) will be more than \( m_1 \) and \( m_2 \). In this case, a boundary condition binds.
References


