Working Paper 2021:14

Department of Economics
School of Economics and Management

Multiple Pricing for Personal Assistance Services

Tommy Andersson
Lina Maria Ellegård
Andreea Enache
Albin Erlanson
Prakriti Thami

November 2021
Revised: April 2022

LUND UNIVERSITY

Multiple Pricing for Personal Assistance Services[*]

Tommy Andersson[†], Lina Maria Ellegård[‡], Andreea Enache[§], Albin Erlanson[¶], and Prakriti Thami[‖]

First version: November 5, 2021. This version: April 12, 2022

Abstract

This paper provides a general theoretical framework that captures the essential features of a Swedish reform where private and public health care providers serve patients with certain functional impairments. Because providers receive a fixed hourly compensation for their services (identical across patient types) and only private providers can reject service requests from patients, private providers avoid the costliest patients, resulting in a monetary deficit for public providers. To partially overcome this problem, a multiple pricing (reimbursement) scheme is proposed and its solution is characterized. The results suggest that there are some fundamental trade-offs, e.g., between the goals of containing costs and restricting choices for patients, but that the suggested pricing scheme may substantially reduce the deficits for public providers without affecting the total budget set by the central government.

Keywords: health care services, public and private providers, multiple pricing, welfare, dumping.

JEL Classification: C61, D47, D78, I11.

1 Introduction

A medical treatment for a patient with multiple health problems will normally involve additional costs relative to the same treatment for a patient who is healthy in all aspects except the one that

[*]We are grateful to Kristian Persson Kern at the Swedish Social Insurance Inspectorate for introducing us to this project and for providing all background information to this paper. Andersson acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation (grant no. P2018–0100). Ellegård acknowledges financial support from Forte–The Swedish Research Council for Health, Working Life and Welfare (grant no. 2017–00877).

[†]Department of Economics, Lund University (corresponding author). E-mail: tommy.andersson@nek.lu.se.

[‡]Department of Economics, Lund University.

[§]Department of Economics, Stockholm School of Economics.

[¶]Department of Economics, University of Essex. E-mail: albin.erlanson@essex.ac.uk

[‖]Department of Economics, Lund University.

the patient is being treated for. If the reimbursement scheme (i.e., the "prices") for the health care providers do not take such variations across patients into account, private providers may avoid the costliest patients.[1] This phenomenon, known as "dumping" (see, e.g., Brown et al., 2014; Ellis, 1998; Newhouse, 1996), is undesirable for two intertwined reasons. First, dumping restricts the choice sets of high-cost patients, who are left with the default (typically public) provider. Being able to choose between different providers may be of substantial intrinsic value for patients, especially when it comes to services that are consumed often or over a long duration of time, such as personal assistance services for people with functional impairments. Second, when prices are based on the average costs in the patient population, dumping inflates total costs since third-party payers have to fund both the services provided to the dumped patients and the excess payments to private providers.

It seems intuitive to think that both these dumping related problems could be solved by introducing a multiple pricing scheme that differentiates prices according to "cost types," making high-cost type patients more attractive to serve while reducing the profit margins. This intuition lies behind the design of prospective payment schemes such as Diagnosis Related Group (DRG) reimbursement in hospital care (Hafsteinsdottir and Siciliani, 2010; Kifmann and Siciliani, 2017) and risk-adjustments of payments to health care plans (Ven et al., 2000). Yet, recent literature has suggested that risk adjustment may induce cream-skimming within risk categories (Brown et al., 2014), and may inflate costs if the risk scores are based on information supplied by the provider (Geruso and McGuire, 2016). In both cases, the conclusion utterly depends on the uncertainty surrounding the health needs of a given patient. While uncertainty is at the core of many aspects of health and health care (Arrow, 1963), it is not always a dominant feature. In particular, long-term care services such as personal assistance for functionally impaired or elderly persons are examples of care settings in which the need of a given patient is predictable with small variation. In 2019, such services accounted for on average 1.3 percent of the gross domestic product in OECD countries in 2019, i.e., a significant share of health expenditures.[2]

This paper sets up a very general theoretical model where patients have the option to choose a private or a public provider of personal assistance services. Both types of providers receive a fixed hourly compensation for supplying a given number of hours of care to patients, i.e., a purely prospective payment system with no cost-reimbursement component. The paper attempts to answer the following question: can the total costs be reduced by introducing a multiple pricing scheme where the hourly compensation is identical for both public and private providers, but where the compensation is allowed to be distinct for different "types" of users? To motivate the model, a Swedish policy for personal assistance services (see Section 2) will serve as a leading example throughout the paper, noting that the model and the results are not restricted to this spe-

[1] On some markets, it is a common practice to use nonlinear pricing schemes where prices are not strictly proportional to the quality or quantity purchased. See Oren (2012) or Wilson (1993) for detailed reviews.

[2] https://stats.oecd.org/Index.aspx?ThemeTreeId=9, Dataset: Health expenditure and financing, Financing: All financing schemes, Function: Long-Term Care (health).

cific application. The model is based on the empirical observation that private providers engage in dumping since the hourly compensation is fixed, but the cost of providing assistance services typically varies widely depending on who receives the service. That is, different patients have different needs, but even if they have identical needs, they may still be differently costly to assist, for example, because they live in different geographical areas. To formalize this observation, it is assumed that patients can be classified to be specific cost types, and that the regulator knows the cost type of each patient.[3]

The considered optimization problem allows for the flexibility to set any number k of prices (one price for each k exogenously given intervals of cost types) instead of just one price for all cost types (as in the current Swedish system).[4] The solution to this problem characterizes the pricing scheme that minimizes total costs for any given number of prices. The results demonstrate that as long as all high-cost types must be served, differentiated pricing schemes may give rise to fundamental trade-offs between the goals of containing costs and allowing patients an effective choice between providers. The regulator reduces total costs by increasing the number of prices, but possibly at the expense of decreasing the share of patients that may effectively choose between providers (i.e., the non-dumped patients). Thus, the paper shows that the trade-off between efficiency and risk-selection (see also Newhouse, 1996) does not only arise in the choice between retrospective (cost-based) and prospective payment systems, but also between different prospective payment systems. The model further highlights that refinements of the pricing scheme may be associated with a trade-off regarding the types of patients that are able to effectively choose between different providers. Specifically, the distribution of patients facing an effective choice may include more high-cost types and fewer low-cost types under a refined pricing scheme.

From a policy perspective, the results can be used to investigate how a more flexible pricing scheme with multiple prices affects the monetary loss of the public default provider under different assumptions on cost functions and cost type distributions.[5] A numerical study demonstrates that marginal adjustments in the policy, such as increasing the number of prices from one (as in the current Swedish system) to two, would reduce the loss of the municipalities by roughly 50 percent. The corresponding number when moving from one to ten prices is roughly 90 percent. A back-of-the-envelope calculation shows that the estimated loss for the public providers of 4.45 billion SEK in 2020 could have been reduced to 2.31 billion SEK if the reimbursement schemes

[3]The former of these assumptions captures the above observation that different users may be differently costly to serve (see also Brown et al., 2014; Chernew et al., 2020; Newhouse, 1996). The latter assumption is realistic in the personal assistance setting, as users need to submit detailed information on determinants of costs (e.g., their disabilities, general health status, home address, etc.) when applying for assistance services.

[4]The problem is solved for an integer number k of prices since this is relevant from a policy perspective. However, the model can also be solved for infinitely many prices as, e.g., Mirrlees (1971) did in his seminal continuous optimal taxation framework. See the discussion in Section 5.

[5]The theoretical results are valid under weak assumptions on the distribution of cost types as it suffices that the cost types are distributed according to a cumulative density function F with a strictly positive density f.

had been based on two prices instead of one. The corresponding loss for ten prices is 0.48 billion SEK.

The remaining part of the paper is outlined as follows. Section 2 provides a brief description of the Swedish system for personal assistance services. Section 3 contains an overview of the related literature. Section 4 introduces the theoretical model. The main theoretical and numerical findings are stated in Sections 5 and 6, respectively. Section 7 concludes the paper. All proofs, as well as some technical remarks, are delegated to the Appendix.

2 Motivating Example: Personal Assistance Services in Sweden

Swedish residents with significant and long-term functional impairments have the right to receive personal assistance with activities of daily living.[6] These activities include, among other things, assistance with personal hygiene, food preparation and simpler medical treatments.[7] Since the enactment of the Personal Assistance Law in 1994 (SFS 1993:387, 1993), eligible assistance users may freely choose the provider of their personal assistance services (for a detailed review of this act, see Clevnert and Johansson, 2007). Giving the users the right to choose their provider was a central goal behind the enactment of the law. It was argued that users, who are highly dependent on their assistant, should be allowed to choose their provider, acknowledging the uniqueness of each user-assistant relationship and the difficulty to govern personal services of this kind (SOU 1991:46, 1991, pp. 111).

Public providers (municipalities in this case) are obliged, by law, to accept all assistance requests, but private providers, by contrast, are allowed to reject assistance requests. Providers are reimbursed by the same hourly compensation from the central government. The annual total compensation is capped by the central government, meaning that any deficits have to be covered by the providers. Between 1996 and 2020, the costs of the state-funded assistance services rose in nominal terms from almost 4.2 billion SEK (0.6 percent of the budget of the central government) to 23.5 billion SEK (1.9 percent of the budget).[8] This dramatic increase in costs is largely driven by the increase in the number of people who have been granted assistance and the increase in the number of hours that have been granted to them (ISF 2021:11, 2021). By carefully studying the estimated personnel costs for the private and municipality providers, another interesting pattern emerges: It is plausible that private firms indeed engage in dumping, i.e., that they avoid the costliest users. As can be seen from Figure 1, the average personnel costs for the private firms

[6] For a comprehensive description of the Swedish health care system in general, see Anell et al. (2012). There are several papers analyzing the impact of different price and choice reforms in Swedish healthcare for example Kastberg and Siverbo (2007), Dackehag and Ellegård (2019), Anell et al. (2018), Agerholm et al. (2015) and Dietrichson et al. (2020). However, none of these studies provide a theoretical framework for analysing multiple pricing schemes.

[7] The decisions about eligibility and granted number of hours are taken by the Social Insurance Agency (ISF 2021:11, 2021, p. 9).

[8] See the 2022 Swedish Budget Bill (Proposition 2021/22:1, 2021).

(white bars) are clearly below the fixed hourly compensation from the central government (grey bars) for all years between 2008 and 2020. In the same time period, the average personnel costs for the municipalities (black bars) are clearly above the fixed hourly compensation from the central government (grey bars).[9] Thus, the figure seems to indicate that private firms on average make a profit on each user, whereas the municipalities on average make a loss on each user.[10]

There are no strong reasons to believe that private and public providers have different cost functions in the production of personal assistance services for a given cost type.[11] Although private providers often can be expected to be more concerned about cost efficiency (Hart et al., 1997), the structure of the personal assistance system leaves little room for efficiency improvements: The service package is standardised and consists of manual tasks with little room for productivity improvements. Moreover, the incentives for such improvements are in all circumstances weak, as providers are paid by the hour. Wage rates are collectively bargained and thus do not depend on whether the employer is private or public. For a user of a given cost type, the cost of providing services should therefore be quite similar. Although wage rates are regulated, they do differ between more and less skilled staff categories. The observed difference in personnel costs indicates that municipalities employ personnel with more experience, more advanced training and education, i.e., exactly the type of personnel that is required to treat users with more severe health issues. In other words, it looks as though private providers earn their profits by dumping high-cost users to the public providers.

3 Related Literature

The payment structure considered in this paper is an instance of a prospective payment system. In such systems, a third-party payer (e.g., the government) sets prices, and possibly the base of payment (e.g., hours granted), in advance (Jegers et al., 2002). In comparison with retrospective payment, which is based on "reasonable costs," prospective payment strengthens the incentives for the providers to improve production efficiency (Shleifer, 1985).[12] However, the theoretical literature (see, e.g., Newhouse, 1996) has long acknowledged that this form of payment generates incentives to screen the pool of potential users for low cost types (risk-selection). In this strand of literature, profit-maximizing, possibly semi-altruistic, health care providers select the

[9] The calculations of the estimated personnel costs is based on data from Statistics Sweden and the Swedish Tax Agency. See also the recent report by the Swedish Social Insurance Inspectorate (ISF 2021:11, 2021)

[10] There are, of course, also non-personnel related costs (e.g., for transportation and medical equipment), but these costs are, for any given user, likely to be almost identical for private and municipality providers.

[11] Accordingly, the government investigation behind the law did not argue that private providers would be able to produce services at lower costs. The investigation did express hopes about yardstick competition (SOU 1991:46, 1991, pp. 112).

[12] Other structures of payment can be considered. The recent surge in availability of detailed patient data, together with increased computational capacity, suggest that individual-level auctions may be an alternative way to stimulate production efficiency while minimizing risk-selection (Montanera et al., 2021). In practice, the notion of auctioning out the care for disabled individuals to the lowest bidder may sound unattractive to the general public.

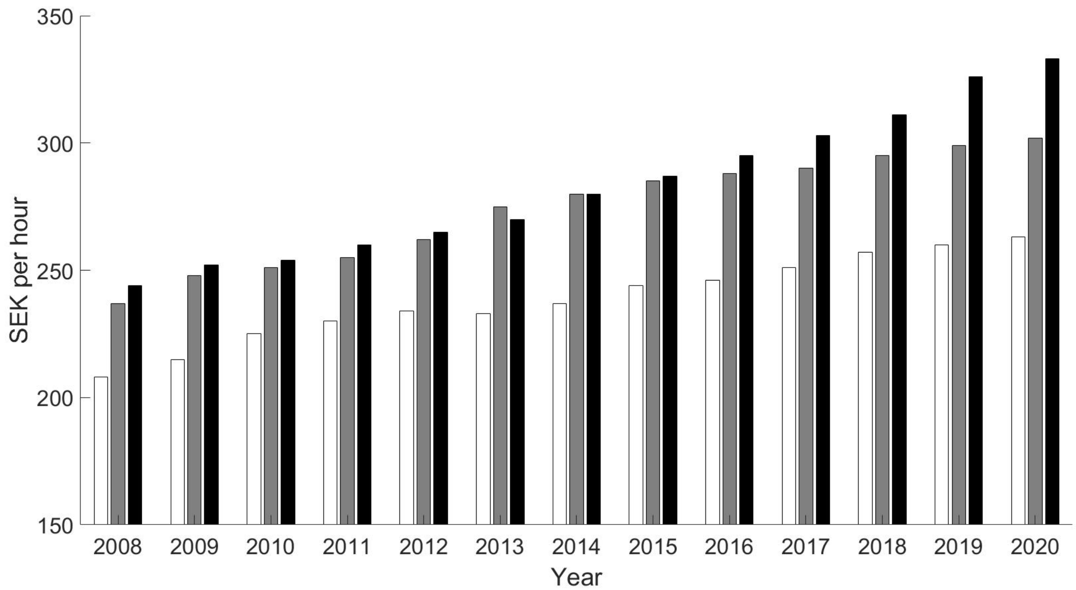

Figure 1: The average personnel costs (in SEK) for the private firms (white bars), the fixed hourly compensation from the central government (grey bars) and the average personnel costs for the municipalities (black bars) between year 2008 and 2020.

quantity and quality of care supplied to users of different cost types (Dranove, 1987; Allen and Gertler, 1991; Ma, 1994; Ellis, 1998; Eggleston, 2000). Providers will attract low cost types ("cream-skimming") and underserve ("skimp") or avoid ("dump") high cost types. Researchers have suggested that incentives for risk-selection can be mitigated by introducing partial cost reimbursement (Barros, 2003; Eggleston, 2005; Kifmann and Lorenz, 2011) or refining the pricing scheme to better reflect variation in risk, a practice known as risk-adjustment (Ven et al., 2000; Ellis et al., 2018).[13]

Notably, this literature generally ignores the cost of providing services to the dumped patients (Newhouse, 1996). By contrast, these costs are central to the model introduced in the present paper where the service package delivered to each user ("quality") is fixed *ex ante* and the cost to provide the service varies across cost types. The main economic problem facing the third-party payer in our setting is that the dumping behavior of private providers inflates costs in return for nothing: Profits only serve as a signal of risk-selection, not as a signal of productivity or ability to innovate. Thus, the regulator seeks to minimize private profits. This objective is also common in the branch of the industrial organization literature that investigates how to optimally regulate profit maximizing firms (Laffont and Tirole, 1993). In the context of this paper, the objective can be motivated by a concern about the dead-weight loss from raising the public funds to finance

[13]Some authors have suggested that the third-party payer should opt for other price levels than the expected costs to address other problems such as adverse selection (Glazer and McGuire, 2000). Such objectives are beyond the scope of the present study.

the services that, under all circumstances, will have to be delivered to the high-cost types. With the exception of Eggleston (2000), we are not aware of any other paper that has considered the dead-weight loss from excess spending in relation to risk-selection.

The study by Brown et al. (2014) is most closely related to this paper. They also investigate the consequences in a healthcare market with prospective payments when changing from one price to multiple (risk-adjusted) prices. Similarly to the model considered in this paper, their focus is on providers and patients play no active role in the analysis, the budget period is fixed and the budget allocated by the third-party payer does not change with the number of prices. In contrast to this paper, they study a market for overall healthcare financed via the federal U.S. program Medicare. This program includes all healthcare services, and providers that enroll patients need to cover all expenses associated with their patients. This significantly increases the uncertainty of the cost associated with a patient. They model this uncertainty by assuming that each patient type has an expected cost with a variance that is increasing with the expected cost (providers can screen patients at a cost). This additional component of their model leads to a similar pattern as in this paper. Namely, providers screen patients and only enroll the low-cost patients in each risk-adjustment class. As in this study, Brown et al. (2014) also find that profits for private firms decrease with the number of prices. Even if their conclusions are similar to ours, the models are different and our optimization problem and objectives are different. Because of these differences, this paper characterizes a different multiple pricing scheme. Further, Brown et al. (2014) do not recognize the trade-off between cost containment (total costs decrease with the number of prices also in their paper) and the share of patients that are able to choose between different providers.

Geruso and McGuire (2016) is another closely related study. The authors point out that the use of risk adjustment to limit cream-skimming in a prospective payment system may reintroduce cost-containment problems, if the risk scores are based on information (such as previous diagnoses or procedures) reported by the provider. In difference to the present study and Brown et al. (2014), Geruso and McGuire (2016) thus find that multiple prices may not lead to lower total costs. Geruso and McGuire's paper fit well to settings where the provider sits on exclusive information about patients' cost type, such as hospital care services. Our model suits better to settings in which the service package can be is determined ex ante by a third party, which is the case for means-tested personal assistance services or home help services more generally.

Another related strand of literature considers multiple prices in the context of differentiating prices across medical procedures. In difference to the model considered in this paper, these studies do not consider a fixed service package. In a study by Hafsteinsdottir and Siciliani (2010) providers are not allowed to dump patients and the third-party payer considers the marginal cost of public funds. However, they do not discuss risk-selection.

4 Model and Preliminaries

For convenience, and without loss of generality, it is assumed that there is a continuum of patients. Each patient can be described by a cost type θ, normalized to belong to the interval $[0, 1]$. The cost types are distributed on this interval according to the cumulative distribution function F with a strictly positive density function f. Patients can be served by the municipality or by a profit maximizing private firm. Both types of providers can serve as many patients as they desire (there are no capacity constraints). Private firms can decline to serve patients. Any patient not served by a private firm must be served by the municipality.

A cost function $c : [0, 1] \to \mathbb{R}_+$ specifies the cost for serving a patient of cost type θ. This function is increasing and differentiable, i.e., $c'(\theta) > 0$ for all $\theta \in [0, 1]$. That the cost function is increasing follows by the definition of cost types since a higher cost type is associated with a higher cost. The assumption that private firms and the municipality have identical cost functions is rather mild since there are no strong reasons to believe that their cost structures are very different for the type of application used to motivate the theoretical framework (see Section 2).

A regulator determines the prices that the municipality and the private providers receive for serving patients. More specifically, the regulator partitions the support of cost types $[0, 1]$ into k sub-intervals. Let this partition be given by the vector $\overline{\theta} = (\overline{\theta}_0, \overline{\theta}_1, \dots, \overline{\theta}_k)$ where $\overline{\theta}_{j-1} < \overline{\theta}_j$ for all $j \in \{1, \dots, k\}$, and $\overline{\theta}_0 = 0$, $\overline{\theta}_k = 1$. It now follows that:

$$[0, 1] = \cup_{j=1}^{k-1} [\overline{\theta}_{j-1}, \overline{\theta}_j) \cup [\overline{\theta}_{k-1}, \overline{\theta}_k].$$

By defining $\overline{\theta}_j = \underline{\theta}_{j+1}$ for all $j = 0, \dots, k-1$, a sub-interval j can also be written as $[\underline{\theta}_j, \overline{\theta}_j)$. Let $p = (p_1, \dots, p_k)$ denote the price vector set by the regulator. Here, $p_j \in [c(\underline{\theta}_j), c(\overline{\theta}_j)]$ represents the price for all cost types in sub-interval j. Note also the difference to the current Swedish system where there is a single price in the entire cost type interval. This situation is represented by the special case when $k = 1$.

There is a budget B available to the regulator. This budget is set by the central government for the fixed time period of interest. For simplicity, it will be assumed that the available amount of money is enough to cover the cost of serving the patients (a situation where the budget is smaller than the total costs can be incorporated in the model by adding a constant κ, that represents the "deficit," to equation (1)). Formally:

$$B = \int_0^1 c(\theta) f(\theta) d\theta. \tag{1}$$

The total amount of money payed out to the private providers and the municipality cannot exceed the budget B. Consequently, the regulator is restricted by the following budget feasibility

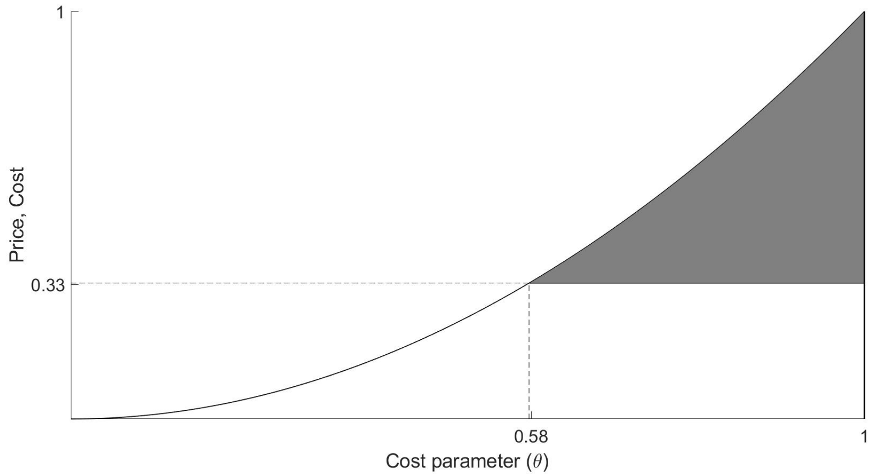

Figure 2: The cost function c is represented by the increasing function. For the single price $p = c^{-1}(0.58) = 0.33$, the private firms will only serve cost types in the interval $[0, 0.58]$. The loss for the municipality is represented by the grey area.

constraint:

$$\sum_{j=1}^{k} p_j[F(\overline{\theta}_j) - F(\underline{\theta}_j)] = p_k - \sum_{j=1}^{k-1}[p_{j+1} - p_j]F(\overline{\theta}_j) \leq B. \tag{2}$$

The objective of the regulator is to determine the prices p to minimize the loss incurred by the municipality on patients that are not served by the private firms. To formally define this objective, consider any given sub-interval $[\underline{\theta}_j, \overline{\theta}_j)$ with an associated price $p_j \in [c(\underline{\theta}_j), c(\overline{\theta}_j)]$. Since the cost function is increasing, there is a cutoff $c^{-1}(p_j)$ for the sub-interval j where the private firms serve all patients on which they can make a profit. That is, the patients in the interval $[\underline{\theta}_j, c^{-1}(p_j)]$. Because private firms engage in dumping, the patients that belong to the interval $(c^{-1}(p_j), \overline{\theta}_j)$ must by served by the municipality. This logic is graphically illustrated in Figure 2 where the cost function is represented by the increasing function. Moreover, there is a single price in the entire cost type interval $[0, 1]$. Note now that if the price is given by $p = c^{-1}(0.58) = 0.33$, the private firms will only serve patients with cost types in the interval $[0, 0.58]$ as these are the only types where the price exceeds the cost of serving them.

This structure also implies that the loss for the municipality in each sub-interval equals the mass of patients in the interval $(c^{-1}(p_j), \overline{\theta}_j)$ multiplied with the loss, i.e., $c(\theta) - p_j$, of each patient in that interval. In Figure 2, this loss is represented by the grey area defined by the difference between the cost function and the price $p = 0.33$ in the interval $(0.58, 1]$. Now, by summing over each sub-interval, the total loss for the municipality is given by:

$$\pi(\overline{\theta}) = \sum_{j=1}^{k} \int_{c^{-1}(p_j)}^{\overline{\theta}_j} [c(\theta) - p_j] f(\theta) d\theta. \tag{3}$$

The regulators objective is to choose prices p to minimize the loss as defined in equation (3), subject to budget feasibility as defined by equation (2). Such objective is in line with the large literature on regulation that appeared in the 1980s (nicely summarized in the book by Laffont and Tirole, 1993). Similar objectives have been considered in the context of prospective payment schemes (Eggleston, 2000).

5 Theoretical Results

As described in the above, the problem for the regulator is to set the prices in order to minimize the loss for the municipality. Given that the sub-intervals are fixed (and not a variable that the regulator optimizes over), the regulator thus needs to solve the following optimization problem:

$$\begin{aligned} \min_{p \in \prod_{j=1}^{k} [c(\underline{\theta}_j), c(\overline{\theta}_j)]} \quad & \sum_{j=1}^{k} \int_{c^{-1}(p_j)}^{\overline{\theta}_j} [c(\theta) - p_j] f(\theta) d\theta \\ & \text{subject to:} \\ & p_k - \sum_{j=1}^{k-1} (p_{j+1} - p_j) F(\overline{\theta}_j) \leq B. \end{aligned} \tag{4}$$

Because this is a convex problem (see the Appendix), it can be analyzed via its Lagrangian and the corresponding Karush–Kuhn–Tucker (KKT, henceforth) conditions. In other words, the KKT conditions arising from the Lagrangian are both necessary and sufficient for an optimal solution. The Lagrangian is defined as follows:

$$L(p, \lambda) = \sum_{j=1}^{k} \int_{c^{-1}(p_j)}^{\overline{\theta}_j} [c(\theta) - p_j] f(\theta) d\theta + \lambda \Big(p_k - \sum_{j=1}^{k-1} (p_{j+1} - p_j) F(\overline{\theta}_j) - B \Big).$$

The KKT conditions for the above Lagrangian are derived in the Appendix. To obtain a characterization of the multiple pricing scheme, it is next observed that the KKT conditions provide the following expression for λ for all $j = 1, \ldots, k$, given the presumption that $c^{-1}(p_j) \in (\overline{\theta}_{j-1}, \overline{\theta}_j)$:

$$\lambda = \frac{F(\overline{\theta}_j) - F(c^{-1}(p_j))}{F(\overline{\theta}_j) - F(\overline{\theta}_{j-1})}. \tag{5}$$

The interpretation of λ reveals important properties of the multiple pricing scheme. Namely, the proportion of patients served by the municipality is constant in each sub-interval j and, furthermore, that λ can be used to derive the prices. The latter finding is illustrated in the following example.

Example 1. Suppose that θ is uniformly distributed on the interval $[0, 1]$, i.e., $F(\theta) = \theta$ for any $\theta \in [0, 1]$, and that the cost function is given by $c(\theta) = \theta^2$. This implies that $c^{-1}(\theta) = \sqrt{\theta}$ and, consequently, that $c^{-1}(p_j) = \sqrt{p_j}$. Hence, condition (5) can be written as:

$$\lambda = \frac{\overline{\theta}_j - \sqrt{p_j}}{\overline{\theta}_j - \overline{\theta}_{j-1}}. \tag{6}$$

From this equation, it follows that $\lambda \in (0, 1)$ for any $p_j \in (\overline{\theta}^2_{j-1}, \overline{\theta}^2_j)$ since p_j is defined by $c(\theta) = \theta^2$ for some $\theta \in (\overline{\theta}_{j-1}, \overline{\theta}_j)$. Solving equation (6) for p_j yields:

$$p_j = (\lambda\overline{\theta}_{j-1} + \overline{\theta}_j(1 - \lambda))^2. \tag{7}$$

Using the same arguments as in the above, the conclusion that $\lambda \in (0, 1)$ and the identity $\underline{\theta}_j = \overline{\theta}_{j-1}$, it now follows that $p_j > c(\underline{\theta}_j)$. This means that private firms will serve some (but not all) patients in each sub-interval j. Furthermore, prices p_j are monotonically increasing in the sense that $p_{j+1} > p_j$ for all $j = 1, \dots, k - 1$. This can also be seen using equation (7) as this monotonicity condition holds if and only if:

$$[\lambda\overline{\theta}_j + \overline{\theta}_{j+1}(1 - \lambda)]^2 > [\lambda\overline{\theta}_{j-1} + \overline{\theta}_j(1 - \lambda)]^2.$$

But this inequality is satisfied by the above conclusion that $\lambda \in (0, 1)$ and by the construction that $\overline{\theta}_{j+1} > \overline{\theta}_j > \overline{\theta}_{j-1}$. □

As it turns out, the main insights from Example 1 hold in general as revealed by the following theorem.

Theorem 1. The optimal price vector p is implicitly given by the KKT conditions in equation (5). Furthermore, $p_j > c(\underline{\theta}_j)$ for each $j = 1 \dots, k$.

The interpretation of the next result is that private firms engage in dumping and, consequently, only serve the patients on which they will be able to make a profit. This finding is in line with what previously has been observed in the literature (see, e.g., Brown et al., 2014; Chernew et al., 2020; Eggleston, 2000; Newhouse, 1996).

Corollary 1. For each price and each sub-interval j, the private firms serve all patients with θ-types in the non-empty interval $[\underline{\theta}_j, c^{-1}(p_j)]$.

Theorem 1 characterizes the multiple price scheme, but it does not reveal anything about potential trade-offs when increasing the number of prices. Before theoretically and numerically investigating this is somewhat more detail, it will be instructive to first continue analyzing Example 1. Figures 2–4 provide a graphical representation of the numerical solution to the optimization problem for the cases with one, two and five prices, respectively. The loss for the municipality is given by the sum of the grey areas in each figure.

By comparing the figures, it is evident that when cost types are discretized into sub-intervals with distinct prices, private firms serve the "lower cost types" within each sub-interval rather than the "lowest cost types" overall. At the same time, private firms will also dump some of the "low cost types" that they would have served under the the uniform price policy and start serving some of the "high cost types" that they would have dumped under the uniform price policy (this also captures the main insights of Corollary 1). This demonstrates that there is a trade-off regarding the types of patients that are able to effectively choose between different providers. Specifically, the distribution of patients facing an effective choice may include more high-cost types and fewer low-cost types under a refined pricing scheme. Further, the loss of the municipality (i.e., the sum of the grey areas in the figures) is decreasing in the number of prices. Thus, the multiple pricing policy that incentivize private firms to serve some of the "higher cost type" patients will not only result in lower losses for the municipality (without affecting the total budget set by the central government), but it will also affect which patients that have an effective choice.

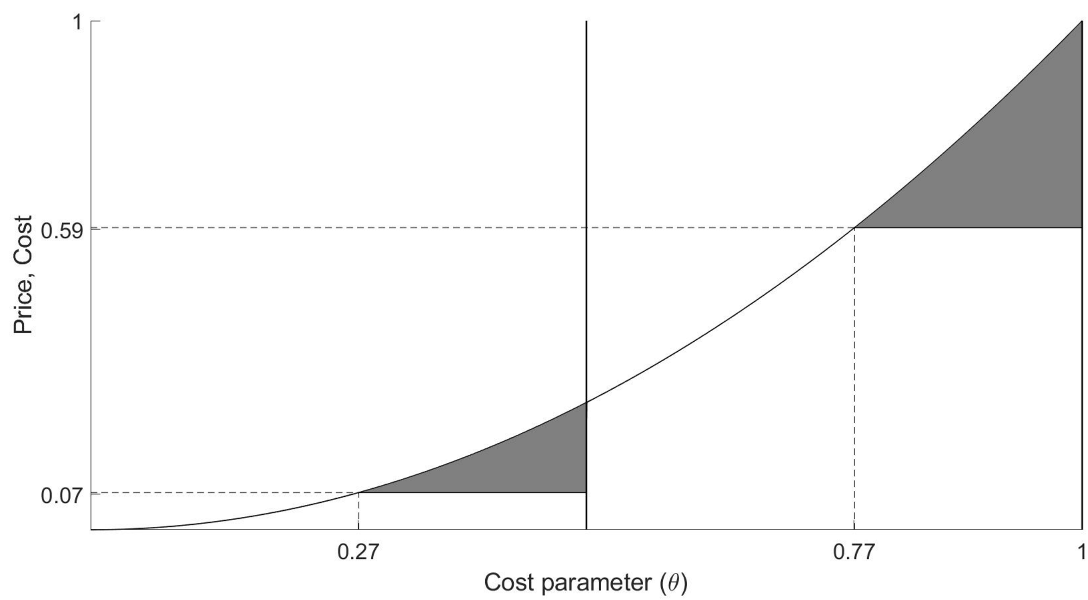

Figure 3: The cost function is represented by the increasing function. The two prices are given by $p^* = (0.07, 0.59)$, and the loss for the municipality is represented by the sum of the two grey areas.

A general insight is that if the regulator is allowed to set infinitely many prices and, more pre-

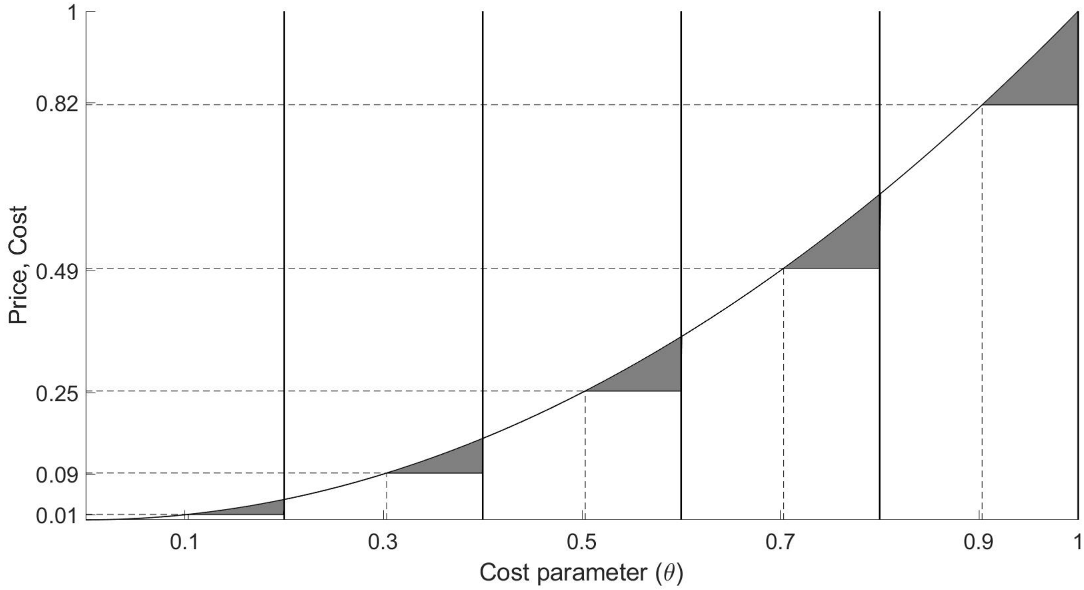

Figure 4: The cost function is represented by the increasing function. The five prices are given by $p^* = (0.01, 0.09, 0.25, 0.49, 0.82)$, and the loss for the municipality is represented by the sum of the five grey areas.

cisely, one price for each cost type θ, the loss for the municipality will be zero as the optimization problem gives a price p_θ for any $\theta \in [0, 1]$ that is equal to the cost $c(\theta)$ of serving a patient with cost type θ. But even if the assumption that there is enough funds available to cover the costs of serving the patients is relaxed, the loss that results from solving the optimization problem will be minimal in the sense that it will equal some constant κ that represents the shortage of money in the system (see also the discussion in Section 4). However, from a policy perspective, it is not feasible to set infinitely many prices as such pricing scheme may be perceived as too complicated by the service providers. So, from a policy perspective, it is important to investigate how loss for the municipality is affected and how the choice sets of the patients are affected when a multiple pricing scheme is implemented. It is, however, difficult to derive such results given the weak assumptions on, e.g., the cost function, the type distributions, and the partition of sub-intervals. Hence, to obtain such results, additional assumptions must be imposed on the model (as in Example 1 and Figures 2–4) or the model needs to be evaluated numerically. This approach is also adopted in the remaining part of the paper. In particular, because the numerical results will reveal (see Section 6) that the marginal benefit for the municipality is highest when the number of prices is increased from one to two, the following two propositions focus on this case.

The first proposition states that if the two cost type intervals contains equally many patients, then the loss of the municipality always decreases when adding an additional price to the statue-quo situation with only one price. Note that this conclusion holds for any increasing cost function and any cumulative density function F with a strictly positive density f.

Proposition 1. Suppose that each sub-interval contains an equal mass of cost types. If the number of prices is increased from one to two, then the loss of the municipality decreases.

The next result reveals something about the choice set for the patients when the number of prices increases from one to two. As it turns out, this depends on the shape of the cost function. More precisely, if the cost function is increasing with an increasing (decreasing) rate, then the proportion of patients served by the municipality increases (decreases) with the number of prices. Consequently, whether or not the choice set expands for the patients depends on the shape of the cost function.

Proposition 2. Suppose that each sub-interval contains an equal mass of cost types and that θ is uniformly distributed on the interval $[0, 1]$. If the number of prices is increased from one to two, then the proportion of patients served by the municipality increases, decreases and is unchanged if $c''(\theta) > 0$, $c''(\theta) < 0$ and $c''(\theta) = 0$, respectively.

6 Numerical Results

If the cost function and type distribution are known to the regulator, the optimization problem can without any difficulties be solved numerically. This is next illustrated by presenting some numerical results under varying assumptions about the cost function, the cost type distribution and the number of sub-intervals.[14] Three potential forms of cost functions and cost type distributions are considered. The cost function is allowed to vary such that:

$$c(\theta) = \theta^{\beta} \text{ where } \beta \in \{0.5, 1, 2\}.$$

The considered cost type distributions $f(\theta)$ are given by the uniform distribution $U(0, 1)$ on the interval $[0, 1]$, the Student's-t distribution $t(1)$ with one degree of freedom, and the Gamma distribution $\Gamma(2, 2)$ with both the shape and scale parameters equal to two. The latter two distributions are truncated in the interval $[0, 1]$. Throughout this section, the sub-intervals for any given problem are partitioned such that they have equal probability mass.

The results are presented in Tables 1 and 2. Each row in the tables presents the result from using the cost function and cost type distribution specified in the corresponding row of the first and second columns. Each column presents the results from using the specified number k of sub-intervals in the objective function or, equivalently, the number of prices optimized over.

The losses for the municipality are displayed in Table 1 and they are presented as a proportion of the total budget B set by the central government. For example, the loss for the quadratic cost

[14] The model was solved using the inbuilt 'fmincon' function from the MATLAB Optimization Toolbox. For more information see: https://se.mathworks.com/help/optim/ug/fmincon.html.

function and the uniform distribution is given by 0.3849. This means that the monetary loss for the municipality is in the order of magnitude of 38.49 percent of the total budget set by the central government. There are a few key takeaways from the Table 1. First, even the introduction of one additional price reduces the loss incurred by roughly half across all specifications. Second, while the exact magnitude of loss varies considerably depending on the specification, loss is found to be monotonically decreasing in the number of prices also for $k > 2$ (Proposition 1 only covers the case when $k = 2$). For instance, given a quadratic cost function and uniformly distributed cost types, loss decreases from being equivalent to approximately 38.5 percent of the budget to 18.8 percent when there are two prices. It decreases further to 7.5 percent when there are five prices and to an even lower 0.75 percent when there are fifty prices.

Table 2 displays the proportion of patients served by the municipality. The results reveal the same pattern as predicted in Proposition 2 but for $k > 2$ prices, i.e., the shape of the cost function determines whether or not the municipality will serve more patients when the number of prices increases.

Table 1: Social loss as a proportion of the total budget

Cost function	Distribution	$k=1$	$k=2$	$k=5$	$k=10$	$k=50$
$c(\theta)=\sqrt{\theta}$	$U(0,1)$	0.1481	0.0795	0.0338	0.0174	0.0036
	$t(1)$	0.1602	0.0865	0.0367	0.0189	0.0039
	$\Gamma(2,2)$	0.0890	0.0500	0.0224	0.0120	0.0027
$c(\theta)=\theta$	$U(0,1)$	0.2500	0.1250	0.0500	0.0250	0.0050
	$t(1)$	0.2724	0.1401	0.0565	0.0283	0.0057
	$\Gamma(2,2)$	0.1607	0.0853	0.0359	0.0184	0.0038
$c(\theta)=\theta^2$	$U(0,1)$	0.3849	0.1888	0.0751	0.0375	0.0075
	$t(1)$	0.4219	0.2231	0.0911	0.0457	0.0092
	$\Gamma(2,2)$	0.2699	0.1345	0.0537	0.0268	0.0054

Finally, the findings from the numerical study are corroborated by a back-of-the envelope calculation using Swedish data. In 2020, the costs of the state-funded assistance services was 23.53 billion SEK and the estimated loss for the municipalities was 4.45 billion SEK (ISF 2021:11, 2021). Suppose now that the cost types are uniformly distributed on $[0, 1]$. The parameters α and β in the cost function $c(\theta) = \alpha\theta^\beta$ are calibrated such that the model (when optimized) reports a loss of 4.45 billion SEK.[15] Using the calibrated cost parameters and continuing with the assumption of uniformly distributed cost types, the optimization problem was solved to identify the prices and the corresponding losses for two, five, ten and fifty prices. If there are only two prices, the prices are given by $p_1 = 192$ and $p_2 = 416$ (in 2020, the uniform price in Swe-

[15]Because θ is uniformly distributed on $[0, 1]$ and the total budget is 23.53 billion SEK, the budget was normalized when solving the optimization problem. This didn't affect the calculations that are presented in this paragraph.

Table 2: Proportion of users served by the municipality

Cost function	Distribution	$k=1$	$k=2$	$k=5$	$k=10$	$k=50$
$c(\theta)=\sqrt{\theta}$	$U(0,1)$	0.5556	0.5408	0.5265	0.5189	0.5085
	$t(1)$	0.5318	0.5272	0.5199	0.5150	0.5072
	$\Gamma(2,2)$	0.5798	0.5633	0.5463	0.5367	0.5213
$c(\theta)=\theta$	$U(0,1)$	0.5000	0.5000	0.5000	0.5000	0.5000
	$t(1)$	0.4709	0.4842	0.4935	0.4967	0.4993
	$\Gamma(2,2)$	0.5427	0.5324	0.5216	0.5156	0.5070
$c(\theta)=\theta^2$	$U(0,1)$	0.4226	0.4592	0.4834	0.4917	0.4984
	$t(1)$	0.3867	0.4375	0.4740	0.4870	0.4975
	$\Gamma(2,2)$	0.4839	0.4915	0.4964	0.4981	0.5009

den was 304 SEK) and the loss is estimated to 2.31 billion SEK. For five, ten and fifty prices, the corresponding losses under the multiple pricing policy are 0.95, 0.48 and 0.10 billion SEK, respectively.

7 Conclusion and Discussion

This paper has modelled a situation where private and municipality health care providers receive the same hourly compensation for serving patients that varies in their cost to serve. Since only private providers can reject service requests, it is empirically plausible that they engage in dumping and, consequently, reject service requests they find unprofitable. As a consequence, private firms make profits at the expense of the municipality budgets (and ultimately the tax payers). To mitigate such unintended consequences, this paper has proposed a multiple pricing scheme (see also Brown et al., 2014; Chernew et al., 2020), and demonstrated that this may substantially reduce the loss for municipalities without affecting the total budget provided by the central government. This conclusion holds even for the smallest possible modification of the pricing scheme where only one additional price is introduced.

It is important to note that the investigated model is tailored to care types such as home help services for disabled or elderly individuals, i.e., settings in which the need of a given patient is recurrent and predictable with small variation. In many other parts of health care, the needs of the patients are less predictable/verifiable, and providers have more freedom to decide what and how much treatment to provide. In such settings, public health care providers responsible for patients dumped by private providers may avoid the deficit problem studied in this paper by instead skimping on the care they provide. The deficit problem studied in this paper may thus be less important than the problem of ensuring sufficient care to high-cost patients.

An ideal study would have access to relevant micro data to estimate the cost function and the cost type distribution. Because we currently don't have access to such data, these important estimations have to be postponed to a future research project. Furthermore, the theoretical model is based on, at least, two simplifying assumptions.

First, the patients are treated as substitutes in the model. In reality, different patients complement each other in different ways for both private providers and the municipality. For example, two identical patients may individually be regarded as "high cost types" simply because they both live in a remote geographical area. But if they both would live in the same remote area, they may jointly be regarded as "two low cost types" by a service provider as the travel cost of serving them is reduced by half if the same provider serves both these patients. Dealing with complementarities in this way is an inherently difficult problem (see, e.g., Amir, 2005; Milgrom and Roberts, 1995).

Second, an implicit assumption in the paper is that the patients first approaches the private firms and if they are rejected, they will be served by the municipality. In reality, municipalities also serve low cost types. Our approach is thus a simplification and a more realistic approach would be to integrate the considered model with the search and matching literature to also model how patients and service providers are matched (e.g., Eeckhout and Kircheri, 2010; Pissarides, 2000; Shimer, 2005). Also this theoretical extension of the model is left for future research.

Appendix: Proofs and Technical Remarks

This Appendix starts by deriving the KKT conditions. It is then demonstrated that the the optimization problem is convex and therefore the KKT conditions are both necessary and sufficient for global optimality. These two steps jointly prove Theorem 1 and Corollary 1. The proofs of Propositions 1 and 2 can be found at the end of the Appendix.

The KKT Conditions

Begin with interchanging the order of integration in the Lagrangian:

$$L(p,\lambda) = \sum_{j=1}^{k} \int_{\overline{\theta}_j}^{c^{-1}(p_j)} [p_j - c(\theta)] f(\theta) d\theta + \lambda\Big(p_k - \sum_{j=1}^{k-1} (p_{j+1} - p_j) F(\overline{\theta}_j) - B\Big).$$

Differentiating L with respect to p_j generates the following first-order conditions (FOCs):

$$\begin{aligned}
\frac{\partial}{\partial p_j} L(p, \lambda) &= \frac{d}{dp_j}\left(\int_{\overline{\theta}_j}^{c^{-1}(p_j)} [c(\theta) - p_j] f(\theta) d\theta\right) - \lambda(F(\overline{\theta}_{j-1}) - F(\overline{\theta}_j)), \\
&= [p_j - c(c^{-1}(p_j))] f(c^{-1}(p_j)) \frac{1}{c'(c^{-1}(p_j))} - [p_j - c(\overline{\theta}_j)] f(\overline{\theta}_j) \cdot 0 + \int_{\overline{\theta}_j}^{c^{-1}(p_j)} f(\theta) d\theta \\
&\quad -\lambda(F(\overline{\theta}_{j-1}) - F(\overline{\theta}_j)), \\
&= F(c^{-1}(p_j)) - F(\overline{\theta}_j) - \lambda(F(\overline{\theta}_{j-1}) - F(\overline{\theta}_j)), \\
&= F(c^{-1}(p_j)) - \lambda F(\overline{\theta}_{j-1}) + (\lambda - 1) F(\overline{\theta}_j).
\end{aligned}$$

Next, set the FOCs equal to zero:

$$\lambda(F(\overline{\theta}_j) - F(\overline{\theta}_{j-1})) = F(\overline{\theta}_j) - F(c^{-1}(p_j)).$$

By simplifying the above condition, the following expression for λ can be obtained, given that $c^{-1}(p_j) \in (\overline{\theta}_{j-1}, \overline{\theta}_j)$:

$$\lambda = \frac{F(\overline{\theta}_j) - F(c^{-1}(p_j))}{F(\overline{\theta}_j) - F(\overline{\theta}_{j-1})}. \tag{8}$$

Finally, it is demonstrated that that $\lambda \in (0, 1)$ which implies that $c^{-1}(p_j) > c(\underline{\theta}_j)$ for each j. First, if $\lambda = 0$ then, by complementary slackness, the budget constraint is not binding. But this is impossible since it is assumed that $B = \int_0^1 c(\theta) f(\theta) d\theta$. Therefore, it must be the case that $\lambda > 0$. Suppose instead that $\lambda = 1$. This implies that, for each j, $F(c^{-1}(p_j)) = F(\overline{\theta}_{j-1})$ and, consequently, that $p_j = c(\underline{\theta}_j)$. At such prices, no private firms are serving any users and not all of the budget is used. Clearly, this cannot be optimal because by increasing one price in some interval, the loss will be reduced. Thus, $0 < \lambda < 1$ and the optimal price vector is such that some patients will be served by private firms in each sub-interval j, i.e., $c^{-1}(p_j) > \underline{\theta}_{j-1}$ for each $j = 1, \ldots, k$.

Necessity of KKT Conditions for Global Optimality

To guarantee that the KKT conditions in equation (8) are also sufficient, it must be demonstrated that the objective function is convex and that the feasible set is convex. To see that the feasible set is convex, take any $\alpha \in [0, 1]$ and any two feasible price vectors p and p'. Then $q = \alpha p + (1-\alpha)p'$ is a feasible price vector, since the price vectors αp and $(1 - \alpha)p'$ satisfies the inequality in equation (4) with αB and $(1-\alpha)B$, respectively. Hence, the convex combination $q = \alpha p + (1 - \alpha)p'$ satisfies equation (4) meaning that the feasible set is convex.

Next, it is established that the loss function is convex. Note that the loss function, defined in equation (3), is the sum of k one-dimensional functions. It suffices to show that each of the k functions is convex. Thus, for any $1 \leq j \leq k$, it needs to be demonstrated that:

$$\varphi(p_j) = \int_{\overline{\theta}_j}^{c^{-1}(p_j)} [p_j - c(\theta)] f(\theta) d\theta,$$

is convex. Since φ has well-defined second-order derivatives, it suffices to investigate the second-order derivative. Recall, from the previous analysis, that:

$$\varphi'(p_j) = F(c^{-1}(p_j)) - F(\overline{\theta}_j).$$

Note that $\varphi' \leq 0$, but that φ is increasing in p_j. This makes sense, since by setting $p_j = \overline{\theta}_j$, the losses in sub-interval j is zero, but this is too costly in general. Next, differentiate φ a second time:

$$\varphi^{''}(p_j) = f(c^{-1}(p_j)) \frac{1}{c'(c^{-1}(p_j))}.$$

Hence, φ is strictly convex if and only if $\varphi^{''}(p_j) > 0$. But this is clearly true since $f(\theta) > 0$ and because $c'(\theta) > 0$ by assumption. Therefore, the problem is convex and the KKT conditions are necessary and sufficient for a global minimum.

Proof of Proposition 1

Let p_j^k denote the price in sub-interval j when there are k prices, and let $\hat{\theta}_j^k$ denote the cost type that determines price p_j^k when there are k prices (i.e., $\hat{\theta}_j^k = c^{-1}(p_j^k)$). Note first that $p_1^2 < p_1^1 < p_2^2$ and $\hat{\theta}_1^2 < \hat{\theta}_1^1 < \hat{\theta}_2^2$ by budget balance. Note next that the loss for the municipality when there are one and two prices are given by:

$$\begin{aligned}
\pi^1 &= \int_{\hat{\theta}_1^1}^{\overline{\theta}_1^1} (c(\theta) - c(\hat{\theta}_1^1)) f(\theta) d\theta, \\
\pi^2 &= \int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} (c(\theta) - c(\hat{\theta}_1^2)) f(\theta) d\theta + \int_{\hat{\theta}_2^2}^{\overline{\theta}_2^2} (c(\theta) - c(\hat{\theta}_2^2)) f(\theta) d\theta,
\end{aligned}$$

respectively. Because π^j represents a loss (expressed as a non-negative number), it needs to be proved that $\pi^2 < \pi^1$, i.e., that:

$$\int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} (c(\theta) - c(\hat{\theta}_1^2)) f(\theta) d\theta + \int_{\hat{\theta}_2^2}^{\overline{\theta}_2^2} (c(\theta) - c(\hat{\theta}_2^2)) f(\theta) d\theta < \int_{\hat{\theta}_1^1}^{\overline{\theta}_1^1} (c(\theta) - c(\hat{\theta}_1^1)) f(\theta) d\theta.$$

By using the facts that $\overline{\theta}_1^1 = \overline{\theta}_2^2 = 1$ and $\hat{\theta}_1^2 < \hat{\theta}_1^1 < \hat{\theta}_2^2$, this inequality can be rewritten as:

$$\int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} (c(\theta) - c(\hat{\theta}_1^2))f(\theta)d\theta < \int_{\hat{\theta}_1^1}^{\hat{\theta}_2^2} (c(\theta) - c(\hat{\theta}_1^1))f(\theta)d\theta + \int_{\hat{\theta}_2^2}^{1} (c(\hat{\theta}_2^2)) - c(\hat{\theta}_1^1))f(\theta)d\theta.$$

Because all three integrals in the above expression are positive, the above inequality holds if:

$$\int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} (c(\theta) - c(\hat{\theta}_1^2))f(\theta)d\theta < \int_{\hat{\theta}_2^2}^{1} (c(\hat{\theta}_2^2)) - c(\hat{\theta}_1^1))f(\theta)d\theta.$$

The remaining part of the proof is also devoted to proving that this inequality holds. By partly evaluating the integrals, using equations (2) and (5) as well as the assumption that $F(\overline{\theta}_1^2) = \frac{1}{2}$, the above inequality can be rewritten as:

$$\int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} c(\theta)f(\theta)d\theta < B\left(\frac{1}{2} - F(\hat{\theta}_1^2)\right). \tag{9}$$

Note next that the following result always hold since $c(\theta)$ is an increasing function:

$$\int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} c(\theta)f(\theta)d\theta < c(\overline{\theta}_1^2)\int_{\hat{\theta}_1^2}^{\overline{\theta}_1^2} f(\theta)d\theta = c(\overline{\theta}_1^2)\left(\frac{1}{2} - F(\hat{\theta}_1^2)\right). \tag{10}$$

But now the proof is completed since the right hand side of equation (9) always is larger than the right hand side of equation (10). This conclusion follows as $\overline{\theta}_1^2 < 1$ and the budget is balanced.

Proof of Proposition 2

Only the first part of the proposition is proved (the other two parts are proved using almost identical arguments) and the same notation as in the proof of Proposition 1 will be used. Note first that because the budget is balanced independently of if there are one or two prices, it must be the case that:

$$B = c(\hat{\theta}_1^1) = \frac{1}{2}c(\hat{\theta}_1^2) + \frac{1}{2}c(\hat{\theta}_2^2) > c(\hat{\theta}_1^2/2 + \hat{\theta}_2^2/2),$$

where the last inequality follows from Jensen's inequality since $c''(\theta) > 0$. But this also means that $\hat{\theta}_1^1 > \frac{1}{2}(\hat{\theta}_1^2 + \hat{\theta}_2^2)$ since $c(\theta)$ is an increasing function. Because θ is uniformly distributed, it follows from equation (5) that $\hat{\theta}_2^2 = \hat{\theta}_1^2 + \frac{1}{2}$ and, therefore, that:

$$\hat{\theta}_1^1 > \frac{1}{2}(\hat{\theta}_1^2 + \hat{\theta}_2^2) = \hat{\theta}_1^2 + \frac{1}{4}. \tag{11}$$

Recall next that the critical cost type where the private firms stop serving patients when there is only one price is given by $\hat{\theta}_1^1$. Because the municipality serves the same proportion of patients in both sub-intervals when there are two prices and the types are uniformly distributed, the proportion of patients served by the municipality increases if:

$$\hat{\theta}_1^1 \geq 2\left(1 - \hat{\theta}_2^2\right) = 2\left(\frac{1}{2} - \hat{\theta}_1^2\right) = 1 - 2\hat{\theta}_1^2. \tag{12}$$

Using equation (11), it follows that equation (12) holds if:

$$\hat{\theta}_1^2 + \frac{1}{4} \geq 1 - 2\hat{\theta}_1^2. \iff \hat{\theta}_1^2 \geq \frac{1}{4}.$$

But this condition must hold by budget balance. To see this, suppose that $\hat{\theta}_1^2 \in [0, 1/4)$. Note next that for any strictly convex function $g(x)$ defined on the closed interval $[\underline{x}, \overline{x}] \subset \mathbb{R}_+$, and any constant $x' \in [0, \overline{x}/2]$, it holds that:

$$\int_{\underline{x}}^{\overline{x}} g(x)dx > (\overline{x} - \underline{x})g(x').$$

Let now g represent c, $[\underline{x}, \overline{x}] = [\underline{\theta}_j^2, \overline{\theta}_j^2]$ and $x' = \hat{\theta}_j^2$ for $j = 1, 2$. Given the assumption that θ is uniformly distributed in the interval $[0, 1]$, i.e., that $f(\theta) = 1$, the left hand side of the above expression can be interpreted as the cost in the interval $[\underline{\theta}_j^2, \overline{\theta}_j^2]$ and the right hand side as the reimbursement payed out by the regulator in the interval $[\underline{\theta}_j^2, \overline{\theta}_j^2]$. Because the inequality holds for $j = 1, 2$, this means that the budget cannot be balanced if $\hat{\theta}_1^2 \in [0, 1/4)$, which contradicts our assumptions. Thus, it must be the case that $\hat{\theta}_1^2 \geq \frac{1}{4}$ and the conclusion follows.

References

Agerholm, J., Bruce, D., de Leon, A. P., and Burström, B. (2015). Equity impact of a choice reform and change in reimbursement system in primary care in Stockholm county council. *BMC Health Services Research*, 15:1–10.

Allen, R. and Gertler, P. (1991). Regulation and the provision of quality to heterogenous consumers: The case of prospective pricing of medical services. *Journal of Regulatory Economics*, 3:361–375.

Amir, R. (2005). Complementarity in economics: An elementary survey. *Southern Economic Journal*, 71:636–660.

Anell, A., Dackehag, M., and Dietrichson, J. (2018). Does risk-adjusted payment influence primary care providers' decision on where to set up practices? *BMC Health Services Research*, 18:179.

Anell, A., Glenngård, A. H., Merkur, S., et al. (2012). Sweden: Health system review. *World Health Organization. Regional Office for Europe*.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5):941–973.

Barros, P. P. (2003). Cream-skimming, incentives for efficiency and payment system. *Journal of Health Economics*, 22(3):419–443.

Brown, J., Duggan, M., Kuziemko, I., and Woolston, W. (2014). How does risk selection respond to risk adjustment? New evidence from the medicare advantage program. *American Economic Review*, 104:3335–3364.

Chernew, M. E., Dafny, L. S., and Pany, M. J. (2020). A proposal to cap provider prices and price growth in the commercial health-care market. *The Hamilton Project. Washington, DC: Brookings Institution*.

Clevnert, U. and Johansson, L. (2007). Personal assistance in Sweden. *Journal of Aging & Social Policy*, 19:65–80.

Dackehag, M. and Ellegård, L. M. (2019). Competition, capitation, and coding: Do public primary care providers respond to increased competition? *CESifo Economic Studies*, 65:402–423.

Dietrichson, J., Ellegård, L. M., and Kjellsson, G. (2020). Patient choice, entry, and the quality of primary care: Evidence from Swedish reforms. *Health Economics*, 29:716–730.

Dranove, D. (1987). Rate-setting by diagnosis related groups and hospital specialization. *RAND Journal of Economics*, 18:417–427.

Eeckhout, J. and Kircheri, P. (2010). Sorting and decentralized price competition. *Econometrica*, 78:539–574.

Eggleston, K. (2000). Risk selection and optimal health insurance-provider payment systems. *Journal of Risk and Insurance*, 67:173.

Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics*, 24(1):211–223.

Ellis, R. P. (1998). Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics*, 17:537–555.

Ellis, R. P., Martins, B., and Rose, S. (2018). Risk adjustment for health plan payment. In McGuire, T. G. and van Kleef, R. C., editors, *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*, pages 55–104. Cambridge: Academic Press.

Geruso, M. and McGuire, T. G. (2016). Tradeoffs in the design of health plan payment systems: Fit, power and balance. *Journal of Health Economics*, 47:1–19.

Glazer, J. and McGuire, T. G. (2000). Optimal risk adjustment in markets with adverse selection: An application to managed care. *American Economic Review*, 90:1055–1071.

Hafsteinsdottir, E. J. G. and Siciliani, L. (2010). DRG prospective payment systems: refine or not refine? *Health Economics*, 19(10):1226–1239.

Hart, O., Shleifer, A., and Vishny, R. W. (1997). The proper scope of government: Theory and an application to prisons. *Quarterly Journal of Economics*, 112:1127–1161.

ISF 2021:11 (2021). Timschablonen inom assistansersättningen, inspektionen för socialförsäkringen. https://isf.se/publikationer/rapporter/2021/2021-12-15-timschablonen-inom-assistansersattningen.

Jegers, M., Kesteloot, K., De Graeve, D., and Gilles, W. (2002). A typology for provider payment systems in health care. *Health Policy*, 60:255–273.

Kastberg, G. and Siverbo, S. (2007). Activity-based financing of health care – Experiences from Sweden. *International Journal of Health Planning and Management*, 22:25–44.

Kifmann, M. and Lorenz, N. (2011). Optimal cost reimbursement of health insurers to reduce risk selection. *Health Economics*, 20:532–552.

Kifmann, M. and Siciliani, L. (2017). Average-cost pricing and dynamic selection incentives in the hospital sector. *Health Economics*, 26:1566–1582.

Laffont, J.-J. and Tirole, J. (1993). *A Theory of Incentives in Procurement and Regulation*. Cambridge: MIT Press.

Ma, C. A. (1994). Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy*, 3:93–112.

Milgrom, P. and Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics*, 19:179–208.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38:175–208.

Montanera, D., Mishra, A. N., and Raghu, T. S. (2021). Mitigating risk selection in healthcare entitlement programs: A beneficiary-level competitive bidding approach. *Information Systems Research*.

Newhouse, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature*, 34:1236–1263.

Oren, S. S. (2012). Nonlinear pricing. In Özer, O. and Phillips, R., editors, *Oxford Handbook of Pricing Management*, pages 492–521. Oxford: Oxford University Press.

Pissarides, C. (2000). *Equilibrium Unemployment Theory* (2nd edition). Cambridge: MIT Press.

Proposition 2021/22:1 (2021). Budgetpropositionen för 2022, Finansdepartementet. https://www.regeringen.se/rattsliga-dokument/proposition/2021/09/prop.-2021221/.

SFS 1993:387 (1993). Lag om stöd och service till vissa funktionshindrade, socialdepartementet. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-1993387-om-stod-och-service-till-vissa_sfs-1993-387.

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113:996–1025.

Shleifer, A. (1985). A theory of yardstick competition. *RAND Journal of Economics*, 16:319–327.

SOU 1991:46 (1991). Handikapp, välfärd, rättvisa, statens offentliga utredningar. lagen.nu/sou/1991:46.

Ven, V. D., Wynand, P. M. M., and Ellis, R. (2000). Risk adjustment in competitive health plan markets. Handbook of Health Economics, Elsevier.

Wilson, R. B. (1993). *Nonlinear Pricing*. Oxford: Oxford University Press.