Multiple Pricing for Personal Assistance Services

Andersson, Tommy; Ellegård, Lina Maria; Enache, Andreea; Erlanson, Albin ; Thami, Prakriti

2022

Document Version:
Other version

Link to publication

Citation for published version (APA):
Andersson, T., Ellegård, L. M., Enache, A., Erlanson, A., & Thami, P. (2022). *Multiple Pricing for Personal Assistance Services*. (Working Papers; No. 2021:14).

Total number of authors:
5

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
Revised: October 2023
Revised: May 2024

LUND UNIVERSITY

Multiple Pricing for Personal Assistance Services*

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

Abstract

Third-party payers often reimburse health care providers based on prospectively set prices. Although a key motivation of prospective payment is to contain costs, this paper shows that this aspect crucially depends on the design of the pricing scheme due to the well-known incentives of patient selection (or "dumping"). This paper provides a general theoretical framework where heterogeneous users are served by either private for-profit or public providers, each paid an hourly compensation by a third-party payer. The private, but not the public providers may select patients. It is demonstrated that this realistic feature of the model implies that total costs depends on the number of prices. The features of the model is illustrated using the Swedish system of personal assistance services as a motivating example. Numerical results show that marginal adjustments to the current uniform pricing scheme would lead to substantial savings.

Keywords: OR in health services, personal assistance, public and private providers, multiple pricing, welfare, dumping.

JEL Classification: C61, D47, D78, I11.

Declarations of interest: None.

---

*We are grateful to two anonymous referees for excellent comments and remarks. We are also 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. P2022–0087). 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. Address: Department of Economics, PO Box 7080, S-220 07 Lund, Sweden.

[‡]Department of Economics, Lund University; Faculty of Business, Kristianstad University. E-mail: lina_maria.ellegard@nek.lu.se.

[§]Department of Economics, Stockholm School of Economics. E-mail: andreea.enache@hhs.se.

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

[‖]Department of Economics, Lund University. E-mail: prakriti.thami@nek.lu.se.

1 Introduction

Health care providers are often reimbursed based on prospectively set prices. One example is the use of hospital prices based on Diagnosis Related Groups (DRG), which originated in the US Medicare and have since spread to numerous other settings in Europe and elsewhere (Busse et al., 2013). Similarly, providers of publicly funded long-term care (LTC) for elderly and disabled persons face prospectively set prices in the US (Medicaid), Australia, France, and Spain to name a few examples (World Health Organization, 2021).[1]

A key advantage of prospective payment, compared to cost-based reimbursement, is that it gives providers incentives to be cost-efficient (Shleifer, 1985). However, the very same feature implies that providers have incentives to select profitable patients (Newhouse, 1996), that is, to avoid (or "dump") patients who are likely to be more costly to serve than implied by the price (Ellis, 1998; Ma, 1994). Considerable research efforts have been put into investigating how pricing schemes can be refined to mitigate the incentives to dump patients (Eggleston, 2000; Hafsteinsdottir and Siciliani, 2010; Kifmann and Siciliani, 2017; McGuire and van Kleef, 2018). In this paper we explore an overlooked implication of dumping, namely that it may increase the total cost of providing the service. Specifically, when the dumped patients are served by a default provider, the third-party payer has to fund both the services provided to the dumped users and the (excess) payments to private providers. The novel insight of this paper is that a refined pricing scheme that mitigates dumping also serves the purpose of containing costs within a prospective payment system. Thus, although a key motivation of using prospective payment is to contain costs, the degree to which this goal is achieved depends on the incentives to dump patients.

We set up a general theoretical framework where heterogeneous patients, henceforth called users, are served by either private for-profit or public providers, each paid an hourly compensation by a third-party payer. A realistic feature of the model is that the private, but not the public providers, may select users. Dumped users are thus served by the public provider.

To motivate the model, the Swedish system for personal assistance services serves as a leading example throughout the paper (see Section 2), noting that the model and the results are not restricted to this specific application. For example, similar pricing schemes for publicly funded home care services are used in Spain, certain local authorities in France, and in the US state of Illinois (World Health Organization, 2021). The considered model is based on the empirical observation that private providers engage in dumping because the hourly compensation is fixed, but the cost of providing assistance services typically varies widely depending on who receives the service. That is, different users have different needs, but even if they have identical needs,

[1] For detailed reviews of nonlinear pricing schemes in other settings, see Oren (2012) or Wilson (1993).

the costs to assist them man differ, for example, because they live in more or less remote areas. To formalize this observation, it is assumed that users can be classified to be specific cost types, and that the regulator knows the cost type of each user.[2]

We use the model to address the following question: can the total costs be reduced by introducing a multiple pricing scheme where the hourly compensation remains identical for both public and private providers, but where the compensation is allowed to be distinct for different "types" of users? The investigated 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).[3] The solution to the optimization problem characterizes the pricing scheme that minimizes total costs for any given number of prices.

The results have notable policy applications, in that they can be used to investigate how a more flexible pricing scheme affects the monetary loss of the public default provider under different assumptions on cost functions and cost type distributions.[4] We provide numerical results demonstrating that marginal adjustments of the current Swedish system, such as increasing the number of prices from one to two, would lead to substantial reductions in the loss of the default providers. Another result of the model is that the subset of users who may effectively choose between private and public providers may include more high-cost types and fewer low-cost types under a refined pricing scheme. Although the investigated model does not formalize the value of allowing users to choose their provider, we note that a differentiated pricing schemes may give rise to a trade-off between the goals of cost containment and user choice. Consequently, the trade-off between incentives for efficiency and disincentives for patient selection, a core result in the literature on retrospective (cost-based) versus prospective payment systems (see Newhouse, 1996), also arises when considering prospective payment systems with more or less refined pricing schemes.

1.1 Relation to Previous Literature

This paper is not the first to demonstrate that there may be a trade-off between the goals of limiting selection and reducing costs in a prospective payment system. Geruso and McGuire (2016)

[2]The latter assumption is realistic in the personal assistance setting because 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. See Section 5 for more details.

[3]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, for example, Mirrlees (1971) does in his seminal continuous optimal taxation framework. See the discussions in Section 4.

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

note that the risk adjustment formulae used to reduce cream-skimming among health insurers often rely on information that is in fact tied to costs, such as previous diagnoses or procedures, to predict risk scores. Hence, more refined prices gives providers incentives to use more costly procedures. Notably, this mechanism, which stems from the information asymmetries between the payer and the provider, is inherently different to the one explored in the present study.

Brown et al. (2014) demonstrate that risk adjustment gives providers incentives to exploit variation in costs within each risk category, to select patients who are "cheap for their risk score." Their focus is on the overpayment to individual providers; they do not consider how risk adjustment affects total costs. In fact, the total costs in their motivating example are lower under risk adjustment than when there is only a single price, even though some providers are overpaid.

Savva et al. (2023) consider yardstick competition between hospitals. They show that further refinement of the DRG prices may lead to a bifurcation of the market, where some hospitals continue to dump high-cost patients and underinvest in cost-reducing technology, while other hospitals stop dumping and invest in such technology. Our model is complementary to that of Savva et al. (2023) and is more fit to inherently labor-intensive services such as personal assistance services and long-term home care, where technological innovation is a less important cost driver.

Most of the theoretical work on prospective payment is tailored to health care services, where the provider holds private information about patients' cost type and has considerable discretion over how much treatment to provide and what technology to use. Although 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 user 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, that is, a significant share of health expenditures.[5] Our model is tailored to a situation in which the service package is determined ex ante by a case manager, and the labor-intensive nature of the service leaves little scope for cost-reducing investments. In such a setting, the core economic problem facing the regulator is that total costs increase when private providers can dump unprofitable patients on a public provider, who cannot refuse to serve them. Since profits only serve as a signal of risk-selection, not as a signal of productivity or ability to innovate, in such a setting, the objective of the regulator in our model is to minimize public losses. In this sense, our model is most closely related to the branch of the industrial organization literature that investigates how to optimally regulate profit-maximizing firms (Laffont and Tirole, 1993).

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

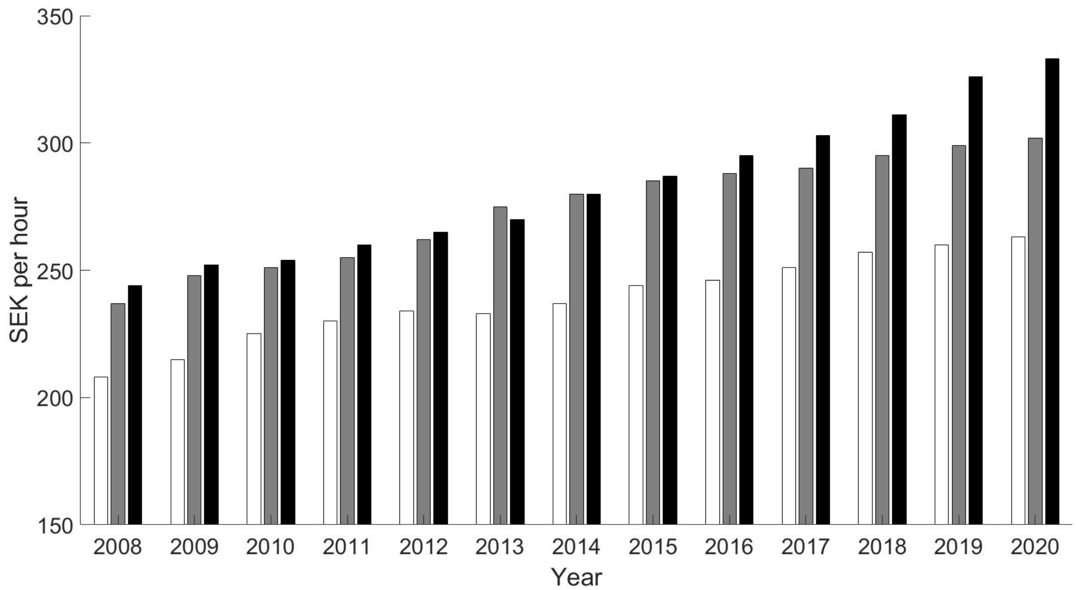

Figure 1: The average personnel costs (in SEK) for the private firms (white bars), the fixed hourly compensation from the central government (gray bars), and the average personnel costs for the municipalities (black bars) between year 2008 and 2020.

1.2 Outline of the Paper

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 introduces the theoretical model. The main theoretical findings are stated in Section 4. Section 5 contains some remarks related to a potential estimation of the model and a back-of-the-envelope calibration of the Swedish system. Section 6 concludes the paper. All proofs, as well as some technical remarks and numerical results, are delegated to Appendices A–C.

2 Motivating Example: Personal Assistance 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

[6]For a comprehensive description of the Swedish health care system in general, see Janlöv et al. (2023). There are several papers analyzing the impact of different price and choice reforms in Swedish healthcare, such as 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 analyzing multiple pricing schemes.

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

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 of governing personal services of this kind (SOU 1991:46, 1991, p. 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. Both types of providers are reimbursed by the same fixed hourly compensation from the central government. In 2020, this compensation was 304 SEK per hour (approximately 30 USD). 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 (Figure 1), another interesting pattern emerges: It is plausible that private firms indeed avoid the costliest users. The figure shows that the average personnel costs for the private firms (white bars) are clearly below the fixed hourly compensation from the central government (gray bars), while the average personnel costs for the municipalities (black bars) are clearly above the fixed hourly compensation from the central government (gray bars).[9] Thus, the data 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 users with a given cost type.[11] Although private providers 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

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

[9]The calculations of the estimated personnel costs are 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, p. 112).

improvements: The service package is standardized and consists of manual tasks with little room for productivity improvements. Wage rates are collectively bargained and thus do not normally 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 experience and 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 on the public providers.

3 Model and Preliminaries

For convenience, and without loss of generality, it is assumed that there is a continuum of users. The cost types θ are distributed on the interval $[0,1]$ according to the cumulative distribution function F with a strictly positive density function f. The cost of serving a user, that is, the user's cost type, is determined by several factors such as the need for high- or low-skilled care, the need for medical equipment and the remoteness of the user's place of residence (location determines travel costs for the provider). Any given user can be described by a vector of such cost-related factors $(\tau_1, \ldots, \tau_n) \in \mathbb{R}^n_+$. These factors determine the cost type $\theta \in [0,1]$ through some function $T : \mathbb{R}^n_+ \to [0,1]$. Note that the cost type does not by itself give sufficient information to infer the user's need for high-skilled care; for instance, the lower travel costs associated with serving urban users mean that a user in need of high-skilled care living in an urban area may have the same cost type as a user in need of low-skilled care who lives in a rural area.

Users can be served by a public provider or by a profit-maximizing private firm. Both types of providers can serve as many users as they desire (there are no capacity constraints). Private firms can decline to serve users. Any user not served by a private firm must be served by the public provider. A cost function $c : [0,1] \to \mathbb{R}_+$ specifies the costs for serving a user of cost type θ. The cost function is increasing and differentiable, that is, $c'(\theta) > 0$ for all $\theta \in [0,1]$. That the cost function is increasing follows by the definition of cost types because a higher cost type is associated with a higher cost. The assumption that private firms and the public provider have identical cost functions is rather mild given that there are no strong reasons to believe that (ceteris paribus) their cost structures are greatly different for the type of application used to motivate the theoretical framework (see Section 2). However, two remarks are in order here.

First, there is nothing in the aforementioned cost specification that separates users in terms of how many hours of assistance they need in a given time period. Suppose that there is a two-dimensional distribution function H describing the number of hours t of service that each user with cost type θ needs. Even if such function were to be included in the cost function, the theoretical results presented in the following would not qualitatively change, as demonstrated in Appendix A. For notational simplicity, we therefore disregard this issue in the main body of the paper.

Second, it is natural that the required skills of the personnel somehow is reflected in the cost type, because highly skilled personnel earn higher wages. As argued previously, there is not a perfect correlation between the cost type and required personnel skills. It is therefore not possible to identify a cutoff cost type θ^c such that *all* users with cost type $\theta \in [0, \theta^c)$ (with cost type $\theta \in [\theta^c, 1]$) need assistance from low-skilled (high-skilled) personnel. Nevertheless, given that the personnel costs constitute a large proportion of the total costs,[12] it is likely that there is a cost type θ^c that approximates such a cutoff (with some classification errors in the proximity to the cutoff), i.e., where most users with cost type $\theta \in [0, \theta^c)$ (with cost type $\theta \in [\theta^c, 1]$) can be served by low-skilled (high-skilled) personnel. We will return to this issue later.

A regulator determines the prices that the public provider and private providers receive for serving users. More specifically, the regulator partitions the support of cost types $[0, 1]$ into k subintervals. Let this partition be given by the vector $\overline{\theta} = (\overline{\theta}_0, \overline{\theta}_1, \ldots, \overline{\theta}_k)$ where $\overline{\theta}_{j-1} < \overline{\theta}_j$ for all $j \in \{1, \ldots, k\}$, 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 $j = 0, \ldots, k-1$, a subinterval j can also be written as $[\underline{\theta}_j, \overline{\theta}_j)$. Let $p = (p_1, \ldots, p_k)$ denote the price vector set by the regulator. The regulator must choose prices between the lowest and highest cost type for each interval. That is, $p_j \in [c(\underline{\theta}_j), c(\overline{\theta}_j)]$, and p_j represents the price for all cost types in subinterval j. Let $\mathcal{P} = \{p \in \mathbb{R}_+^k : c(\underline{\theta}_j) \leq p_j \leq c(\overline{\theta}_j)$ for each $j = 1, \ldots, k$ $\}$. Note also the difference in 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 users (a situation where the budget is smaller than the total costs can be incorporated in the model by adding a constant κ, which represents

[12] In the Swedish system for assistance service, for example, the hourly compensation was 304 SEK and the average personnel costs for a private supplier was 263 SEK (for year 2020, see Figure 1). Thus, the (average) hourly personnel costs constituted around 85 percent of the hourly reimbursement. According to a recent survey, the average profit margin among private providers was 3.9 percent in 2023. See https://assistanskoll.se/assistans_kostnadsslagen.php#rormar.

the "deficit," to equation (1)). Formally:

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

The total amount of money paid out to the private providers and the public provider cannot exceed the budget B. Consequently, the regulator is restricted by the following budget feasibility 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 public provider on users who are not served by the private firms. To formally define this objective, consider any given subinterval $[\underline{\theta}_j, \overline{\theta}_j)$ with an associated price $p_j \in [c(\underline{\theta}_j), c(\overline{\theta}_j)]$. Because the cost function is increasing, there is a cutoff $c^{-1}(p_j)$ for the subinterval j where the private firms serve all users on whom they can make a profit, that is, the users in the interval $[\underline{\theta}_j, c^{-1}(p_j)]$. Because private firms engage in dumping, the users who belong to the interval $(c^{-1}(p_j), \overline{\theta}_j)$ must be served by the public provider.[13] 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 users with cost types in the interval $[0, 0.58]$ because these are the only types where the price exceeds the cost of serving them.

Now suppose that there is a cutoff cost type θ^c such that most of the users with cost type $\theta \in [0, \theta^c)$ only need assistance from low-skilled personnel. If $\theta^c \geq 0.58$ in Figure 2, private providers will only employ low-skilled personnel. If $\theta^c < 0.58$, they will also employ some high-skilled personnel. As will become apparent later, the latter situation will normally be the case when there are multiple prices, but it need not be the case when there is a single price. Indeed, the simplest way to engage in dumping is to focus solely on easy-to-serve users that only require low-skilled personnel.

This structure also implies that the loss for the public provider in each subinterval equals the mass of users in the interval $(c^{-1}(p_j), \overline{\theta}_j)$ multiplied by the loss, that is, $c(\theta) - p_j$, of each user in that interval. In Figure 2, this loss is represented by the gray 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

[13] Recall that the number of hours needed per user is not taken into account in the main body of the paper (see Appendix A for details about extending the model). If this measure were included in the cost function, the cutoff value $c^{-1}(p_j)$ would change, but the general insight would still remain true, that is, there is some threshold cutoff value that the private providers do not operate beyond.

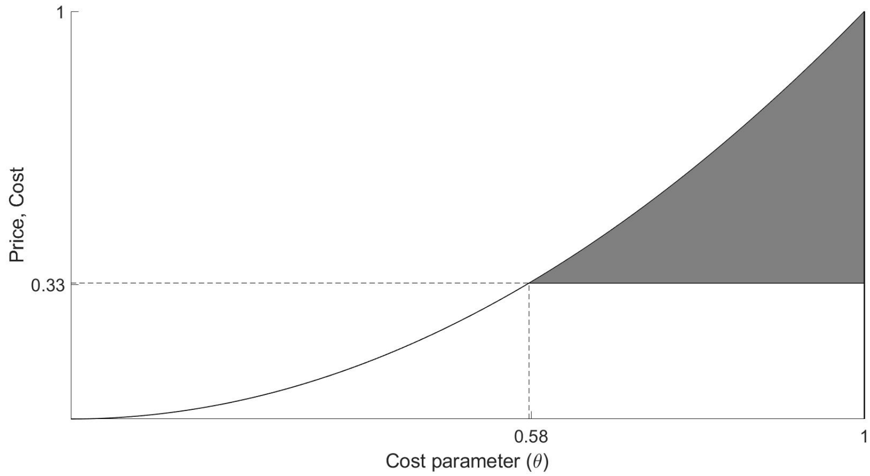

Figure 2: The cost function c is represented by the increasing function. For the single price $p = c^{-1}(0.58) = 0.33$, the private firms will only serve cost types in the interval $[0, 0.58]$. The loss for the public provider is represented by the gray area.

each subinterval, the total loss for the public provider is given by:

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

The regulator's 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 an objective is in line with the large body of 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).

4 Theoretical Results

As described in the previous section, the problem for the regulator is to set the prices to minimize the loss for the public provider. Given that the subintervals are fixed (and not a variable that the regulator optimizes over), the regulator thus needs to solve the following optimization problem:

$$\min_{p\in\mathcal{P}} \sum_{j=1}^{k} \int_{c^{-1}(p_j)}^{\overline{\theta}_j} [c(\theta) - p_j] f(\theta) d\theta$$

subject to:

$$p_k - \sum_{j=1}^{k-1} (p_{j+1} - p_j) F(\overline{\theta}_j) \leq B. \tag{4}$$

Because this is a convex problem (see Appendix B), it can be analyzed via its Lagrangian and the corresponding Karush–Kuhn–Tucker (KKT) 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 preceding Lagrangian are derived in the Appendix B. 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 users served by the public provider is constant in each subinterval 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 $[0, 1]$, that is, $F(\theta) = \theta$ 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 $p_j \in (\overline{\theta}_{j-1}^2, \overline{\theta}_j^2)$ because 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) users in each subinterval j. Furthermore, prices p_j are monotonically increasing in the sense that $p_{j+1} > p_j$ for all $j = 1, \ldots, k-1$. This can also be seen using equation (7) because this monotonicity condition holds if and only if:

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

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

As it transpires, 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 $j = 1\ldots,k$.

The interpretation of the next result is that private firms engage in dumping and, consequently, only serve the users on whom they will be able to make a profit. This finding is in line with what has previously been observed in the literature (see, e.g., Brown et al., 2014; Chernew et al., 2020; Eggleston, 2000; Newhouse, 1996). Note that because $c^{-1}(p_j) > \underline{\theta}_j$ by Theorem 1, there are cost types θ in the interior of the interval $[\underline{\theta}_j, c^{-1}(p_j)]$.

Corollary 1. For each subinterval j, private firms serve all users with θ-types in the 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 in 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 public provider is given by the sum of the gray areas in each figure.

By comparing the figures, it is evident that when cost types are divided into discrete subintervals with distinct prices, private firms serve the "lower cost types" within each subinterval 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). Related to the previous discussions on personnel skill levels among private providers, recall that when there is only one price (as in the current Swedish system), private providers may (but need not) only employ low-skilled personnel. This will typically not be the case when there are multiple prices, because private providers always have incentives to operate in all cost type intervals as can be seen in Figures 3 and 4 (see also Appendix C for numerical illustrations and, in particular, Table 3). Taken together, these insights demonstrate that there is a trade-off regarding the types of users who are able to effectively choose between different providers.[14] Specifically, the distribution of users 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 public provider (i.e., the sum of the gray areas in the figures) is decreasing in the number of prices. Thus, the multiple pricing policy will not only affect which users that have an effective choice, but also reduce losses for the public provider (without inflating the total budget set by the central government).

A general insight is that if the regulator is allowed to set infinitely many prices and, more precisely, one price for each cost type θ, the loss for the public provider will be zero because 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 user with cost type θ. But even if the assumption that there are enough funds available to cover the costs of serving the users 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 3). However, from a policy perspective, it is not feasible to set infinitely many prices because such a pricing scheme may be perceived as too complicated by the service providers. It is therefore important to investigate how the loss for the public provider, as well as the users' choice sets, are affected when implementing a multiple pricing scheme. It is, however, difficult to derive such results given the weak assumptions on, for example, the cost function, the type distributions, and the partition of subintervals. Hence, to obtain such results, additional assumptions must be imposed on the model (as in Example 1

[14]From a welfare perspective, it is only relevant to discuss this type of trade-off if the users actually value the possibility to choose between private and public providers (which is implicitly assumed in this paper). If so, the multiple pricing policy not only determines the set of users who can efficiently choose their service provider (as explained in Figures 2–4), but also how the welfare between the different sets of users is traded off (depending on the user types who can effectively choose their service providers). Note, however, that this type of trade-off does not reveal anything about aggregated welfare effects in the considered model. It may be the case that the aggregated welfare loss for the users who lose their option to choose is smaller than the aggregated welfare gain for the users who get the right to choose, or vice versa. To formally model the user value of choice, the framework introduced by Liran and Finkelstein (2011) may be useful.

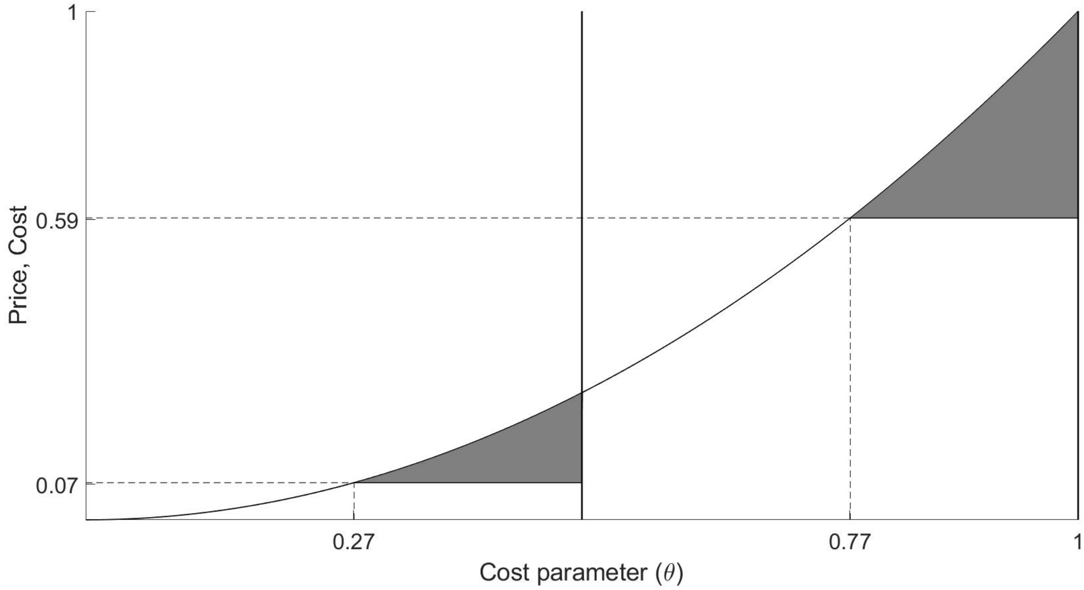

Figure 3: The cost function is represented by the increasing function. The two prices are given by $p^* = (0.07, 0.59)$, and the loss for the public provider is represented by the sum of the two gray areas.

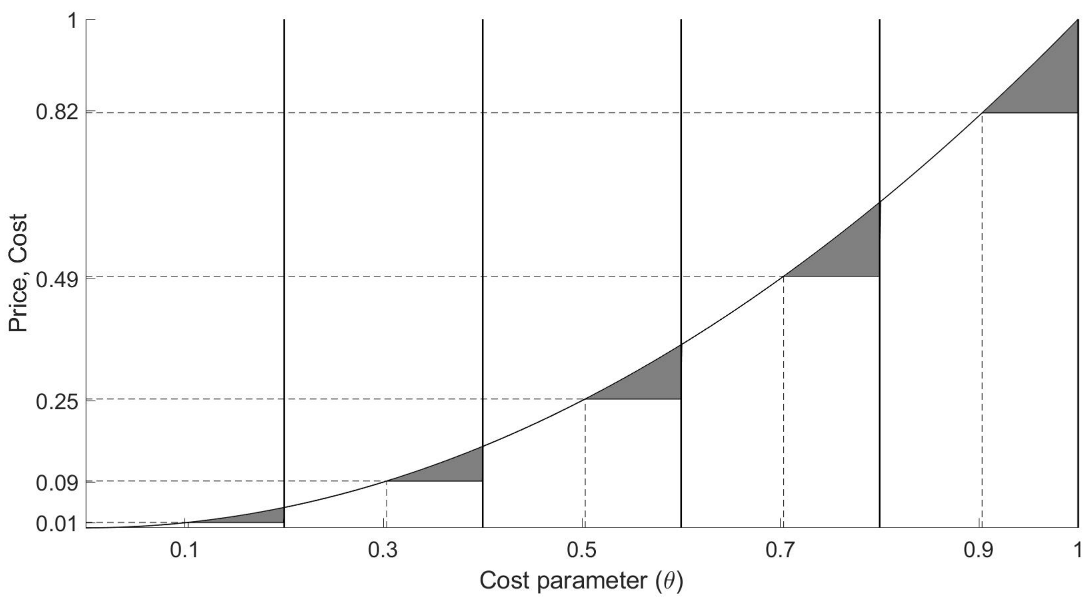

Figure 4: The cost function is represented by the increasing function. The five prices are given by $p^* = (0.01, 0.09, 0.25, 0.49, 0.82)$, and the loss for the public provider is represented by the sum of the five gray areas.

and Figures 2–4) or the model needs to be evaluated numerically. This approach is also adopted in the remaining part of the paper. The following two propositions focus on the case when the number of prices is increased from one to two.

The first proposition states that if the two cost type intervals contain equally as many users, then the loss of the public provider 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 the cost function is increasing and consider a cumulative density function F with a strictly positive density f where each subinterval $[\underline{\theta}_j, \overline{\theta}_j]$ contains an equal mass of cost types. Then, if the number of prices is increased from one to two, the loss of the public provider decreases.

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

Proposition 2. Suppose that each subinterval 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 users served by the public provider increases, decreases and is unchanged if $c''(\theta) > 0$, $c''(\theta) < 0$, and $c''(\theta) = 0$, respectively.

5 Estimation and Calibration

A question that naturally arises is how the presented model can be estimated and calibrated, and how it can be used to predict deficit reductions when introducing multiple prices. From Section 4 it is clear that to solve minimization problem (4), the regulator must have information about:

(a) the cost type distribution,

(b) the two-dimensional distribution function H, which describes the number of hours t of service needed by each user with cost type θ,[15]

(c) the cost function.

To obtain (a) and (b), user-level information on the assistance needs, place of residence, and so on is required. Such information is routinely gathered by the case handler when users apply for home care services (World Health Organization, 2021), and can in principle be shared with the

[15] Recall the discussion in Section 3. See also the theoretical analysis in Appendix A

regulator. Together with the cost factors τ, which are either known or possible to estimate using wage statistics and travel cost estimates, the regulator has all essential information needed to partition users into different cost types and to define the two-dimensional time distribution function H.[16] To determine the cost types, machine learning techniques may be appropriate because these methods can handle complex nonlinear relationships and, consequently, are particularly useful when mapping a diverse set of characteristics (such as age, medical condition, and living area) to cost types (see, e.g., Mullainathan and Spiess, 2017, for a list of potential methods). Once these types and functions have been obtained, the cost function (c) can be estimated using standard econometric techniques (for a thorough survey of cost function estimation in health care, see, e.g., Mihaylova et al., 2011).

Once the information specified in (a)–(c) is known to the regulator, the optimization problem can be solved numerically. Appendix C presents some illustrative numerical results, whereas the remaining part of this section focuses on a back-of-the-envelope calibration. Lacking access to the required registry data, we have to make additional assumptions, noting that the results of this exercise consequently should be interpreted with caution. More precisely, the time distribution function H is approximated by a single constant function for all cost types θ. The assumption is supported by Figure 5, which displays the average number of hours per user and week for 12 different age intervals. The number of hours is essentially constant across age groups, and there are only small differences between public (dashed line) and private providers (dotted line). Thus, the function H will be approximated using the average across all age intervals (solid line). Because this constant will play no role in the minimization problem, that is, it is a constant that can be factored our in the integrals, it will be ignored in the calibration exercise.

Because we do not have access to a good proxy for the cost type distribution, the calibrations are performed assuming that the cost types distribution is uniformly distributed on the interval $[0, 1]$ for a varying number of subintervals (Appendix C provides some illustrative numerical results under varying assumptions about the cost type distribution). Furthermore, we assume that the cost function can be represented by a piece-wise cubic hermite interpolating polynomial (PCHIP) function.[17] The parameters in this function are calibrated such that the model (when optimized) reports (i) a total budget of 23.53 billion SEK, (ii) a loss to public providers of 4.45

[16]In Sweden, applicants for personal assistance services provide detailed information including basic demographics and a motivation for needing assistance, along with medical documentation outlining specific needs. The Swedish Social Insurance Agency (SSIA) evaluates applications and determines assistance hours per week. This information, including patient needs and hours allocated, is shared with providers, whether public or private, who are also legally obliged to document service implementation. Thus, the necessary information for regulation is readily available in Swedish registers. Detailed information and application forms are available at the website of SSIA (`www.forsakringskassan.se/privatperson`).

[17]The function was identified using the inbuilt 'pchip' function in MATLAB. For more information see: `https://se.mathworks.com/help/matlab/ref/pchip.html`.

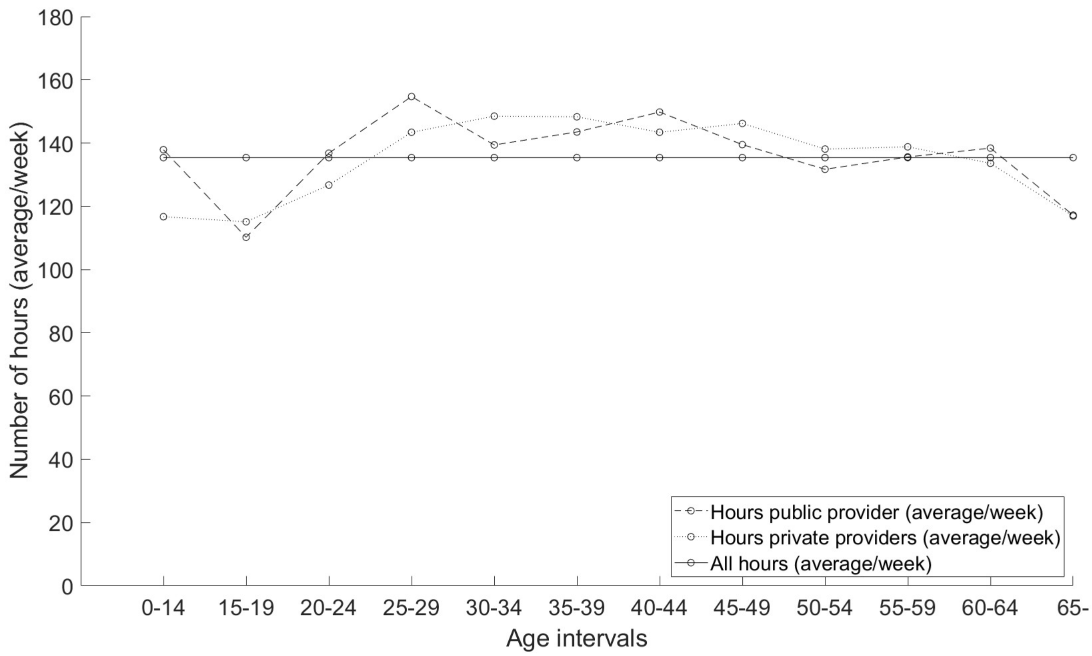

Figure 5: The average number of hours per week for the public providers (dashed line), private providers (dotted line), and all providers (solid line) per age interval in December 2022. Data from the Swedish Social Insurance Agency, publicly available at www.forsakringskassan.se/statistik-och-analys.

billion SEK, (ii) and an average profit margin among private providers of 3.9 percent. As previously reported, these numbers corresponds to the current Swedish situation.

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 to the public providers for two, five, ten, and fifty prices. If there are only two prices, the prices are given by $p_1 = 283$ SEK and $p_2 = 326$ SEK (recall that in 2020, the uniform price, that is, the fixed hourly compensation was 304 SEK) and the loss is estimated at 4.08 billion SEK. For five, ten, and fifty prices, the corresponding losses are 3.51, 1.89, and 0.42 billion SEK, respectively.

6 Discussion and Conclusion

This paper has modeled a situation where for-profit and public providers receive the same hourly compensation for serving users who differs in their cost to serve. Because only private providers can reject service requests, it is plausible that they dump unprofitable users to the public providers. As a consequence, private firms make profits at the expense of the budgets of the public providers (and ultimately the taxpayers). To mitigate such unintended consequences, in this paper we have proposed a multiple pricing scheme (see also Brown et al., 2014; Chernew et al., 2020), and demonstrated that it may substantially reduce the loss for the public providers without affecting the total budget provided by the central government. The conclusion holds even for the smallest possible modification of the current pricing scheme, that is, the addition of only one additional price.

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

An ideal study would have access to relevant micro data to estimate the cost function and the cost type distribution. Because we lack 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 users are treated as substitutes in the model. In reality, different users complement each other in different ways for both private and public providers. For example, two identical users may individually be regarded as "high cost types" simply because they both live in a remote geographical area. But if they both live in the same remote area, they may jointly be regarded as "two low cost types" by a service provider because the travel cost of serving them is reduced by half if the same provider serves both these users. 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 users first approaches the private firms and if they are rejected, they will be served by the public provider. In reality, public providers of personal assistance services also serve low cost types. Our approach is thus a simplification. So, how realistic is the assumption? To answer the question, note first that users are not automatically assigned to a provider, i.e., they have to contact a provider (public or private) themselves. Note next that a large majority of users are served by private providers.[18] For example, in 2023, 72 percent of all users chose to hire private providers. Given that public providers are not allowed to reject service requests, this implies that a large majority of all users prefer private over public provision. With that noted, we acknowledge that a more realistic approach would be to integrate the model with the search and matching literature (e.g., Eeckhout and Kircheri, 2010; Pissarides,

[18] See https://assistanskoll.se/assistans-statistik.php#three.

2000; Shimer, 2005) to also model how users and service providers are matched. Also, this theoretical extension of the model is left for future research.

The focus has been on reducing the monetary loss for the public care providers given the observation that private providers reject service requests from the costliest users. Because there is a value for users to chose their own care provider, it can be argued that there is a welfare loss, say ν, for each user who does not have this opportunity (i.e., the users whose cost type makes them unprofitable for the private provides). For each subinterval of cost types, this type of welfare loss can be calculated by multiplying ν by the mass of all such users in the interval (the total welfare loss is then essentially a sum of integrals defined on the relevant part of each subinterval). This type of welfare loss can be added to the loss function (3) and new optimality conditions, that look similar to equation (5), can be calculated given some additional assumptions that guarantee that the problem is convex. In such an approach, it is not clear that the results presented in this paper continue to hold with the same generality. A more detailed investigation of alternative welfare functions that also take, for example, "the value of choice," into account is left for future research.

Finally, given the focus on solving the deficit problem, one may argue that the problem can be solved by prohibiting private provision. Although this is a valid point, there are strong practical reasons to analyze the deficit problem. Private firms provide the majority of personal services in Sweden as well as internationally, suggesting that a proposed ban of private provision would be unfeasible. Our approach of making small adjustments to the existing framework is a much easier and pragmatic way to get around political deadlocks and legislative constrains. It is also in line with the highly influential minimalist approach to market design, pioneered by the Nobel Laureate Alvin E. Roth and Tayfun Sönmez (see Sönmez, 2023), in which the key objective is to fix broken markets, but not necessarily by abolishing them or by preventing specific actors from participating. With that said, the solution suggested in this paper does not claim to be normative, but rather presents an easy-to-implement solution to a prevailing problem in the current system. In fact, the Swedish Social Insurance Inspectorate (an independent supervisory agency for the Swedish social insurance system) recommended to the Swedish government that the model presented in this paper should be "developed, tested and evaluated to determine if this is a suitable solution for personal assistance pricing" (ISF 2021:11, 2021, p. 38).

A Appendix: Notes on the Cost Function

This appendix demonstrates that it is possible to incorporate a time dimension, represented by a function H (previously introduced in Section 3, but here described in more detail), in addition to the cost type θ in the cost function c, without altering the analysis. This will, in fact, only change the distribution F.

Suppose that each user with cost type θ demands different units of time t and that the cost is linear in time for the provider. This gives us a cost function c that depends on both θ and t as $c(\theta, t) = t\theta$. There is an upper bound T of the demand that any user may have for the service (simply because there are only 24 hours per day). Hence, time demands t belong to $[0, T]$ for some $T > 0$. Analogously, to the distribution of cost types θ with cumulative density function F, introduce now a two-dimensional probability distribution $H : [0,1] \times [0,T] \to [0,1]$ with a corresponding probability density function h capturing the distribution over cost types θ and time demands t. This gives the following (expected) cost function:

$$\begin{aligned} B &= \int_0^1 \int_0^T c(\theta, t) h(\theta, t) dt\, d\theta = \int_0^1 \int_0^T c(\theta) t h(\theta, t) dt\, d\theta \\ &= \int_0^1 c(\theta) \int_0^T t h(\theta, t) dt\, d\theta = \int_0^1 c(\theta) g(\theta) d\theta, \text{ where } g(\theta) = \int_0^T t h(\theta, t) dt. \end{aligned}$$

Let now $D = \int_0^1 g(\theta) d\theta$ and $\hat{f}(\theta) = \frac{g(\theta)}{D}$. Then:

$$\frac{B}{D} = \int_0^1 c(\theta) \frac{g(\theta)}{D} d\theta = \int_0^1 c(\theta) \hat{f}(\theta) d\theta,$$

where $\hat{f}$ is a probability density function corresponding to cumulative density function $\hat{F}$. Thus, the (expected) cost function can be written with a time dimension exactly as how it was modelled in the main body of the paper in equation (1), that is:

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

The only difference is that one needs to renormalize the budget constraint with a constant $\frac{1}{D}$.

Note, finally, that, if the two random variables are independently distributed, that is, if $h(\theta, t) = f(\theta) h_1(t)$ for some h_1, then:

$$g(\theta) = \int_0^T t h(\theta, t) dt = \int_0^T t f(\theta) h_1(t) dt = f(\theta) \int_0^T t h_1(t) dt = f(\theta) \mathbf{E}_t[t],$$

and it follows that:

$$B = \int_0^1 c(\theta) g(\theta) d\theta = \int_0^1 c(\theta) f(\theta) \mathbf{E}_t[t] d\theta = \mathbf{E}_t[t] \int_0^1 c(\theta) f(\theta) d\theta.$$

B 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 this 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:

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

Next, set the first-order conditions 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 $\lambda \in (0,1)$ which implies that $c^{-1}(p_j) > \underline{\theta}_j$ for each j. First, if $\lambda = 0$ then, by complementary slackness, the budget constraint is not binding. But this is impossible given the assumption 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 users will be served by private firms in each subinterval 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 $(1-\alpha)p'$ satisfy 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. Because 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 subinterval j are 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

To prove that the loss always decreases when the number of prices increases from one to two, it will be demonstrated that there exists feasible (but not necessarily optimal) prices such that the loss is always lower in the latter case compared to the former.

We start by computing the loss π_1 with only one price p. Budget feasibility and optimality imply that this price equals B, that is:

$$p(F(1) - F(0)) = B \iff B = p.$$

This gives a loss of:

$$\pi_1 = \pi(p) = \pi(B) = \int_{c^{-1}(B)}^{1} (c(\theta) - B) f(\theta) d\theta.$$

Consider now the case with two prices. Denote by π_2 the loss with the two optimally chosen prices p_1^* and p_2^*. By assumption, both subintervals have an equal mass of cost types, that is, $F(\overline{\theta}_1) = \frac{1}{2}$. Budget feasibility and optimality of prices then imply:

$$\frac{1}{2}(p_1^* + p_2^*) = B \iff p_1^* = 2B - p_2^*,$$

To complete the proof, we will derive an upper bound on π_2 and show that it is lower than π_1, thereby proving that $\pi_2 < \pi_1$. There are three cases to consider.

Case (i). Suppose that $F(c^{-1}(B)) < \frac{1}{2}$, implying that $c^{-1}(B) \in (0, \overline{\theta}_1)$. Then:

$$\begin{aligned}
\pi_1 &= \int_{c^{-1}(B)}^{\overline{\theta}_1} (c(\theta) - B) f(\theta) d\theta + \int_{\overline{\theta}_1}^{1} (c(\theta) - B) f(\theta) d\theta \\
&> \int_{c^{-1}(p_1)}^{\overline{\theta}_1} (c(\theta) - p_1) f(\theta) d\theta + \int_{c^{-1}(p_2)}^{1} (c(\theta) - p_2) f(\theta) d\theta \geq \pi_2,
\end{aligned}$$

where $p_1 = 2B - c(\overline{\theta}_1)$ and $p_2 = c(\overline{\theta}_1)$. Note that $c^{-1}(p_2) = \overline{\theta}_1$ and that $p_1 = 2B - c(\overline{\theta}_1) > c(0)$ since $B = \int_0^1 c(\theta) f(\theta) d\theta > \frac{1}{2}[c(0) + c(\overline{\theta}_1)]$. Hence, p_1 and p_2 are feasible prices for the optimization problem with two subintervals. The inequality between the two lines in the above follows from one observation: The increase in loss for the lower interval $[c^{-1}(p_1), \overline{\theta}_1]$ is always smaller than the savings in loss for the higher interval $[\overline{\theta}_1, 1]$. Thus, $\pi_2 < \pi_1$, as desired.

Case (ii). Suppose that $F(c^{-1}(B)) = \frac{1}{2}$. Here, almost identical arguments as in Case (i) can be used, simply by setting $p_1 = c(\overline{\theta}_1) - \varepsilon$ and $p_2 = c(\overline{\theta}_1) + \varepsilon$ for some "small" $\varepsilon > 0$. We omit repeating the preceding arguments once more.

Case (iii). Suppose that $F(c^{-1}(B)) > \frac{1}{2}$, implying that $c^{-1}(B) \in (\overline{\theta}_1, 1]$. We can then obtain an analogous upper bound on π_2 by again decreasing and increasing the prices in the two intervals, respectively.

$$\pi_1 = \int_{c^{-1}(B)}^{\overline{\theta}_1} (c(\theta) - B) f(\theta) d\theta + \int_{c^{-1}(B)}^{1} (c(\theta) - B) f(\theta) d\theta = \int_{c^{-1}(B)}^{1} (c(\theta) - B) f(\theta) d\theta$$
$$> \underbrace{\int_{c^{-1}(p_1)}^{\overline{\theta}_1} (c(\theta) - p_1) f(\theta) d\theta}_{=0} + \int_{c^{-1}(p_2)}^{1} (c(\theta) - p_2) f(\theta) d\theta \geq \pi_2$$

where $p_1 = c(\overline{\theta}_1)$ and $p_2 = 2B - c(\overline{\theta}_1)$. Note that $p_2 = 2B - c(\overline{\theta}_1) < c(1)$ since $B = \int_0^1 c(\theta) f(\theta) d\theta < \frac{1}{2}[c(\overline{\theta}_1) + c(1)]$. Hence, p_1 and p_2 are feasible prices for the optimization problem with two subintervals. Because $c^{-1}(B) > c^{-1}(p_1) = \overline{\theta}_1$ the first terms in the integrals are zero for both rows. Furthermore, $p_2 = 2B - c(\overline{\theta}_1) > B$ and, therefore, the strict inequality follows. Thus, $\pi_2 < \pi_1$, as desired.

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 whether 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 because $c''(\theta) > 0$. However, 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:

$$\hat{\theta}_1^1 > \frac{1}{2}(\hat{\theta}_1^2 + \hat{\theta}_2^2) = \hat{\theta}_1^2 + \frac{1}{4}. \tag{9}$$

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

$$\hat{\theta}_1^1 \geq 2\left(1 - \hat{\theta}_2^2\right) = 2\left(\frac{1}{2} - \hat{\theta}_1^2\right) = 1 - 2\hat{\theta}_1^2. \tag{10}$$

Using equation (9), it follows that equation (10) 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 $[0, 1]$, namely, 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 paid 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.

C Appendix: Numerical Results

This appendix presents some numerical results under varying assumptions about the cost function, the cost type distribution, and the number of subintervals. 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 parameters α and β in the cost function are calibrated such that the model (when optimized) reports a loss of 4.45 billion SEK (i.e., the deficit in Sweden as of 2020).[19] 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 subintervals for any given problem are partitioned such that they have equal probability mass. Furthermore, the function H is assumed to be constant (see Section 5 for further discussions). To analyze the proportion of low- and high-

[19]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.

skilled personnel employed by public and private providers, all calculations are also conducted for three different cutoff cost types and, more precisely, for $\theta^c \in \{0.4, 0.6, 0.8\}$. Throughout this section, the subintervals for any given problem are partitioned such that they have equal probability mass.

The results are presented in Tables 1–3. 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. For Table 3, each row also presents the results for given cutoff cost types. Each column in the tables presents the results from using the specified number k of subintervals in the objective function or, equivalently, the number of prices in the given optimization problem.

The losses for the public provider are displayed in Table 1 and are presented as a proportion of the total budget B set by the central government. For example, the loss for the quadratic cost function and the uniform distribution is given by 0.3849. This means that the monetary loss for the public provider 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 Table 1. First, even the introduction of one additional price reduces the loss incurred by roughly half across all specifications. Second, although 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 users served by the public provider. The results reveal the same pattern as predicted in Proposition 2 but for $k > 2$ prices, that is, the shape of the cost function determines whether or not the public provider will serve more users when the number of prices increases. Table 3 presents the proportion of patients served by high-skilled personnel employed by private providers. As expected from the previous discussions in Sections 3 and 4, the proportion of high-skilled personnel among the private providers decreases in θ^c, but increases in the number of prices k.

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

Table 2: Proportion of users served by the public provider

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

Table 3: Proportion of patients served by high-skilled personnel employed by private providers

Cost function	Distribution	θ^c	$k=1$	$k=2$	$k=5$	$k=10$	$k=50$
	$U(0,1)$	0.4	0.0444	0.2296	0.2841	0.2886	0.2949
	$t(1)$	0.6	0.0000	0.1296	0.1894	0.1924	0.1966
	$\Gamma(2,2)$	0.8	0.0000	0.0000	0.0947	0.0962	0.0983
	$U(0,1)$	0.4	0.0000	0.2364	0.2036	0.2425	0.2517
$c(\theta)=\sqrt{\theta}$	$t(1)$	0.6	0.0000	0.0483	0.1039	0.1455	0.1496
	$\Gamma(2,2)$	0.8	0.0000	0.0000	0.0369	0.0485	0.0690
	$U(0,1)$	0.4	0.2260	0.2425	0.3629	0.3706	0.3823
	$t(1)$	0.6	0.0107	0.2184	0.2627	0.2685	0.2773
	$\Gamma(2,2)$	0.8	0.0000	0.0360	0.0991	0.1390	0.1506
	$U(0,1)$	0.4	0.1000	0.2500	0.3000	0.3000	0.2999
	$t(1)$	0.6	0.0000	0.1500	0.2000	0.2000	0.2000
	$\Gamma(2,2)$	0.8	0.0000	0.0000	0.1000	0.1000	0.1000
	$U(0,1)$	0.4	0.0446	0.2579	0.2194	0.2516	0.2559
$c(\theta)=\theta$	$t(1)$	0.6	0.0000	0.0698	0.1145	0.1510	0.1521
	$\Gamma(2,2)$	0.8	0.0000	0.0000	0.0422	0.0503	0.0701
	$U(0,1)$	0.4	0.2631	0.2733	0.3827	0.3875	0.3941
	$t(1)$	0.6	0.0478	0.2338	0.2776	0.2811	0.2861
	$\Gamma(2,2)$	0.8	0.0000	0.0514	0.1090	0.1453	0.1553
	$U(0,1)$	0.4	0.1774	0.2704	0.3100	0.3050	0.3009
	$t(1)$	0.6	0.0000	0.1704	0.2066	0.2033	0.2006
	$\Gamma(2,2)$	0.8	0.0000	0.0000	0.1033	0.1017	0.1003
	$U(0,1)$	0.4	0.1288	0.2813	0.2311	0.2565	0.2568
$c(\theta)=\theta^2$	$t(1)$	0.6	0.0000	0.0932	0.1223	0.1539	0.1527
	$\Gamma(2,2)$	0.8	0.0000	0.0000	0.0461	0.0513	0.0703
	$U(0,1)$	0.4	0.3218	0.3143	0.4029	0.4014	0.3986
	$t(1)$	0.6	0.1066	0.2543	0.2927	0.2916	0.2892
	$\Gamma(2,2)$	0.8	0.0000	0.0719	0.1191	0.1505	0.1569

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.

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.

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

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.

Busse, R., Geissler, A., Aaviksoo, A., Cots, F., Häkkinen, U., Kobel, C., Mateus, C., Or, Z., O'Reilly, J., Serdén, L., Street, A., Tan, S. S., and Quentin, W. (2013). Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals? *BMJ*, 346:f3197.

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: 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.

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.

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

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.

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

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.

Janlöv, N., Blume, S., Glenngård, A., Hanspers, K., Anell, A., and Merkur, S. (2023). *Sweden: Health System Review 2023*. Number 4 in Health Systems in Transition. OECD Observatory on Health Systems and Policies.

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 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.

Liran, E. and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, 25:115–138.

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

McGuire, T. G. and van Kleef, R. C. (2018). *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*. Cambridge: Academic Press.

Mihaylova, B., Briggs, A., O'Hagan, A., and Thompson, S. G. (2011). Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, 20:897–916.

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.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31:87–106.

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/.

Savva, N., Debo, L., and Shumsky, R. A. (2023). Hospital reimbursement in the presence of cherry picking and upcoding. *Management Science*, 69:6417–7150.

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.

Sönmez, T. (2023). Minimalist market design: A framework for economists with policy aspirations. *Boston College Working Paper*.

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

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

World Health Organization (2021). *Pricing long-term care for older persons*. World Health Organization.