



Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

Journal of Health Economics

journal homepage: [www.elsevier.com/locate/econbase](http://www.elsevier.com/locate/econbase)



## Optimal public rationing and price response

Simona Grassi<sup>a,\*</sup>, Ching-to Albert Ma<sup>b</sup>

<sup>a</sup> Faculty of Business and Economics, Department of Economics and Econometrics and Institut d'Economie et de Management de la Santé, University of Lausanne, Building Internef, CH-1015 Lausanne, Switzerland

<sup>b</sup> Department of Economics, Boston University, United States

### ARTICLE INFO

#### Article history:

Received 13 January 2010  
Received in revised form 2 May 2011  
Accepted 22 August 2011  
Available online xxx

#### JEL classification:

D61  
H42  
H44  
I11

#### Keywords:

Rationing  
Price response  
Means-testing  
Cost effectiveness

### ABSTRACT

We study optimal public health care rationing and private sector price responses. Consumers differ in their wealth and illness severity (defined as treatment cost). Due to a limited budget, some consumers must be rationed. Rationed consumers may purchase from a monopolistic private market. We consider two information regimes. In the first, the public supplier rations consumers according to their wealth information (means testing). In equilibrium, the public supplier must ration both rich and poor consumers. Rationing some poor consumers implements price reduction in the private market. In the second information regime, the public supplier rations consumers according to consumers' wealth and cost information. In equilibrium, consumers are allocated the good if and only if their costs are below a threshold (cost effectiveness). Rationing based on cost results in higher equilibrium consumer surplus than rationing based on wealth.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Public supply of health care services is very common. Because of limited budgets, free health care for all is infeasible. The limited public supply is usually distributed by nonprice rationing. Rationed consumers often can turn to the private market and purchase at their own expense. In this paper we study optimal public rationing policies and price responses in the private market.

The design of rationing policies should take into account private market reactions; otherwise, unintended consequences may arise. For example, expansions in Medicaid and similar programs for the indigent may actually reduce consumers' purchases in the private market, a phenomenon called "crowd out" (see Cutler and Gruber, 1996; Gruber and Simon, 2008). The literature has not investigated the mechanism behind it. By explicitly considering private market responses, we exhibit a mechanism for crowd out.

Two mechanisms are often used for distributing public health services. The first is *means testing*, supply based on wealth or income. For example, Medicaid in the United States and many state programs target the indigent. The second is *cost effectiveness*, supply based on a ratio of benefit to cost. For example, in most

European countries and Canada, a medical service is covered by national insurance only if its benefit–cost ratio is higher than a threshold. Cost-effectiveness rationing and means-testing rationing yield different price responses in the private market. Crowd out can be avoided under cost-effectiveness rationing. Furthermore, we show that optimal cost-effectiveness rationing results in higher equilibrium consumer utility than means testing.

In our model, consumers are heterogeneous in two dimensions: they have different wealth levels, and they have different illness severities. Wealth heterogeneity is a natural assumption, and it means that rich consumers are more willing to pay for services than poor consumers. Illness severity heterogeneity is also natural. Each illness severity is associated with a treatment cost and a benefit. For convenience, we simply let severity be the treatment cost. Consumers' treatment benefits are increasing in severity, but at a decreasing rate. Our assumption on cost and benefit is similar to common ones in the health economics literature (see for example, Ellis, 1998).

We consider rationing in two information regimes. In the first, rationing is based on consumers' wealth; means-testing rationing policies belong to this regime. In the second, rationing is based on consumers' wealth and cost; cost-effectiveness rationing policies belong to this regime. In each regime, we study equilibria of the following extensive form. First, the public supplier chooses a rationing scheme. Second, the private firm, unable to observe

\* Corresponding author. Tel.: +41 (0)21 692 34 74.

E-mail addresses: [Simona.Grassi@unil.ch](mailto:Simona.Grassi@unil.ch) (S. Grassi), [ma@bu.edu](mailto:ma@bu.edu) (C.-t.A. Ma).

consumers' wealth levels, sets its prices according to consumers' costs of provision. Third, consumers who are rationed by the public supplier may purchase from the private firm. The public supplier aims to maximize aggregate consumer utility, while the private market consists of a profit-maximizing monopolist.

Rationing determines whom among consumers are entitled to public provision. In the first regime with wealth-based rationing, in equilibrium the public supplier must ration both poor and rich consumers, and implement price reduction in the private sector. What is the intuition behind this result? If poor consumers are supplied, then only rich consumers will be in the private market. The private firm cream-skims rich consumers by setting a high price. The public supplier can mitigate cream-skimming by rationing some poor consumers, making them available to the private market. The firm may then find it attractive to set a low price when costs are low. Rationing some poor consumers always yields a first-order gain in the form of price reductions.

In the second information regime, rationing can be based on both wealth and cost information. Clearly, the public supplier's equilibrium payoff must be higher compared to rationing based only on wealth. Surprisingly, in equilibrium the public supplier rations consumers according to cost information alone, ignoring wealth information altogether. The most efficient use of the public budget is to serve those consumers with the highest benefit–cost ratio. Using rationing to implement price reduction is suboptimal because cost effectiveness is already achieved. The private market is an option for higher-cost consumers who are willing to pay for the good, and remains so even if it sets a high price.

Clearly, if the public supplier can pick one piece of information for rationing, it will choose cost rather than wealth information. Once cost information is available, wealth information does not improve the design of optimal rationing. Crowd out – higher prices in the private sector – is not a concern when the public supply can be based on costs. In equilibrium, poor and rich consumers are treated equally because public supply is only based on costs.

Our information assumptions are plausible. The public sector has access to wealth information through tax returns. It may well have access to cost information because of service provisions. The firm has access to cost information. Dumping and cream-skimming are common problems in the health market. These problems are based on the premise that firms get to select less costly patients, so we follow a well recognized assumption in the literature.

In Grassi and Ma (forthcoming), we study a similar model, but the public rationing and private price schemes are chosen simultaneously. That model offers a longer term perspective on the interaction, because public rationing and private price schemes must be mutual best responses. In Grassi and Ma (forthcoming), cost effectiveness is an equilibrium when rationing is based on wealth and cost. If rationing is based on wealth, the game has a continuum of equilibria, all of which differ from the equilibrium here. In the equilibrium with the highest welfare, all poor consumers are supplied in the public sector while all rich are rationed and available in the market. Price reduction is never implemented there.

A common result in the literature of public provision of private goods is that the public sector serves poor consumers while the private sector serves rich consumers. This is the theme in Besley and Coate (1991) and Epple and Romano (1996). In our model, when rationing is based on wealth, the private sector will serve some poor consumers. Contrary to the standard result, a complete separation of the poor and rich does not obtain. In both Besley and Coate (1991) and Epple and Romano (1996), taxes and income redistributions are a concern, while we study nonprice rationing under a fixed budget. Also, while both assume a perfectly competitive private market, we consider a monopolistic private market.

A competitive private market is a common assumption in the literature. Barros and Olivella (2005) consider doctors working in the public sector who self-refer patients to their private practices. Prices paid by patients in the private sector are fixed, while doctors only refer low-cost patients. Iversen (1997) studies waiting-time rationing when there is a private market. Hoel and Sæther (2003) consider the effect of competitive supplementary insurance on a national health insurance system. Also the extensive literature on rationing by waiting times either assume away the private sector, or use a perfectly competitive private market (see for example, Gravelle and Siciliani, 2008, 2009). In fact, when the private market pricing rule is fixed, one only can study how it influences public policies. By contrast, we study how public policies influence private market responses.

Cost effectiveness as a criterion to allocate scarce resources has been advocated for a long time (see for example, Weinstein and Zeckhauser, 1973, or Garber and Phelps, 1997). Hoel (2007) discusses how cost effectiveness should be modified when treatments are also available in a competitive market. Following Hoel (2007), we study cost effectiveness when a private market exists, but we believe we are the first to derive cost effectiveness as the optimal rationing policy given a monopolistic private market.

Section 2 lays out the model. Section 3 and its subsections describe the firm's choice of the profit-maximizing prices and the equilibrium rationing when the public supplier observes only consumers' wealth level. Section 4 and its subsections focus on the information regime where wealth and cost levels are observed by the public supplier. The last section contains some concluding remarks. Appendix A contains proofs.

## 2. The model

### 2.1. Consumer utility and benefit

There is a set of consumers. Each consumer's wealth is either  $w_1$  or  $w_2$ , with  $0 < w_1 < w_2$ . Let  $m_i > 0$  be the mass of consumers with wealth  $w_i$ ,  $i = 1, 2$ . We call consumers with wealth  $w_1$  poor consumers, and consumers with wealth  $w_2$  rich consumers.

Each consumer may consume, at most, one unit of a health care good or treatment. Consumers differ in illness severity, and the cost of providing the good increases with severity. We use treatment cost to measure severity. Accordingly, we let the monetary cost of providing the good vary on the positive interval  $[\underline{c}, \bar{c}]$ , with a distribution function  $G : [\underline{c}, \bar{c}] \rightarrow [0, 1]$  and an associated density  $g$ . Let  $\gamma$  be the expected value of  $c$ . We identify a consumer by his wealth and provision cost, and call him either a rich or poor type- $c$  consumer. The lower support  $\underline{c}$  can be interpreted as the minimum severity level above which treatment may be warranted.

A type- $c$  consumer receives a health benefit from the treatment. This benefit varies according to severity. Let the function  $H : [\underline{c}, \bar{c}] \rightarrow \mathbb{R}_+$  denote the utility benefits, so a type- $c$  consumer receives a utility  $H(c)$  from treatment. We let the function  $H$  be strictly increasing and concave. A sicker consumer receives more benefit from treatment, but this benefit increases at a nonincreasing rate.<sup>1</sup>

If a type- $c$  consumer with wealth  $w_i$  pays a price  $p$  for the good, his utility is  $U(w_i - p) + H(c)$ , while if he does not consume the good (and pays nothing), his utility is  $U(w_i)$ . The function  $U$  is strictly increasing and strictly concave. We can use a general utility function where the utilities from consuming the good at price  $p$ , and from not consuming the good, are  $U(w - p, H(c))$  and  $U(w, 0)$ ,

<sup>1</sup> In Grassi and Ma (2009), the benefit  $H(c)$  is constant and normalized to 1. All the results presented here remain valid under the assumption of a constant benefit.

respectively. A separable utility function simplifies the analysis, but it does mean that rich and poor consumers receive the same utility from treatment.

A rich or poor type- $c$  consumer's willingness to pay is denoted by  $\tau_i(c)$ , and defined implicitly by

$$U(w_i - \tau_i) + H(c) = U(w_i), \quad i = 1, 2, \quad (1)$$

so  $\tau_i(c)$  is the maximum price a type- $c$  consumer with wealth  $w_i$  is willing to pay.

## 2.2. Public supplier and rationing policies

A public supplier has a budget  $B$  which is insufficient to provide the good for free to all consumers, so we assume  $B < (m_1 + m_2)\gamma$ . We consider two information regimes. In the first, the public supplier can use a nonprice rationing mechanism based on wealth. In the second, the public supplier uses a nonprice rationing mechanism based on both wealth and cost. The first regime corresponds to a means-test policy regime. For example, in the U.S., indigent consumers qualify for health insurance provided by Medicaid. The second regime includes a cost-effectiveness criterion that is commonly used in European countries. For example, all consumers are covered under a national insurance or health service, but services are only provided when they satisfy cost-effectiveness criteria.

When rationing is based on consumers' wealth, a rationing policy is a pair of fractions  $(\theta_1, \theta_2)$ ,  $0 \leq \theta_i \leq 1$ ,  $i = 1, 2$ . For each wealth class  $w_i$ , the public supplier rations  $\theta_i m_i$  consumers, and supplies  $(1 - \theta_i)m_i$  consumers. When rationing is based on consumers' wealth and costs, a rationing policy is a pair of functions  $(\phi_1, \phi_2)$ ,  $\phi_i : [c, \bar{c}] \rightarrow [0, 1]$ ,  $i = 1, 2$ . The value  $\phi_i(c)g(c)$  is the density of consumers with wealth  $w_i$  and cost  $c$  who are rationed. For each wealth class, the mass of rationed consumers with cost less than  $c$  is  $m_i \int_c^{\bar{c}} \phi_i(x)g(x)dx$ , while the mass of supplied consumers is  $m_i \int_c^{\bar{c}} [1 - \phi_i(x)]g(x)dx$ .

The public supplier's payoff is the sum of consumer utilities. We focus on the optimal public supply, not the optimal regulation of the entire market. Therefore, it is natural to assume that the public supplier is concerned with consumer surplus. We consider an unweighted sum of consumer surplus, but will discuss how our results will change when the poor's utility is given a higher weight than the rich's.

We now write down the benchmark rationing policies when there is no private supply. First, the aggregate consumer utility when rationing is based on wealth is

$$\sum_{i=1}^2 m_i \left\{ \theta_i U(w_i) + (1 - \theta_i) \int_c^{\bar{c}} [U(w_i) + H(c)]g(c)dc \right\},$$

where, for each wealth class, the rationed consumers have utility  $U(w_i)$ , while supplied consumers have utility  $U(w_i) + H(c)$ . The budget constraint is

$$\sum_{i=1}^2 m_i \int_c^{\bar{c}} [1 - \theta_i]cg(c)dc = \sum_{i=1}^2 m_i(1 - \theta_i)\gamma = B,$$

which says that the expected cost of supplying consumers is equal to the budget. Any rationing policy  $(\theta_1, \theta_2)$  that exhausts the budget is optimal. When rationing is based on wealth alone, supplying a poor consumer yields the same expected benefit as supplying a rich consumer.

Second, the aggregate consumer utility, when rationing is based on wealth and cost, is

$$\sum_{i=1}^2 m_i \left\{ \int_c^{\bar{c}} \phi_i(c)U(w_i)g(c)dc + \int_c^{\bar{c}} [1 - \phi_i(c)][U(w_i) + H(c)]g(c)dc \right\}.$$

The aggregate consumer utility simplifies to

$$\sum_{i=1}^2 m_i U(w_i) + \sum_{i=1}^2 m_i \int_c^{\bar{c}} [1 - \phi_i(c)]H(c)g(c)dc.$$

The budget constraint is

$$\sum_{i=1}^2 m_i \int_c^{\bar{c}} [1 - \phi_i(c)]cg(c)dc = B.$$

An optimal rationing policy based on wealth and cost is a pair  $(\phi_1, \phi_2)$  that maximizes the aggregate consumer utility subject to the budget constraint.

The optimal policy is the familiar *cost effectiveness principle*. Consider the benefit less the cost adjusted by the multiplier of the budget constraint:  $H(c) - \lambda c$ . It is optimal to supply a type- $c$  consumer if and only if this is positive.<sup>2</sup> We have interpreted  $c$  as the severity threshold for warranted treatment, so we let  $H(c)$  be sufficiently high. From this and the concavity of  $H$ , we have  $H(c) - \lambda c > 0$  if and only if  $c < c^B$  where  $c^B$  exhausts the budget if consumers with cost lower than  $c^B$  are supplied:  $(m_1 + m_2) \int_c^{c^B} cg(c)dc = B$ . As severity increases, the health benefit increases but at a nonincreasing rate, so it is not cost effective to treat very severe cases. Also, the cost effectiveness principle gives equal treatment to the rich and poor consumers because they receive the same benefit. This implies that wealth information is not required for optimal rationing.

## 2.3. Private market and consumers' willingness to pay

There is a private market which we model as a monopoly. The firm observes a consumer's cost  $c$ , but not his wealth  $w_i$ . To maximize profits, and given the public supplier's rationing policy, the private firm chooses prices as a function of costs. Because a consumer buys, at most, one unit of the indivisible good, price discrimination in the form of quantity discount is infeasible.<sup>3</sup> In this subsection, we present properties of the consumer's willingness-to-pay functions, as well as the monopolist's pricing strategy in a benchmark case of zero public supply.

Recall that the willingness to pay,  $\tau_i$ , in (1) is implicitly defined by  $U(w_i - \tau_i) + H(c) = U(w_i)$ ,  $i = 1, 2$ . Because  $U$  is strictly concave,  $\tau_1(c) < \tau_2(c)$  for each  $c$ ; a rich type- $c$  consumer is willing to pay more for the good than a poor type- $c$  consumer. From  $H$  and  $U$  strictly increasing, the willingness to pay,  $\tau_i(c)$ , is strictly increasing. Indeed, for  $i = 1, 2$  we have

$$\tau_i'(c) = \frac{H'(c)}{U'(w_i - \tau_i(c))} > 0. \quad (2)$$

<sup>2</sup> The Lagrangean is  $L = m_1[1 - \phi_1(c)]H(c) + m_2[1 - \phi_2(c)]H(c) + \lambda[B - m_1[1 - \phi_1(c)]c - m_2[1 - \phi_2(c)]c]$ , and the first-order derivative with respect to  $\phi_i$  is  $-m_i[H(c) - \lambda c]$ . It is optimal to set  $\phi_i = 0$  if and only if  $H(c) > \lambda c$ .

<sup>3</sup> In our model, if the firm managed to observe consumers' wealth and costs, it would extract all consumer surplus. In this case, the existence of the private market would be irrelevant to strategic consideration of rationing. Also, it is implausible to assume that the firm has no information about cost. A firm must eventually learn about cost, and it may renege on provision if the cost turns out to be higher than price.

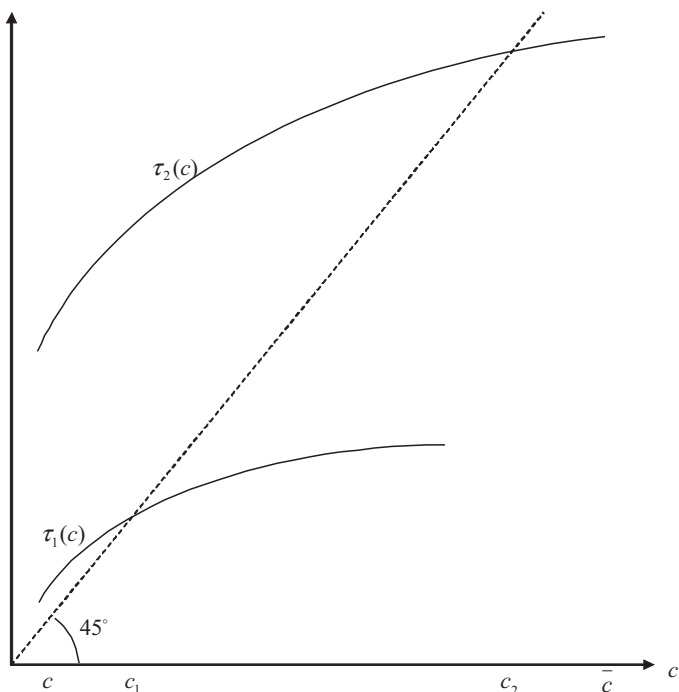


Fig. 1. Willingness to pay  $\tau_1$  and  $\tau_2$ .

Also, the willingness to pay,  $\tau_i(c)$ , is strictly concave.<sup>4</sup> Furthermore, at each  $c$ , we have  $\tau'_1(c) < \tau'_2(c)$ : the rich consumer's willingness to pay function is both higher and increasing faster than the poor consumer.<sup>5</sup>

Next we define a cost threshold. From our assumption that  $H(\underline{c})$  is sufficiently high, we also have  $\tau_1(\underline{c}) > \underline{c}$ : the firm is able to sell the good to consumers with low severities. We assume that  $\tau_1(c)$  is sufficiently concave and that  $\bar{c}$  is sufficiently large so that at some  $c_1 < \bar{c}$ , we have  $\tau_1(c_1) = c_1$ . In sum, we assume that at low severity levels, a poor consumer's willingness to pay is higher than the treatment cost, but there will be a cost sufficiently high (at  $c_1$ ) at which the benefit  $H(c)$  is not worthwhile to him. We can also analogously define  $c_2$  by  $\tau_2(c_2) = c_2$  (if there is such a  $c_2 \leq \bar{c}$ ).

Fig. 1 illustrates the properties of the two willingness-to-pay functions. There, the two increasing and concave functions graph the  $\tau_1$  and  $\tau_2$  for the poor and rich consumers. We assume that before  $c$  reaches  $\bar{c}$ , the concave function  $\tau_1$  must cut the 45-degree cost line from above.

At any  $c$ , the two willingness to pay,  $\tau_1(c)$  and  $\tau_2(c)$ , are the firm's candidate profit-maximizing prices. Clearly, if  $c \geq c_1$ , the firm cannot sell to poor consumers, because their willingness to pay is lower than cost. Therefore, at any  $c \geq c_1$ , the firm sets the price at  $\tau_2(c)$ , selling only to rich consumers. At cost  $c < c_1$ , there are two candidate prices,  $\tau_1(c)$  and  $\tau_2(c)$ . If the firm sells to both rich and poor consumers, it charges the lower price  $\tau_1(c)$ , but if it sells only to rich consumers, it charges the higher price  $\tau_2(c)$ . There is the usual trade-off between selling to less consumers at a higher price-cost margin and selling to more consumers at a lower

price-cost margin. When there is no public supply, the profits from these two prices are

$$\pi(\tau_1(c); c \leq c_1) \equiv (m_1 + m_2)[\tau_1(c) - c] \quad (3)$$

$$\pi(\tau_2(c); c \leq c_1) \equiv m_2[\tau_2(c) - c]. \quad (4)$$

By the strict concavity of  $\tau_i$ , the profit functions in (3) and (4) are strictly concave in  $c$ .

The cream-skimming literature typically hypothesizes that firms prefer to treat less severe patients. In most pricing models, a firm's profit is decreasing in cost. To rule out profits increasing in costs, we assume that both  $\tau'_1(c)$  and  $\tau'_2(c)$  are smaller than 1. In this case, the derivatives of (3) and (4) with respect to  $c$  are negative, so that profit does decrease with severity. From the expression for  $\tau'_i(c)$  in (2), if  $H'$  is smaller than  $U'$ , the assumption that  $\tau'_i(c) < 1$  is valid.

Next consider the difference between the profits from setting a low price and a high price, namely the difference between (3) and (4). After simplification, this difference is  $m_1[\tau_1(c) - c] - m_2[\tau_2(c) - \tau_1(c)]$ , and its derivative is  $m_1[\tau'_1(c) - 1] - m_2[\tau'_2(c) - \tau'_1(c)] < 0$ , because  $\tau'_2(c) > \tau'_1(c)$  and  $\tau'_1(c) < 1$ . Hence, the profit functions (3) and (4) cross, at most, once.

We will analyze situations in which the firm will find it optimal to reduce the price from  $\tau_2(c)$  to  $\tau_1(c)$  at some cost. Our interest is how rationing implements a price reduction. This issue would be moot if the price always stayed high at  $\tau_2(c)$ . We therefore assume that

$$(m_1 + m_2)[\tau_1(\underline{c}) - \underline{c}] > m_2[\tau_2(\underline{c}) - \underline{c}], \quad (5)$$

which says that at the lowest cost  $\underline{c}$ , the firm's optimal price is  $\tau_1(\underline{c})$  to sell to both poor and rich consumers. Because at  $c = c_1$ ,  $\tau_1(c_1) = c_1$ , so  $0 = (m_1 + m_2)[\tau_1(c_1) - c_1] < m_2[\tau_2(c_1) - c_1]$ , the firm's optimal price is the high price  $\tau_2(c_1)$  at  $c_1$ . Our assumption (5) implies that there must exist a unique  $c_m$  between  $\underline{c}$  and  $c_1$  such that  $(m_1 + m_2)[\tau_1(c_m) - c_m] = m_2[\tau_2(c_m) - c_m]$ , which simplifies to

$$m_1[\tau_1(c_m) - c_m] = m_2[\tau_2(c_m) - \tau_1(c_m)]. \quad (6)$$

The cost level  $c_m$  is where price reduction occurs. At cost  $c > c_m$ , the firm will charge the high price  $\tau_2(c)$ , but at  $c < c_m$ , it will charge the low price  $\tau_1(c)$ . Fig. 2 illustrates the determination of  $c_m$ . The two downward sloping, concave graphs are the profit functions (3) and (4), and their intersection defines  $c_m$ .

#### 2.4. Extensive forms

We consider the following extensive-form games:

**Stage 0:** For each consumer who has either wealth  $w_1$  or  $w_2$ , Nature draws a cost realization according to the distribution  $G$ . The private firm observes a consumer's cost realization, but not his wealth. Under rationing based on wealth, the public supplier observes a consumer's wealth, but not the cost realization. Under rationing based on wealth and cost, the public supplier observes a consumer's wealth and cost.

**Stage 1:** Under rationing based on wealth, the public supplier sets a rationing policy  $(\theta_1, \theta_2)$ ,  $0 \leq \theta_i \leq 1$ , supplying  $(1 - \theta_i)m_i$  of consumers with wealth  $w_i$ ,  $i = 1, 2$ . Under rationing based on wealth and cost, the public supplier sets a rationing policy  $(\phi_1, \phi_2)$ ,  $\phi_i : [\underline{c}, \bar{c}] \rightarrow [0, 1]$ , supplying  $[1 - \phi_i(c)]m_i$  of consumers with wealth  $w_i$  and cost  $c$ .

**Stage 2:** The firm sets a price for each cost realization.

**Stage 3:** Consumers who are rationed by the public supplier may purchase from the firm at prices set at Stage 2.

<sup>4</sup> From (2), we have  $\tau''_i(c) = ((U'(w_i - \tau_i)H''(c) + H'(c)U''(w_i - \tau_i)\tau'_i(c))/(U'(w_i - \tau_i(c))^2) < 0$ .

<sup>5</sup> By definition,  $U(w_1 - \tau_1(c)) + H(c) = U(w_1) < U(w_2) = U(w_2 - \tau_2(c)) + H(c)$ , so  $U(w_1 - \tau_1(c)) < U(w_2 - \tau_2(c))$ , and  $U'(w_1 - \tau_1(c)) > U'(w_2 - \tau_2(c))$ . From (2), it follows that  $\tau'_1(c) < \tau'_2(c)$ .

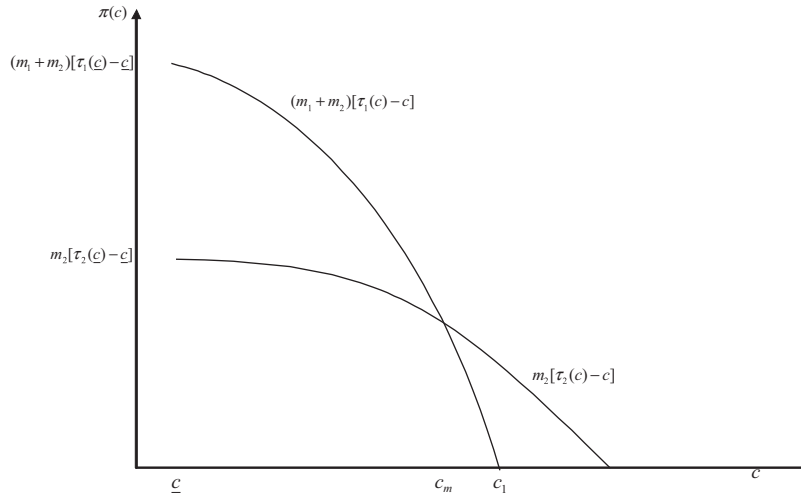


Fig. 2. Profits from low and high prices.

We study subgame-perfect equilibria. In Stage 1 the public supplier sets the rationing policies. A subgame in Stage 2 is a continuation game given the rationing policy in Stage 1. An equilibrium in Stage 2 refers to the equilibrium of the continuation subgame defined by a rationing policy in Stage 1.

### 3. Equilibrium rationing and prices in wealth-based rationing

#### 3.1. Equilibrium prices

In this subsection, we derive the equilibrium in Stage 2. Given a rationing policy  $(\theta_1, \theta_2)$ , only  $\theta_1 m_1$  of poor consumers and  $\theta_2 m_2$  of rich consumers are available to the firm. The firm may set a low

$$m_1 \left[ (1 - \theta_1) \left\{ U(w_1) + \int_{\underline{c}}^{\bar{c}} H(c) dG \right\} + \theta_1 \left\{ \int_{\underline{c}}^{c_r} [U(w_1 - \tau_1(c)) + H(c)] dG + \int_{c_r}^{\bar{c}} U(w_1) dG \right\} \right] \\ + m_2 \left[ (1 - \theta_2) \left\{ U(w_2) + \int_{\underline{c}}^{\bar{c}} H(c) dG \right\} + \theta_2 \left\{ \int_{\underline{c}}^{c_r} [U(w_2 - \tau_1(c)) + H(c)] dG + \int_{c_r}^{\bar{c}} [U(w_2 - \tau_2(c)) + H(c)] dG \right\} \right].$$

price  $\tau_1(c)$ , selling to both rich and poor consumers, or a high price  $\tau_2(c)$ , selling only to rich consumers. These strategies yield profits:

$$\pi(\tau_1(c); c \leq c_1) \equiv (m_1 \theta_1 + m_2 \theta_2) [\tau_1(c) - c] \quad (7)$$

$$\pi(\tau_2(c); c \leq c_1) \equiv m_2 \theta_2 [\tau_2(c) - c]. \quad (8)$$

These profit functions are both decreasing and concave, as in the case when the firm has access to all consumers (compare with (3) and (4)).

Recall that  $c_m$  is the cost threshold at which the equilibrium price switches from  $\tau_2(c)$  to  $\tau_1(c)$  when the firm has access to the entire market of consumers. Analogously, we can characterize the equilibrium in Stage 2 by the cost level  $c_r$  at which the price switches from  $\tau_2(c)$  to  $\tau_1(c)$  under the rationing policy  $(\theta_1, \theta_2)$ . If there is such a cost level  $c_r$  between  $\underline{c}$  and  $c_1$ , it is given by  $(m_1 \theta_1 + m_2 \theta_2) [\tau_1(c_r) - c_r] = m_2 \theta_2 [\tau_2(c_r) - c_r]$ , which simplifies to

$$m_1 \theta_1 [\tau_1(c_r) - c_r] = m_2 \theta_2 [\tau_2(c_r) - \tau_1(c_r)]; \quad (9)$$

otherwise we set  $c_r$  at  $\underline{c}$ .

As the cost drops below  $c_1$ , a price reduction is worthwhile only if there are enough poor consumers relative to rich ones. If there are few poor consumers in the market, the cost has to be much lower

than  $c_1$  for a price reduction to occur. In an extreme, if only the rich consumers are rationed and all the poor are supplied, the firm will never reduce the price. We summarize the firm's equilibrium prices in Stage 2 by the following (the proof omitted):

**Lemma 1.** Given a rationing policy  $(\theta_1, \theta_2)$ , if  $c_r$  in (9) is greater than  $\underline{c}$ , in equilibrium the firm sets the high price  $\tau_2(c)$  if  $c > c_r$ , and the low price  $\tau_1(c)$  if  $c < c_r$ . Otherwise, in equilibrium the firm always sets the high price  $\tau_2(c)$ .

#### 3.2. Equilibrium rationing

Given the equilibrium prices in Stage 2, the aggregate consumer utility is:

$$[m_1 U(w_1) + m_2 U(w_2)] + [m_1 (1 - \theta_1) + m_2 (1 - \theta_2)] \int_{\underline{c}}^{\bar{c}} H(c) dG \\ + m_2 \theta_2 \int_{\underline{c}}^{c_r} \Delta(c) dG, \quad (10)$$

In this expression, terms involving  $(1 - \theta_i)$  are consumers' utilities when they receive the public supply at no charge. Terms involving  $\theta_i$  are the market outcomes. For poor consumers, if their costs are below  $c_r$ , they purchase at  $\tau_1(c)$ , which actually leaves them no surplus (see definition of  $\tau_i(c)$  in (1)). Similarly, for rich consumers, if their costs are above  $c_r$ , they purchase at price  $\tau_2(c)$ , earning no surplus. However, if rich consumers' costs are below  $c_r$ , they earn a surplus  $U(w_2 - \tau_1(c)) + H(c) - U(w_2) \equiv \Delta(c) > 0$  since the price  $\tau_1(c)$  is lower than their willingness to pay,  $\tau_2(c)$ .

Using the definitions of the willingness to pay,  $\tau_i$ ,  $i = 1, 2$ , we simplify the aggregate consumer utility to

$$[m_1 U(w_1) + m_2 U(w_2)] + [m_1 (1 - \theta_1) + m_2 (1 - \theta_2)] \int_{\underline{c}}^{\bar{c}} H(c) dG \\ + m_2 \theta_2 \int_{\underline{c}}^{c_r} \Delta(c) dG, \quad (10)$$

where  $c_r \geq \underline{c}$  characterizes the firm's equilibrium price strategy. The first term is the consumers' utility from wealth. The middle term is the total expected benefit from public supply, while the last term is the sum of rich consumers' incremental surplus  $\Delta(c)$  when they purchase at price  $\tau_1(c)$ .

We introduce a new notation  $\beta \equiv B/\gamma$ . Because  $B$  denotes the available budget, and  $\gamma$  the expected cost,  $\beta$  is the number of supplied consumers. In equilibrium the budget  $B$  must be exhausted. Hence, we replace  $m_1(1 - \theta_1) + m_2(1 - \theta_2)$  by  $\beta$ , and simplify (10) to

$$\left[ m_1 U(w_1) + m_2 U(w_2) + \beta \int_{\underline{c}}^{\bar{c}} H(c) dG \right] + m_2 \theta_2 \int_{\underline{c}}^{c_r} \Delta(c) dG. \quad (11)$$

An equilibrium is a rationing policy  $(\theta_1, \theta_2)$  and the equilibrium price-reduction cost threshold in (9) that maximize (11), subject to the budget constraint

$$m_1(1 - \theta_1) + m_2(1 - \theta_2) = \beta \frac{B}{\gamma} (< m_1 + m_2), \quad (12)$$

and the boundary conditions  $\underline{c} \leq c_r$ , and  $0 \leq \theta_i \leq 1, i = 1, 2$ .

**Proposition 1.** *In equilibrium, the public supplier rations consumers in each wealth class:  $\theta_1 > 0$  and  $\theta_2 > 0$ , while the firm charges the low price  $\tau_1(c)$  when the consumer's cost is below a threshold  $c_r^*$ , where  $\underline{c} < c_r^* < c_1$ .*

**Proposition 1** (whose proof is in Appendix A) says that for any budget, the public supplier must ration some poor consumers and some rich consumers, and price reduction must occur. By assumption, if the firm has access to all consumers, it will reduce the price from  $\tau_2(c)$  to  $\tau_1(c)$  at  $c < c_m$ . The public supplier can always implement price reduction by setting  $\theta_1 = \theta_2 > 0$ , which maintains the same ratio of rich to poor consumers as in the full market (compare (6) and (9)). Some surplus in the private market must be available to consumers.

If  $\theta_1 = 0$ , then all poor consumers are supplied, and the price must remain high at all costs. If  $\theta_2 = 0$ , all rich consumers are supplied, so they do not participate in the private market. In either case, trade surplus in the private market cannot be realized, but this cannot happen in equilibrium. Therefore, we must have  $\theta_1 > 0$  and  $\theta_2 > 0$ , and cost reduction.

How does the public supplier set the rationing policy? What sort of trade-off is involved? The public supplier's objective is to maximize the consumer surplus in (11). Without the constant terms, the objective function is

$$m_2 \theta_2 \int_{\underline{c}}^{c_r} \Delta(c) dG. \quad (13)$$

This is the incremental surplus enjoyed by rationed rich consumers buying at price  $\tau_1(c)$ ; all of them have costs below the price-reduction cost threshold  $c_r$ . Obviously, the public supplier would like threshold  $c_r$  to be high, and would like  $\theta_2$  to be high. In that case, more rich consumers can realize more surplus from the market. But these two goals, raising the price-reduction cost threshold and rationing more rich consumers, are incompatible.

Consider rationing more rich consumers. This increases  $\theta_2$ . Some of the budget is now available to supply poor consumers, so  $\theta_1$  decreases. In other words, there are more rich consumers and less poor consumers in the market. The firm finds it less profitable to reduce price, so cost must fall lower before price reduction happens in equilibrium. The value of the cost threshold  $c_r$  decreases as  $\theta_2$  increases. Raising both  $\theta_2$  and  $c_r$  is impossible.<sup>6</sup> The basic trade-off is between a bigger range of cost reduction for fewer rich consumers and a smaller range of cost reduction for more rich consumers.

<sup>6</sup> If  $\theta_2$  increases, then  $\theta_1$  must decrease due to the budget constraint (12). From (9), when  $\theta_2$  increases and  $\theta_1$  decreases,  $c_r$  must decrease. This is because for all  $c$ ,  $\tau_2(c) - \tau_1(c)$  is increasing in  $c$  whereas  $\tau_1(c) - c$  is decreasing in  $c$ .

Changes in  $\theta_2$  and  $c_r$  are constrained by the budget as well as the equilibrium in Stage 2. We use (9) and (12) to eliminate  $\theta_1$  and obtain

$$m_2 \theta_2 = K \frac{\tau_1(c_r) - c_r}{\tau_2(c_r) - c_r}, \quad (14)$$

where  $K \equiv m_1 + m_2 - \beta > 0$ . Substituting (14) into (13), we now can characterize the equilibrium by the choice of  $c_r$  that maximizes

$$K \left[ \frac{\tau_1(c_r) - c_r}{\tau_2(c_r) - c_r} \right] \int_{\underline{c}}^{c_r} \Delta(c) dG \quad (15)$$

subject to the boundary conditions.

The objective function in (13) is a product of  $[\tau_1(c_r) - c_r]/[\tau_2(c_r) - c_r]$  and  $\int_{\underline{c}}^{c_r} \Delta(c) dG$ . They are, respectively, the ratio of price-cost margins at low and high prices, and the incremental consumer surplus. The total effect on (15) as  $c_r$  changes depends on the proportional changes in the product components as  $c_r$  changes. The first is decreasing in  $c_r$  while the second is increasing.

We now present the characterization of the equilibrium in the following proposition (whose proof is in Appendix A):

**Proposition 2.** *If the budget  $B$  is sufficiently large, the equilibrium price-reduction cost threshold  $c_r^*$  is the unique solution of*

$$\left[ \frac{1 - \tau_2'(c_r)}{\tau_2(c_r) - c_r} - \frac{1 - \tau_1'(c_r)}{\tau_1(c_r) - c_r} \right] + \frac{\Delta(c_r)g(c_r)}{\int_{\underline{c}}^{c_r} \Delta(c) dG} = 0 \quad (16)$$

and the equilibrium rationing policy  $(\theta_1, \theta_2)$  can be recovered from (12) and (14):

$$\theta_1 = \frac{m_1 + m_2 - \beta}{m_1} \left[ \frac{\tau_2(c_r^*) - \tau_1(c_r^*)}{\tau_2(c_r^*) - c_r^*} \right] < 1 \quad \text{and}$$

$$\theta_2 = \frac{m_1 + m_2 - \beta}{m_2} \left[ \frac{\tau_1(c_r^*) - c_r^*}{\tau_2(c_r^*) - c_r^*} \right] < 1. \quad (17)$$

If the budget is small, either  $\theta_1$  or  $\theta_2$  may be equal to 1, and the public supplier may ration an entire wealth class. If  $\theta_j = 1$ , then  $\theta_i = 1 - (\beta/m_j)$ ,  $i, j = 1, 2$ , and  $i \neq j$ , and the value of  $c_r^*$  then is obtained from (9) with  $\theta_i = 1$ .

Eq. (16) in Proposition 2 is the first-order condition for the maximization of (15) when the boundary conditions for  $\theta_i \leq 1$  do not bind. If a boundary condition on  $\theta_i \leq 1$  binds, then the constraint set uniquely determines the optimum.

The trade-off is between rationing rich consumers so they enjoy the incremental surplus in the private market and rationing poor consumers to implement more price reduction. The optimal trade-off is achieved by differentially supplying rich and poor consumers. With a large budget, the manipulation of this ratio is easier. This corresponds to the first part of Proposition 2 when the boundary conditions  $\theta_i \leq 1$  are slack. With a small budget, the manipulation is to withhold supply to a whole class of consumers. This corresponds to a binding boundary condition.

In Fig. 3, we graph the downward-sloping budget line (12), and the dotted lines for the boundary conditions for  $m_i \theta_i$ . The feasible set is the triangle formed by the boundary conditions and the budget line. A bigger budget means more consumers can be supplied (a higher  $\beta$ ), so the budget line shifts downward. The upward-sloping line graphs the combinations of  $m_1 \theta_1$  and  $m_2 \theta_2$  that implement a price reduction at cost threshold  $c_r^*$ .

In Fig. 3, the boundary conditions  $\theta_i \leq 1$  do not bind. The price-reduction cost threshold  $c_r^*$  (in (16)) is implemented by the policy in ( ) (the intersection between the two solid lines in the figure). Here, there is enough budget to implement  $c_r^*$ . The cost threshold  $c_r^*$  is independent of the budget, as is the ratio between  $\theta_1$  and  $\theta_2$ .

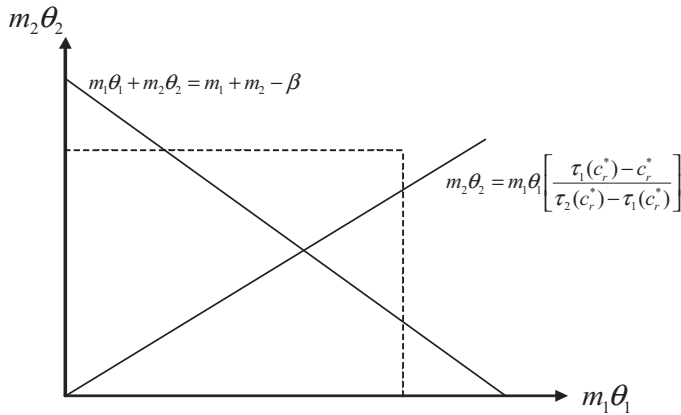


Fig. 3. Budget constraint, cost threshold, and boundary conditions.

The second part of Proposition 2 is about the equilibrium when the budget is small. Suppose that the ratio of rich to poor consumers in the market should decrease to favor price reduction, so this requires supplying more rich consumers than poor ones. With a small budget, this may mean rationing all poor consumers so all of them are in the private market. A boundary condition binds.

In general, the equilibrium cost threshold  $c_r^*$  may be higher or lower than  $c_m$ . Nevertheless, if  $\theta_2 = 1$ , we must have  $\theta_1 < 1$ , and  $c_r^* < c_m$ . Rationing all rich consumers means that the budget must be spent on poor consumers. With less poor consumers in the market, price reduction is less often. Then public supply reduces transactions in the private market. This explains crowd out.

The public supplier's objective is to maximize the sum of poor and rich consumers' utilities. If there is any equity concern, more weight will be given to poor consumers. In this case, rationing will

$$\int_{\underline{c}}^{\bar{c}} \{m_1[1 - \phi_1(c)][U(w_1) + H(c)] + m_2[1 - \phi_2(c)][U(w_2) + H(c)]\}dG(c) + \int_{\underline{c}}^{\bar{c}} m_1 \phi_1(c) \{ [1 - p(c)]U(w_1) + p(c)[U(w_1 - \tau_1(c)) + H(c)] \} dG(c) + \int_{\underline{c}}^{\bar{c}} m_2 \phi_2(c) \{ [1 - p(c)][U(w_2 - \tau_2(c)) + H(c)] + p(c)[U(w_2 - \tau_1(c)) + H(c)] \} dG(c).$$

favor the poor, so fewer poor consumers will be in the market. The equilibrium price-reduction cost threshold will fall, so prices tend to be higher. Equity concern tends to reduce the likelihood of price reduction, and generates a larger extent of crowd out.

#### 4. Equilibrium rationing and prices in wealth-cost based rationing

##### 4.1. Equilibrium prices

We begin with the equilibrium prices given a rationing policy  $(\phi_1, \phi_2)$ ,  $\phi_i : [\underline{c}, \bar{c}] \rightarrow [0, 1]$ . Again, there are only two possible equilibrium prices in the private market, the low price  $\tau_1(c)$  and the high price  $\tau_2(c)$ . For any  $c > c_1$ , the firm's unique best response is  $\tau_2(c)$ . For any  $c$  between  $\underline{c}$  and  $c_1$ , the firm chooses between the low price,  $\tau_1(c)$ , and the high price,  $\tau_2(c)$ . The firm's profit from the low price is  $[m_1 \phi_1(c) + m_2 \phi_2(c)][\tau_1(c) - c]$ ; the profit is  $m_2 \phi_2(c)[\tau_2(c) - c]$  from the high price. The firm sets the low price if  $[m_1 \phi_1(c) + m_2 \phi_2(c)][\tau_1(c) - c] \geq m_2 \phi_2(c)[\tau_2(c) - c]$ , or

$$m_1 \phi_1(c)[\tau_1(c) - c] \geq m_2 \phi_2(c)[\tau_2(c) - \tau_1(c)]. \quad (18)$$

It sets the high price if (18) is violated, and it may randomize between  $\tau_1(c)$  and  $\tau_2(c)$  if (18) holds as an equality. These are the equilibrium prices in Stage 2.

We now define an indicator function for equilibria when  $c < c_1$ . Let  $p : [\underline{c}, c_1] \rightarrow [0, 1]$ . Given a policy  $(\phi_1, \phi_2)$ , we set  $p(c) = 1$  if (18) holds as a strict inequality,  $p(c) = 0$  if (18) is violated, and  $p(c)$  to a number between 0 and 1 if (18) holds as an equality. When  $p(c)$  takes the value 0, the firm chooses the high price, so there is no price reduction. When  $p(c)$  takes the value 1, the firm chooses the low price, so there is a price reduction.

**Lemma 2.** For  $c$  between  $\underline{c}$  and  $c_1$ , any equilibrium in Stage 2 is given by a function  $p : [\underline{c}, c_1] \rightarrow [0, 1]$  satisfying the following two inequalities:

$$p(c)\{m_1 \phi_1(c)[\tau_1(c) - c] - m_2 \phi_2(c)[\tau_2(c) - \tau_1(c)]\} \geq 0 \quad (19)$$

$$[1 - p(c)]\{m_1 \phi_1(c)[\tau_1(c) - c] - m_2 \phi_2(c)[\tau_2(c) - \tau_1(c)]\} \leq 0 \quad (20)$$

Lemma 2 (whose proof is in Appendix A) defines a price-reduction function  $p(c)$  to indicate the equilibrium in Stage 2. The term inside the curly brackets of (19) and (20) is the profit difference between charging the low price and the high price (see (18)). The inequalities (19) and (20) are "complementary" conditions for price reduction. When the firm charges the low price,  $p(c)$  must be equal to 1 for (19) and (20) to hold simultaneously; conversely, when the firm charges the high price,  $p(c)$  must be equal to 0.

For ease of exposition, we extend the function  $p$  from the domain  $[\underline{c}, c_1]$  to  $[\underline{c}, \bar{c}]$ , and set  $p(c) = 0$  for  $c > c_1$ . This simply says that there is no price reduction for  $c > c_1$ . This extensions allows us to write payoffs in a simpler way.

##### 4.2. Equilibrium rationing

We begin with the public supplier's payoff given the equilibrium prices:

$$\int_{\underline{c}}^{\bar{c}} \{m_1[1 - \phi_1(c)][U(w_1) + H(c)] + m_2[1 - \phi_2(c)][U(w_2) + H(c)]\}dG(c) + \int_{\underline{c}}^{\bar{c}} m_1 \phi_1(c) \{ [1 - p(c)]U(w_1) + p(c)[U(w_1 - \tau_1(c)) + H(c)] \} dG(c) + \int_{\underline{c}}^{\bar{c}} m_2 \phi_2(c) \{ [1 - p(c)][U(w_2 - \tau_2(c)) + H(c)] + p(c)[U(w_2 - \tau_1(c)) + H(c)] \} dG(c).$$

In this expression, the first integral is the sum of utilities of supplied consumers; each consumer gets the benefit  $H(c)$  without incurring any cost. The second integral is the sum of utilities of rationed poor consumers. A poor type- $c$  consumer will encounter a price reduction with probability  $p(c)$ . If there is no price reduction, the poor consumer does not buy, so his payoff is  $U(w_1)$ . If there is a price reduction, the poor consumer buys at price  $\tau_1(c)$ , hence the term  $U(w_1 - \tau_1(c)) + H(c)$ . The last integral is the sum of utilities of rationed rich consumers. If there is no price reduction, the rich consumer buys at  $\tau_2(c)$ , hence the term  $U(w_2 - \tau_2(c)) + H(c)$ . If there is a price reduction, he buys at  $\tau_1(c)$ , hence the term  $U(w_2 - \tau_1(c)) + H(c)$ .

The gain in utility when consumers participate in the market is due to the rich consumers purchasing at the low price  $\tau_1(c)$ . Poor consumers either do not buy or buy at their reservation price  $\tau_1(c)$ , gaining no surplus from the private market. We use the definitions of  $\tau_1(c)$  and  $\tau_2(c)$  to simplify the payoff to:

$$m_1 U(w_1) + m_2 U(w_2) + \int_{\underline{c}}^{\bar{c}} \{m_1[1 - \phi_1(c)] + m_2[1 - \phi_2(c)]\}H(c)dG(c) + \int_{\underline{c}}^{\bar{c}} m_2 \phi_2(c)p(c)\Delta(c)dG(c). \quad (21)$$

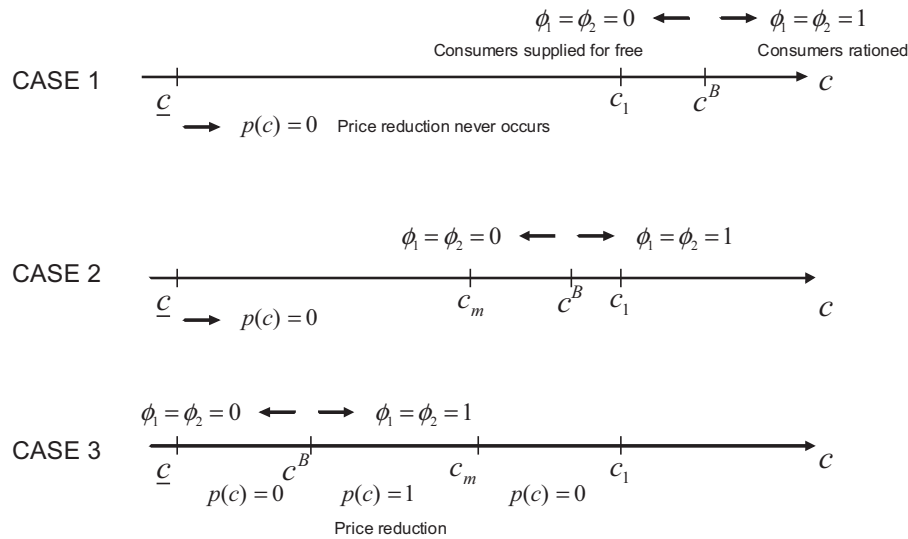


Fig. 4. Equilibrium rationing and price reduction.

In (21), the first integral is consumers' utility gain from the public supply, and the second integral is the incremental gain of rationed rich consumers who purchase in the private market at the low price  $\tau_1(c)$ . (Recall  $\Delta(c) \equiv U(w_2 - \tau_1(c)) + H(c) - U(w_2)$ .)

The optimal rationing policy is one that maximizes (21) subject to the budget constraint, and the equilibrium prices in the private market. By Lemma 2, the equilibrium price in Stage 2 is characterized by the price-reduction function  $p(c)$ . Ignoring the constant terms in (21), we write down the maximization program for the public supplier's equilibrium policy: choose a policy  $(\phi_1, \phi_2)$  and a function  $p$  to maximize

$$\int_{\underline{c}}^{\bar{c}} \{m_1[1 - \phi_1(c)] + m_2[1 - \phi_2(c)]\}H(c)dG(c) + \int_{\underline{c}}^{\bar{c}} m_2\phi_2(c)p(c)\Delta(c)dG(c) \quad (22)$$

subject to

$$B - \int_{\underline{c}}^{\bar{c}} \{m_1[1 - \phi_1(c)] + m_2[1 - \phi_2(c)]\}cdG(c) \geq 0 \quad (23)$$

$$p(c)\{m_1\phi_1(c)[\tau_1(c) - c] - m_2\phi_2(c)[\tau_2(c) - \tau_1(c)]\} \geq 0 \quad (24)$$

$$[1 - p(c)]\{m_1\phi_1(c)[\tau_1(c) - c] - m_2\phi_2(c)[\tau_2(c) - \tau_1(c)]\} \leq 0, \quad (25)$$

and the boundary conditions  $0 \leq \phi_i(c) \leq 1, i = 1, 2, 0 \leq p(c) \leq 1$ , each  $c$  in  $[\underline{c}, \bar{c}]$ , and  $p(c) = 0$  for  $c > c_1$ . Inequality (23) is the budget constraint. For completeness, we have rewritten the two inequalities in Lemma 2 as (24) and (25).

**Proposition 3.** *In the optimal rationing policy based on wealth and cost, the public supplier rations consumers if and only if their costs are above a threshold. That is, in an equilibrium,*

$$\begin{aligned} \phi_1(c) = \phi_2(c) = 0 & \text{ for } c < c^B \\ \phi_1(c) = \phi_2(c) = 1 & \text{ for } c > c^B, \end{aligned}$$

where the cost threshold  $c^B$  is defined by  $\int_{\underline{c}}^{c^B} (m_1 + m_2)cdG(c) = B$ .

Proposition 3 (whose proof is in Appendix A) says that equilibrium rationing coincides with cost effectiveness when the private market is absent. Rich and poor consumers are treated equally, and a type- $c$  consumer is given public supply if and only if the

net benefit is high:  $H(c) > \lambda c$ , where  $\lambda$  is the multiplier of the budget constraint (23). Because the benefit from consumption  $H(c)$  is concave, the benefit–cost ratio,  $H(c)/c$ , is higher at low costs and decreases with  $c$ , so low-cost consumers get the public supply. The cost level  $c^B$  in the proposition refers to one at which supplying the good to consumers with costs below  $c^B$  will exhaust the budget. The presence of the firm does not change the cost-effectiveness principle. What is behind this result?

Unlike the regime when rationing is based only on wealth, implementing cost effectiveness is possible when rationing is based on wealth and cost. The firm sets the high price  $\tau_2(c)$  when there are many rich consumers, but the low price  $\tau_1(c)$  if there are few rich consumers. How does the firm's best response interact with cost effectiveness? If the public supplier provides for the rich, price reduction is irrelevant. If the public supplier provides for the poor, price reduction cannot be an equilibrium: without poor consumers, the firm will set the high price. Cost effectiveness, however, calls for equal treatment to the rich and the poor. At each cost level, the public supplier either provides for both rich and poor consumers, or none at all. The ratio between rich and poor consumers in the private market is the same as if the firm had access to all consumers. If a price reduction occurs, it follows the same fashion as if there was not any public supply. Crowd out does not happen in equilibrium.

Fig. 4 shows the three cases that make up the proof of Proposition 3. Price reduction happens if and only if cost falls below  $c_m$ . In Case 1, the budget is large so that it is cost effective to supply all consumers with costs up to a threshold above  $c_1$ . In Case 2, the budget is medium sized, and may cover some consumers with cost above  $c_m$ . There is still no price reduction at cost  $c$  between  $c_m$  and  $c_1$  because  $c_m$  is the minimum cost level at which price reduction begins to be profitable. In Case 3, the budget is small. Here, price reduction occurs at  $c < c_m$ .

Clearly, the public supplier's equilibrium payoff – aggregate consumer utility – under rationing based on cost and wealth cannot be lower than rationing based on wealth alone. In Proposition 3 the optimal rationing rule is based only on cost. Once cost information is available, wealth information does not improve the public supplier's payoff. We summarize by the following:

**Corollary 1.** *Equilibrium aggregate consumer utility is higher under rationing based on cost than wealth. If the public supplier must pick between cost and wealth information to administer rationing, it optimally will choose cost information.*



Finally, we comment on equity concern. Actually, under rationing based on wealth and cost, the public supplier rations rich and poor consumers equally. If there is an explicit constraint on supplying poor consumers more, then the cost effectiveness principle cannot be applied directly. When public supply favors the poor, fewer of them will be in the market, and the rich will be less likely to experience a price reduction.

## 5. Concluding remarks

We have presented a model to study the effect of rationing on prices in the private market. Public policies should take into account market responses. We show that if rationing is based on wealth information, the optimal policy must implement a price reduction in the private market. This is achieved by leaving some poor consumers in the private market. If the public supplier observes consumers' wealth and cost, optimal rationing is based on cost effectiveness; wealth information is not necessary. Our model sheds light on crowd out, and the design of public programs when private market responses are important.

We assume two wealth classes to make the model tractable. Extending the model and deriving the equilibrium rationing scheme for many, or a continuum of wealth classes involve complex computation. Many possible price reduction configurations must be considered. We believe that our basic result is robust. In other words, some consumers with lower wealth will be rationed to implement more price reductions.

We have used a separable utility assumption. In Grassi and Ma (forthcoming), we list some factors that may influence the results when utility functions are not separable in money and benefits. A secondary effect from consumption on the marginal utility of income will have to be considered. If income effects are small, our results extend to the general utility function.

We have assumed a fixed budget. Extending the model to consider an optimal budget is fairly straightforward. We have obtained the optimal policies in the two information regimes, so that the optimized aggregate consumer utility is available. Once the cost of public funds is specified, the usual optimization steps can be taken to characterize the optimal budget.

The analysis here is limited to free public supply, but obviously a fixed user fee can be included. Due to risk aversion, publicly provided health insurance usually does not impose significant copayments. Nevertheless, a general analysis of optimal monetary subsidy may be fruitful.

## Acknowledgement

We thank seminar participants at Universitat Autònoma de Barcelona, University of Bern, Universidad Carlos III de Madrid, Universit de Lausanne, Michigan State University, and Universit degli Studi di Milano; and conference participants at XXIV Jornadas de Economía Industrial in Vigo, ECARES-CEPR in Brussels, and Seminars in Health Economics and Policy in Crans Montana; and especially Pedro Barros for comments and suggestions. We also thank Editor Tony Cuyler, and two referees for their advice.

## Appendix A.

**Proof of Proposition 1.** Because all terms in square brackets in the objective function (11) are constant, we alternatively can write the objective function as  $m_2\theta_2 \int_{\underline{c}}^{c_r} \Delta(c)dG$ .

The boundary conditions  $\underline{c} \leq c_r$  and  $0 \leq \theta_2$  do not bind. If either  $c_r = \underline{c}$  or  $\theta_2 = 0$  at a solution, then the optimized value is  $m_2\theta_2 \int_{\underline{c}}^{c_r} \Delta(c)dG = 0$ . We show that a rationing policy with

$\theta_1 = \theta_2 = k > 0$  does strictly better. This policy satisfies the budget constraint (12) for some  $0 < k < 1$ . Moreover, from (6) and (9), we have  $c_r = c_m > \underline{c}$  by assumption. Therefore, the rationing policy  $\theta_1 = \theta_2 = k$  is feasible, and yields a payoff  $m_2k \int_{\underline{c}}^{c_m} \Delta(c)dG > 0$ . This implies that at a solution  $c_r > \underline{c}$  and  $\theta_2 > 0$ . Because  $c_r > \underline{c}$ , it follows from (9) that  $\theta_1$  must be bounded away from 0.

**Proof of Proposition 2.** The steps for simplifying the objective function into (15) are already laid out before the Proposition. We differentiate the logarithm of (15) to get the first-order condition (16). We now show that the solution to this first-order condition is unique. The objective function (15) is the product of  $[\tau_1(c_r) - c_r]/[\tau_2(c_r) - c_r]$  and  $\int_{\underline{c}}^{c_r} \Delta(c)dG$ . We show that the derivative of the square-bracketed term is negative. The derivative of its logarithm is the term in square brackets in (16), and that is negative. To see this, note that  $\tau_2(c_r) - c_r > \tau_1(c_r) - c_r$ , and  $1 - \tau_2'(c_r) < 1 - \tau_1'(c_r)$ , and all these terms are positive. So we have  $\{[1 - \tau_2'(c_r)]/[\tau_2(c_r) - c_r]\} < \{[1 - \tau_1'(c_r)]/[\tau_1(c_r) - c_r]\}$ . Obviously the integral in (15) is increasing in  $c_r$ , and its derivative is the other term in (16). We conclude that (16) has a unique solution.

The equilibrium rationing policy in (17) is obtained by solving (12) and (14) simultaneously at  $c_r = c_r^*$ .

If  $\beta$  is sufficiently large, the right-hand side values in (17) will be less than 1, and the omitted boundary conditions  $\theta_i \leq 1$  are satisfied. Otherwise, a boundary condition binds.

**Proof of Lemma 2.** Consider any equilibrium prices in Stage 2. In this equilibrium, at cost  $c$  the firm will charge either  $\tau_1(c)$  or  $\tau_2(c)$  depending on whether (18) is satisfied. If we have defined  $p$  by the method just before the statement of the Lemma, inequalities (19) and (20) are satisfied.

Conversely, let a function  $p : [\underline{c}, c_1] \rightarrow [0, 1]$  satisfy inequalities (19) and (20). We show that it characterizes a best response pricing strategy to any policy  $(\phi_1, \phi_2)$ . Suppose that  $p(c) = 1$ . Inequality (20) is satisfied by any  $\phi_1(c)$  and  $\phi_2(c)$ . Inequality (19) requires the term inside the curly brackets to be positive, and this means that (18) is satisfied. Next, suppose that  $p(c) = 0$ . Inequality (19) is always satisfied. Inequality (20) requires the term inside the curly brackets in (20) to be negative, and this means that (18) is violated. Last, if  $p(c)$  is a number strictly between 0 and 1, both (19) and (20) must hold as equalities, so that (18) must be an equality. Each value of  $p(c)$  satisfying (19) and (20) corresponds to an equilibrium price.

**Proof of Proposition 3.** We use pointwise optimization to solve for the optimal rationing policy. We consider a relaxed program in which constraint (25) is omitted; we will show that in the solution of the relaxed program constraint (25) is satisfied. To simplify notation, we multiply (24) by  $g(c)$ , so that  $g(c)$  can be ignored for pointwise optimization. Let  $\lambda$  denote the multiplier for the budget constraint (23), and  $\mu(c)$  the multiplier for (24) at  $c$ . The Lagrangian is

$$L = m_1[1 - \phi_1(c)]H(c) + m_2[1 - \phi_2(c)]H(c) + m_2\phi_2(c)p(c)\Delta(c) \\ + \lambda\{B - m_1[1 - \phi_1(c)]c - m_2[1 - \phi_2(c)]c\} \\ + \mu(c)p(c)\{m_1\phi_1(c)[\tau_1(c) - c] - m_2\phi_2(c)[\tau_2(c) - \tau_1(c)]\},$$

where we have omitted the boundary conditions on  $\phi_i$  and  $p$ .

For  $c > c_1$ ,  $p(c) = 0$ , so there is no need to optimize over  $p$ , and the first-order derivatives are

$$\frac{\partial L}{\partial \phi_1} = -m_1H(c) + \lambda m_1 c \quad (26)$$

$$\frac{\partial L}{\partial \phi_2} = -m_2H(c) + \lambda m_2 c. \quad (27)$$

For  $c < c_1$ , the first-order derivatives are

$$\frac{\partial L}{\partial \phi_1} = -m_1 H(c) + \lambda m_1 c + \mu(c) p(c) m_1 [\tau_1(c) - c] \quad (28)$$

$$\begin{aligned} \frac{\partial L}{\partial \phi_2} = & -m_2 H(c) + \lambda m_2 c - \mu(c) p(c) m_2 [\tau_2(c) - \tau_1(c)] \\ & + m_2 p(c) \Delta(c) \end{aligned} \quad (29)$$

$$\begin{aligned} \frac{\partial L}{\partial p} = & m_2 \phi_2(c) \Delta(c) + \mu(c) \{m_1 \phi_1(c) [\tau_1(c) - c] \\ & - m_2 \phi_2(c) [\tau_2(c) - \tau_1(c)]\}. \end{aligned} \quad (30)$$

We consider three cases, according to the size of the budget.

*Case 1* is when the budget is large:  $c^B > c_1$ ; that is, the budget is sufficient to cover costs up to a level where poor consumers' willingness to pay equals cost  $\tau_1(c_1) = c_1$ . To prove the proposition, we set  $\lambda = H(c^B)/c^B$ . Now consider  $c > c^B$ . Because  $H(c)/c$  is decreasing with  $c$ , the first-order derivatives (26) and (27) become (after dividing each term by  $c$ ),  $-m_1(H(c)/c) + m_1(H(c^B)/c^B)$ , and  $-m_2(H(c)/c) + m_2(H(c^B)/c^B)$ , respectively. Both are strictly positive. Hence it is optimal to set  $\phi_i(c) = 1$ . Next, consider  $c_1 < c < c^B$ . Then the first-order derivatives (26) and (27) become strictly negative, and it is optimal to set  $\phi_i(c) = 0$ .

Now consider  $c < c < c_1$ . We claim that  $\phi_i(c) = p(c) = 0$ . At these values, the derivatives (28), (29), and (30) are negative. At  $\phi_i(c) = 0$ , the derivative (30) is zero; hence it is optimal to set  $p(c) = 0$ . At  $p(c) = 0$ , (28) and (29) reduce to  $-m_1(H(c)/c) + m_1(H(c^B)/c^B)$ , and  $-m_2(H(c)/c) + m_2(H(c^B)/c^B)$ , respectively, and both are strictly negative. It is optimal to set  $\phi_i(c) = 0$ . Finally, the omitted constraint (25) is satisfied since  $\phi_i(c) = 0$ .

*Case 2* is when the budget  $c^B$  is lower, between  $c_m$  and  $c_1$ ,  $c_m < c^B < c_1$ . Recall that  $c_m$  is the cost level at which the firm will set the low price  $\tau_1(c)$  if it has access to all consumers ( $m_1[\tau_1(c_m) - c_m] = m_2[\tau_2(c_m) - \tau_1(c_m)]$ , see also (6)). Again, we set  $\lambda = H(c^B)/c^B$ . For  $c > c_1$ , the first-order derivatives (26) and (27) are  $-m_1(H(c)/c) + m_1(H(c^B)/c^B)$  and  $-m_2(H(c)/c) + m_2(H(c^B)/c^B)$ , respectively. Both are strictly positive. Hence it is optimal to set  $\phi_i(c) = 1$ .

Next, consider  $c^B < c < c_1$ . We set  $\mu(c)$  to satisfy

$$m_2 \Delta(c) + \mu(c) \{m_1 [\tau_1(c) - c] - m_2 [\tau_2(c) - \tau_1(c)]\} = 0. \quad (31)$$

Because  $c > c^B > c_m$ , we have  $m_1[\tau_1(c) - c] < m_2[\tau_2(c) - \tau_1(c)]$ . Therefore,  $\mu(c) > 0$ . We claim that  $p(c) = 0$ ,  $\phi_i(c) = 1$ . Given  $p(c) = 0$ , first-order derivatives (28) and (29) are  $-m_1(H(c)/c) + m_1(H(c^B)/c^B)$  and  $-m_2(H(c)/c) + m_2(H(c^B)/c^B)$ , respectively. Both are strictly positive. Hence, it is optimal to set  $\phi_i(c) = 1$ . Given  $\phi_i(c) = 1$ , by the choice of  $\mu(c)$  satisfying (31), the derivative (30) is zero. Hence, setting  $p(c) = 0$  is optimal. Obviously, the omitted constraint (25) is satisfied since  $\phi_i(c) = 1$ .

Next, consider  $c < c < c^B$ . We claim that  $\phi_i(c) = p(c) = 0$ . Given  $p(c) = 0$ , the first-order derivatives (28) and (29) are both negative

when  $c < c^B$ . Hence it is optimal to set  $\phi_i(c) = 0$ . Next, given that  $\phi_i(c) = 0$ , the derivative (30) is zero. Hence it is optimal to set  $p(c) = 0$ . Again, the omitted constraint (25) is satisfied since  $\phi_i(c) = 0$ .

*Case 3* is when the budget is small,  $c^B < c_m$ . We set  $\lambda = H(c^B)/c^B$ . For  $c > c_1$ , we use the same argument as in Case 1 and Case 2, and  $\phi_i(c) = 1$ . For  $c_m < c < c_1$ , we claim that  $\phi_i(c) = 1$  and  $p(c) = 0$ . We show this by the same argument in Case 2. When  $\mu(c)$  is set to be sufficiently large, the first-order derivative (30) is zero, so that  $p(c) = 0$  is optimal when  $\phi_i(c) = 1$ . When  $p(c) = 0$ , setting  $\phi_i(c) = 1$  is optimal. The omitted constraint (25) is satisfied because  $\phi_i(c) = 1$  and  $c > c_m$ .

Next, for  $c^B < c < c_m$ , we claim that  $p(c) = 1$  and  $\phi_i(c) = 1$ . We set  $\mu(c) = 0$ . When  $\phi_i(c) = 1$ , first-order derivative (30) becomes

$$\frac{\partial L}{\partial p} = m_2 \Delta(c) > 0,$$

and it is optimal to set  $p(c) = 1$ . Given  $p(c) = 1$  and  $\mu(c) = 0$ , first-order derivatives (28) and (29) are strictly positive since  $c^B < c$ . Hence, it is optimal to set  $\phi_i(c) = 1$ . The omitted constraint (25) is satisfied because  $p(c) = 1$ .

Finally, for  $c < c < c^B$ , we claim that  $\phi_i(c) = p(c) = 0$ . Given  $p(c) = 0$ , the first-order derivatives (28) and (29) are strictly negative because  $c < c^B$ . Hence it is optimal to set  $\phi_i(c) = 0$ . Given  $\phi_i(c) = 0$ , the first-order derivative (30) is zero. It is optimal to set  $p(c) = 0$ . The omitted constraint (25) is satisfied because  $\phi_i(c) = 0$ .

## References

- Barros, P.P., Olivella, P., 2005. Waiting lists and patient selection. *Journal of Economics & Management Strategy* 14 (3), 623–646.
- Besley, T.J., Coate, S., 1991. Public provision of private goods and the redistribution of income. *American Economic Review* 81 (4), 979–984.
- Cutler, D.M., Gruber, J., 1996. Does public insurance crowd out private insurance? *Quarterly Journal of Economics* 111 (2), 391–430.
- Ellis, R., 1998. Creaming, skimming, and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics* 17 (5), 537–555.
- Epple, D., Romano, R., 1996. Public provision of private goods. *Journal of Political Economy* 104 (1), 57–84.
- Garber, A.M., Phelps, C.E., 1997. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 16, 1–31.
- Grassi, S., Ma, C.-t.A., 2009. Optimal Public Rationing and Price Response. Boston University Economics Working Papers Series.
- Grassi, S., Ma, C.-t.A., 2010. Public rationing and private sector selection. *Journal of Public Economic Theory*, forthcoming.
- Gravelle, H., Siciliani, L., 2008. Ramsey waits: allocating public health service resources when there is rationing by waiting. *Journal of Health Economics* 27 (5), 1143–1154.
- Gravelle, H., Siciliani, L., 2009. Third degree waiting time discrimination: optimal allocation of a public sector healthcare treatment under rationing by waiting. *Health Economics Letters* 18 (8), 977–986.
- Gruber, J., Simon, K., 2008. Crowd-out 10 years later: have recent public insurance expansions crowded out private health insurance? *Journal of Health Economics* 27, 201–217.
- Hoel, M., 2007. What should (public) health insurance cover? *Journal of Health Economics* 26, 251–262.
- Hoel, M., Sæther, E.M., 2003. Public health care with waiting time: the role of supplementary private health care. *Journal of Health Economics* 22, 599–616.
- Iversen, T., 1997. The effect of a private sector on the waiting time in a national health service. *Journal of Health Economics* 16, 381–396.
- Weinstein, M., Zeckhauser, R., 1973. Critical ratios and efficient allocation. *Journal of Public Economics* 2, 147–157.