A Public Health Care Puzzle

Timo Seppälä
Centre for Health Economics at STAKES, RUESG and HECER

Discussion Paper No. 163
May 2007
ISSN 1795-0562
A Public Health Care Puzzle*

Abstract

It is a well documented fact that people do delay their doctor's visit when severe symptoms emerge. This causes extra costs for a publicly provided health care system. The other burden of a public health care system is over utilisation. A great deal of the over utilisation is by the patients that are not in need of a medical treatment but could easily survive with some self-medication. This constructs a puzzle: a part of the patients that should seek medical care stay inactive while the system is utilised mainly by those who are not in need of medical care. This paper contemplates interaction between patient behaviour and government's actions. A patient is equipped with hyperbolic preferences and he is either naive or sophisticated while time-consistent patient provides the benchmark. The government's possible actions are reduced to set a consultation fee, a deductible from it and a budget balancing tax rate. The redemption is accepted for patients whose diagnoses reveal a disease that needs to be treated with some non-self-medication methods. The main results establish that fairly small changes in the fee and the deductible can cause substantial changes in patients' behaviour. Naives are affected the most and actually their delays can become so long that counterintuitive change in the fee or deductible would actually decrease the total costs. Finally, an assumption about the partition of the types of individuals in the economy is crucial when choosing the consultation fee, the deductible and the tax rate, since assuming the absence of time-consistents or sophisticates is never as harmful as assuming the absence of naives in a case where they all exist.

JEL Classification: H31, H55, I10, D90.

Keywords: Hyperbolic preferences, self-control problems, patient behaviour, health care costs.

Timo Seppälä

Centre for Health Economics at STAKES
P.O. Box 22 (Lintulahdenkuja 4)
FI-00531 Helsinki
FINLAND

e-mail: timo.seppala@stakes.fi

* The author thanks the Research Unit of Economic Structures and Growth (RUESG) at the University of Helsinki and the Yrjö Jahnsson Foundation and the Finnish Cultural Foundation for financial support. I'm grateful to my advisors Erkki Koskela and Rune Stenbacka for helpful comments. I have also benefited from the insights of participants at HECER labour and public economics seminar.
1 Introduction

While Kenkel (2000) points out the importance of preventive actions as a part of medical expenditure formation for health care systems, it is not only about whether people take preventive actions to keep themselves in shape to avoid costly treatments but also do they seek medical help in time when they notice some symptoms. The following questions arise. Why do people visit a doctor when he or she has only minor symptoms, and on the other hand, why do people not consult a doctor when some symptoms, sometimes even very clear ones, have been lasting very long, and it is quite obvious that one is in need of a diagnosis and of some medical treatment? These phenomena are causing large expenses for economies supporting a public health care system.\(^1\) Firstly, it is obvious that utilising a health care system when it is not necessary causes sunk costs. Secondly, since usually costs of a medical treatment are not linear nor constant in time if a disease is serious, delaying a visit to a doctor causes the costs to increase fast and those can be very high once procrastinator finally gets the treatment.\(^2\) In the public debate, it has been claimed that since people do not pay the market price from visiting a doctor, these services are overutilised and costs accrue unnecessary high for the economy. The second problem, i.e. not to visit a doctor when one should, has got much less public attention while it is empirically well documented fact that people do delay their doctor visits and in this manner cause higher health care costs than what would be attained if people visited a doctor when it is necessary.\(^3\) The public health care puzzle emerges, people who should exploit the public health care system stay inactive whereas it is exploited by those who could easily survive without any medical consultation. In other words, medical costs are increased from the both ends.

What then makes people to delay doctor’s visits when a public health care system covers the expenditures? A naive answer would lean on easy verbalism: “Time is money”, i.e. observed procrastinators just happen to have a higher marginal cost for time compared to those who visit a doctor immediately after noticing symptoms. While a neat and clean answer, the problem is that it is not in line with empirical findings. Caplan (1995) finds that women with increasing symptoms are more likely to delay than women whose symptoms either decreased or remained the same. Meechan et al.

---

1In here, a public health care system refers to a system in which citizens do not have to pay directly anything or only a very small nominal payment to get a medical consultation and treatment when necessary. Hence all costs from the medical care are beared by a local or national government and funded indirectly via taxation.

2About the formation of treatment costs see eg. Butler et al. (1995).

(2002) find, in addition, that women who discovered their breast symptom by chance or through breast self-examination had a shorter delay than women experiencing breast pain. There was also a trend for women who had a family member with breast cancer to have a longer delay time before seeing their general practitioner. Mohamed et al. (2005) show in their empirical study that delays are associated with demographic variables such as lower education, socioeconomic status and older age, as well as, with psychological factors including e.g. severe anxiety and psychosis. None of these interesting findings point out that lackness of time would be the reason to delay the doctor’s visit once symptoms emerge but rather show clearly how anxiety, anticipatory feelings, fears, and other things cause procrastination. However, it is not an easy task to build up a theoretical model that would explain why the delayed visits occur. Especially, behaviour that is present in data is hard to explain with conventional models since those models are incapable of establishing time-inconsistent behaviour. Fortunately, recent literature shows us that where the conventional models seem to be inadequate to explain the phenomenon, steering the models into the direction of behavioral economics facilitates the task. For starters, Caplin and Leahy (2001) develop a model that explains the patient’s time inconsistent behaviour with anticipatory feelings. Köszegi (2003) then concentrates solely on patient’s decision-making and applies the Caplin and Leahy (2001) model to explain delayed doctor’s visits, and shows that anxiety can lead the patient to avoid doctor’s visits or other easily available information about his wealth. This means that the belief about future health is taken into a utility function and the patient is assumed to be information averse. That is, he procrastinates doctor’s visit the longer the stronger is his belief that information from doctor’s visit is bad, i.e. utility decreasing. On the other hand, there is also other explanation that does not utilise anticipatory feelings but endogeneous determination of time preference. Namely, Becker and Mulligan (1997) (BM) show that if time discounting is affected by e.g. mortality, uncertainty, and addictions delayed visits to a doctor will occur without including anticipatory feelings in to a model.

Hence, in some cases where an agent’s actions might seem to be caused by anticipatory feelings or suchlike, the correct answer for observed behaviour is rather endogeneous time-preference. In their

---

4 Note here also that ‘time is money’ explanation is not good when the focus is on public health care system usage. Those who have higher cost to wait usually utilise private sector. In this paper we do not concentrate on private sector at all but we are solely interested in studying behaviour of those who have already chosen to use the public health care.

5 For empirical evidence about preproperation and procrastination in completing pleasant and onerous tasks see e.g. Ainslie (1992) and further references therein.

6 We use he as a personal pronoun for a generic individual. The choice was made by tossing a fair coin.
model discounting occurs due to an imperfect ability to imagine the future. They consider a case where an individual has a possibility to improve his capacity to appreciate the future. This action is costly, and hence it is subject to individual optimisation. A discount factor, \( \delta \), is a function of \( I \), where \( I \) is an investment in improving one’s capacity to appreciate the future. Endogenous discount factor \( \delta (I) \) satisfies the following properties: \( \delta (I) > 0 \), \( \delta' (I) \geq 0 \), and \( \delta'' (I) \leq 0 \) for all \( I \geq 0 \). In the light of current paper, an interesting implication of Becker and Mulligan model is that differences in health could cause differences in time preference. This could happen since greater health reduces mortality and raises future utility levels, and so those with greater health have also a greater incentive to invest in improving one’s capability of value the future than those with poor condition of health. Thus, discounting happens to be milder for people in good shape than what it is for ill persons. Kenkel (2000) notes that while in theoretical studies insured people are less likely to take preventive actions due to \textit{ex ante} moral hazard there is little empirical evidence on that from systems which support private insurance.\textsuperscript{7} However, studying public health care systems reveals that theoretical predictions are supported by data and \textit{ex ante} moral hazard can be shown to exists. Picone et al. (2004) study the effects of risk, time preference, and expected longevity on demand for medical tests. They build up a theoretical model and test it with the data from Health and Retirement Study. At the focal point of the study are women and how they obey national recommendations to treat themselves against breast cancer. Interesting results establish that 1) nearly a third of the woman in the data set are quite myopic and they are less likely to demand for medical tests, 2) those with longer expected life are more likely to engage in preventive behaviour, and 3) uninsured women are less likely to seek for medical tests.

While plenty of studies in the field are about valubleness of preventive screening programs, in this paper we consider a slightly different setup where the main focus is on interaction between patient behaviour and cost formation of the public health care system where the costs are effected by the chosen behaviour. Ideas for modeling are captured from the discussed studies. We are keen to contemplate: 1) the patient behaviour in a case where one first notices (some) symptoms and then decides whether to seek medical help immediately or postpone the task to later periods, 2) the effects of the delays on costs for the public health care system, and finally, 3) the government’s

\textsuperscript{7}It is called \textit{ex ante} moral hazard to emphasise the effect of an insurance on action people takes \textit{before} his state of health is known. By this terminology \textit{ex post} moral hazard is then a term for the action people takes after getting to know his state of health (see Kenkel (2000)).
optimal actions to balance to health care budget and avoid unnecessary high medical costs.

In our simple setup there are three patient types: time-consistent, hyperbolic naif, and hyperbolic sophisticated. We lean on empirical findings about myopic patients in Picone et al. (2004) and hence we assume that patients can have hyperbolic preferences. The intertemporal utility function we use is the same that is commonly used for hyperbolic preferences, and hence gets the following form: \( U_t = u_t + \beta \sum_{\tau=1}^\infty \delta^\tau u_{t+\tau} \), where \( \beta \) is short-run discount factor and \( \beta < 1 \) for hyperbolic patient and \( \beta = 1 \) for the time-consistent, while \( \delta^\tau \) is a conventional exponential long-run discount factor.

We use basic definitions for naivety and sophistication, hence, naif is the one who does not realise his future self-control problems (i.e. does not know the future \( \beta \) to be \(< 1 \)) while sophisticated is the one who knows perfectly his future self-control problems.\(^8\) We revise the ideas from BM and consider that while the patient cannot invest in improving one’s capacity to appreciate the long distance future (i.e. \( \delta \) is the same for all) the short-run discount factor is a function of symptoms \( s \).

We let \( \beta (s^*) < \beta (s_*) \), where \( s \) is the level of the symptoms and \( s_* < s^* \). In the language of BM this means \( I (s^*) < I (s_*) \).\(^9\) We take the investments given, automatic and non-monetary.\(^10\) Another crucial point in our modeling is that we assume patients to have a subjective positive probability for symptoms’ disappearance (or persistence). When time elapses it approaches its real probability of 0 (1).\(^11\) This means that the patient first considers it possible that symptoms vanish away along passage of time but if those stay persistent then this subjective probability will be updated towards more persisting symptoms. Naturally this increases the patient’s willingness to seek for a treatment for the disease, and consequently he cannot delay visiting a doctor indefinitely.

Delayed visits to a doctor are under our special interests, since those can be very costly for public health care system and costs of the delayed visits are rarely discussed in the recent literature.\(^12\) We model the costs with a function which shape follows findings in Butler et al. (1995). The function is simplified version of that but it captures the basic and necessary properties for the cost formation.

\(^8\) See eg. O’Donoghue and Rabin (1999).
\(^9\) This idea is supported by BM since according to their study those people who consider their future to be pleasant are ready to invest in improving one’s capacity to appreciate future more than those with onerous future prognoses.
\(^10\) Notice that severity of symptoms has only effect on a hyperbolic patient’s preferences and not at all on a time-consistent patient’s preferences. Implicitely this means that we assume symptoms to be effective only for those who "suffer" from reversing preferences, that is, those who have hyperbolic preferences.
\(^11\) This is quite natural assumption, since people have a tendency to give probability for symptoms to vanish away. However, for persisting symptoms it happens that probability to vanishing decreases while persistence gets heavier probability.
\(^12\) Most of the studies discuss about screening programs and their efficiency and cost effectiveness but not the cost effects of delayed visits to the public health care system.
For the first two stages of the disease the costs are the same but after that they start to increase dramatically.\textsuperscript{13} The government’s focus is on cost formation since we assume that the health care system is publicly financed. Under contemplation is then the government’s possibilities to set a consultation fee and reimbursement from that as well as the optimal tax rate.

The main results establish that fairly small increases (decreases) in the fee (the deductible) can cause substantial changes in patients’ behaviour. Naïfs are affected the most and actually their delays can become so long that a decrease (increase) instead of an increase (decrease) in the fee (the deductible) would decrease the total costs from the system and hence the tax rate resulting in higher expected intertemporal utility from social planner’s viewpoint. Finally, an assumption about the partition of the types of individuals in the economy is crucial when choosing the consultation fee, the deductible and the tax rate, since assuming the absence of time-consistents or sophisticates is never as harmful as assuming the absence of naïfs in a case where they all exist. The policy relevance of the results is then immediate. Using a conventional assumption of the sole existence of time-consistent individuals can lead not only on substantial imbalances of the health care budget but also on completely falsely selected corrective actions. In addition, while the direct costs for the public health care system can increase due to interaction of the chosen policy and the existence of time-inconsitents, the indirect costs for the whole social security system can increase even more as the unnecessary long delays can cause losses in patient’s productivity and through that in several other dimensions.

The rest of the paper is organised as follows. In Section 2 we construct the model. Section 3 analyses patient behaviour whereas government’s actions are analysed in Section 4. Finally, Section 5 concludes by discussing the implications of the results.

\section{The model}

We consider a publicly financed health care system which costs are initially fully covered through taxation. To revise financing, the government has a possibility to establish a fee, \( b \geq 0 \), for a consultation visit to a doctor, while treatments are still free of charge. This is to say that after establishing the fee all medical treatments and the part of the system costs that is not covered by the

\textsuperscript{13}In clinical studies it has been found that tumor size, for example, doubles in from one week to 300 days. On average the doubling time is 150 days.
fee are still financed through taxatation. Finally, we suppose that if a medical examination reveals a serious disease the government will repay \((1 - \gamma)\) fraction of the fee to a patient, \(\gamma \in [0, 1]\)\(^{14}\). We assume then that cost minimising behaviour of a patient who observes symptoms in some period \(t\) for the first time would be waiting until period \(t + k\) before seeking medical help.\(^ {15}\) We, hence, assume that diseases in the model are such that even for serious diseases \(k\) periods waiting time does not increase the treatment costs.\(^ {16}\) On the other hand, we assume that no asymptomatic diseases exist in the economy. This is to say that those who have serious disease have also the symptoms always.

The timing of events is the following. In the beginning of each period the patient notices the level of symptoms. After that he decides whether to visit a doctor in the current period or postpone it for a subsequent self. The patient’s every period task is then to maximise his intertemporal utility given his perceptions about his future behaviour.

For notational ease we use \(\mathbb{N}^n_m\) to denote a sequence of natural numbers from \(m\) to \(n\).

**Symptoms** We denote symptoms in period \(t\) by \(s_t\), \(s_t \geq 0\), and \(s_t = 0\) refers to no-symptoms. We assume that the incidence of having symptoms is \(\eta \in (0, 1)\). By \(t^*\) we refer to the period when an individual noticed the symptoms for the first time. We assume that in the first \(k\) periods after the symptoms emerged the probability of having a serious disease, \(\rho_{tr}\) (\(tr\) as true) is \(\rho \in (0, 1)\) and if the symptoms still occur in \((k + 1)\)th period from the first observation the disease is severe with certainty.\(^ {17}\) Formally,

\[
\rho_{tr}(n|t^*) = \begin{cases} 
\rho, & \text{if } n \leq k \\
1, & \text{if } n > k 
\end{cases}
\]

\(^{14}\)In here, a disease is serious if it does not heal without a medical treatment or prescribed medicines. The opposite for the serious disease is called here as harmless. The fraction \(\gamma\) is then deductible fraction for the patient that must be paid for visiting a doctor no matter what.

\(^{15}\)As it will be seen later in this paper, it is clear that if symptoms are severe enough the assumption fails to hold. However, we see this assumption plausible since when considering e.g. symptoms that are typical for the flu it is rather optimal to wait some time after the emerged symptoms than immediately visit a doctor, and if the symptoms stay at a constant level or get worse then seek for a medical care. In other words, since some symptoms lead rather to a set of diseases than for a certain disease more time can clarify the symptoms and hence severity of a disease.

\(^{16}\)For example, a breast cancer has a cost function that shows to be of this form. I.e the treatment costs do not increase between Stage 0 and Stage 1 at detection. Naturally, it is implausible to consider any waiting time to be optimal for cancers but for some milder diseases, the kind of where medical treatment is not always necessary, \(k\) periods waiting time can be optimal if the patient really visits a doctor when the symptoms are still present after \(k\) periods.

\(^{17}\)More natural assumptions would be \(s_{t+1} - s_t > 0\) with probability of \(\rho_{sev} = \rho_{sev}(s)\) for which \(\rho_{sev}(s) > 0\), \(\rho_{sev}(s) < 0\). So, that probability of having more extreme symptoms would be function of this period symptoms, and with the property that the harder are the symptoms the more probable it is that they get even worse when time elapses.
where \( n \) is the time distance from the period \( t^* \). \(^{18}\) Naturally, the probability of getting cured without seeing a doctor is given by the complement, i.e. by \( 1 - \rho_{tr}(n|t^*) \).

We assume that there is a cure for all existing serious diseases, and hence if the patient visits a doctor in period \( t \) he will get rid of the symptoms from the next period \( t + 1 \) onwards. \(^{19}\)

**Patients** The patient \( i \) is a hyperbolic discounter and infinitely living, and his type \( \varphi_i \) belongs in the type set \( \Phi = \{N,S,T\} \), where \( N \) is naif, \( S \) is sophisticated, and \( T \) is time-consistent. \(^{20}\)

Adopting the fashion of discounting and the definitions for naivety and sophistication from the recent literature on hyperbolic preferences causes all the types to be identical in all manners except in their discounting and perceptions about future behaviour. We denote the discount factors by \( \beta \in (0,1] \), \( \delta \in (0,1] \), and perception about future behaviour by \( \hat{\beta} \geq \beta \). We use a multiselves approach, and thus we model the patient as a sequence of autonomous temporal selves. These selves are indexed by the respective periods in which they control the choice variable, and we denote a self in period \( t \) by \( \text{self}(t) \). The patient’s intertemporal utility for \( \text{self}(t) \) is of the following form:

\[
U_t = u_t + \beta \sum_{\tau=1}^{\infty} \delta^\tau u_{t+\tau},
\]

where \( u_t = u(C_t, H_t) = C_t H_t \) is an instantaneous utility loosely following Picone et al. (1998), \( \beta \) is the time-inconsistent discount factor, and \( \delta \) the conventional time-consistent exponential discount factor. \(^{21}\)

In the instantaneous utility \( C_t \) is current consumption and \( H_t \) is the state of the health. We assume that

\[
H_t = H_o - s_t,
\]

where \( H_o \) is full health, and \( u_t \) must be positive for the patient to be alive. For simplicity we assume that \( s_t \in \{0, s\} \forall t \), and if \( s_t = s \) we denote \( H_t = H_s \).

To keep things as simple as possible we assume that there is no possibility to save and hence if

\(^{18}\)It should be obvious that \( \rho_{tr}(n|t) \) is defined only for \( n \geq t^* \).

\(^{19}\)Without loss of generality, we assume that no matter whether the disease is serious or not the patient gets rid of the symptoms in the next period after visiting the doctor.

\(^{20}\)The agent is said to be hyperbolic discounter when he discounts a time interval in near future heavier than in the distant future. Eg. if one prefers 1 apple today over 2 apples tomorrow but at the same time he prefers 2 apples after 101 days over 1 apple after 100 days, these choices imply that the agent might have hyperbolic preferences.

\(^{21}\)The form of the intertemporal utility function becomes from Strotz (1955-56), Phelps and Pollak (1968), and finally from Laibson (1994).
the gross-of-tax income is denoted by \( w^g \) and the tax rate by \( \tau^g \). It directly follows

\[
C_t = w^g (1 - \tau^g) - M_t \\
= w - M_t,
\]

where \( w \) is the periodical net-of-tax salary and \( M_t \) is medical expenses in period \( t \) that is given by

\[
M_t = \begin{cases} 
0, & \text{if } s = 0 \\
 b, & \text{if self}(t) \text{ visits a doctor in } t \text{ but disease is not serious} \\
b(1 - \gamma), & \text{if self}(t) \text{ visits a doctor in } t \text{ and disease is serious.}
\end{cases}
\]

and hence \( C_t \) is consumption in period \( t \) on other thing from medical commodities.

We make a novel assumption by assuming that symptoms have an effect on the magnitude of \( \beta \), hence \( \beta \) is considered to be a function of \( s \). We assume that \( \beta'(s) \leq 0 \) and \( \beta''(s) > 0 \) if \( \beta'(s) < 0 \). Notably, \( \beta \) is also a function of \( \varphi \) and \( \beta(s, \varphi) = \beta(s) \) for \( \varphi \in \{N, S\} \) and \( \beta(s, \varphi) = 1 \) for \( \varphi = T \).

Every type is now fully described by a 3-tuple \((\beta(s), \beta(s), \delta)\), i.e. \((\beta(s), 1, \delta)\) is \( N \), \((\beta(s), \beta(s), \delta)\) is \( S \), and \((1, 1, \delta)\) is \( T \). In words, the naif patient is completely unaware and the sophisticated patient fully aware of the reversals in their future preferences while the time-consistent patient does not "suffer" from the reversing preferences at all.

About the patient’s medical knowledge we assume that he does not have proper skills to diagnose true severity of a disease after observing the symptoms but instead has a subjective probability on persistence or severity of a disease. The subjective probability is conditional on how long the symptoms have lasted. This probability is denoted by \( \rho_{sub}(n|t^*) \) (sub as subjective) that is defined by \( \rho_{sub}(n|t^*) \equiv \rho_n \), where \( \rho \in [0, 1] \), \( n \in \mathbb{N}_{1-t^*} \), and when \( t \) and \( t^* \) are clear from the context we simplify by denoting \( \rho_{sub}(n|t^*) \equiv \rho_n \). The subjective probability \( \rho_{sub}(n|t^*) \) tells us that if symptoms still appear in period \( t^* + n \) the subjective probability that symptoms occur also in the subsequent period \( t^* + n + 1 \) is \( \rho_{sub}(n|t^*)^\frac{1}{n} \). In the same manner, we define the complement, i.e. the patient considers that the symptoms will vanish away between periods \( t^* + n \) and \( t^* + n + 1 \) with probability of \( 1 - \rho_{sub}(n|t^*) \).

When contemplating a patient’s behaviour we adopt and use the concepts given in O'Donoghue

\footnote{Note that \( \rho_{sub}(n|t^*) \) is increasing in \( n \), implying that even though the patient is incapable of knowing that \( k \) periods persisted symptoms actually imply a serious disease, he will learn it when time elapses.}
and Rabin (1999) (OR 1999). The patient’s behaviour is then fully described with a strategy, \( \sigma_\varphi \equiv \{\sigma^\varphi_k\}_{k=1}^{\infty} \), that tells us when a patient of type \( \varphi \) is willing to visit a doctor, \( \sigma^\varphi_k = v \), and when he is willing to delay the visit, \( \sigma^\varphi_k = d \). We get a solution concept that is called a perception-perfect strategy, which is stated in the following definition, where \( U^\varphi_t (n) \) is an expected intertemporal utility for self\((t)\) of type \( \varphi \) in a case where the patient plans to visit a doctor in period \( n \).

**Definition 1** *(Revised from OR 1999)* A perception-perfect strategy for

i) \( T \) is a strategy \( \sigma^T \equiv (\sigma^T_1, \sigma^T_2, \sigma^T_3, \ldots) \) that satisfies for all \( t \) \( \sigma^T_t = v \) iff \( U^T_t (t) \geq U^T_t (\kappa) \) for all \( \kappa > t \);

ii) \( N \) is a strategy \( \sigma^N \equiv (\sigma^N_1, \sigma^N_2, \sigma^N_3, \ldots) \) that satisfies for all \( t \) \( \sigma^N_t = v \) iff \( U^N_t (t) \geq U^N_t (\kappa) \) for all \( \kappa > t \);

iii) \( S \) is a strategy \( \sigma^S \equiv (\sigma^S_1, \sigma^S_2, \sigma^S_3, \ldots) \) that satisfies for all \( t \) \( \sigma^S_t = v \) iff \( U^S_t (t) \geq U^S_t (\kappa') \) where \( \kappa' \equiv \min_{\kappa > t} \{ \kappa | \sigma^S_\kappa = v \} \).

About \( T \)’s and \( N \)’s strategies Def.(1) states that \( \sigma^\varphi_k = v \) if and only if there is no other visiting period in the future providing higher utility from self\((\varphi)\)’s viewpoint.\(^{23}\) For \( S \) the strategy is a bit different since he, unlike \( N \), understands that his future intertemporal preferences are subject to change. From this it follows that self\((\varphi)\) of \( S \) completes the visit in ongoing period if and only if there does not exist a self in tolerable time who would complete the visit. This tolerance is defined next.

**Definition 2**  
A tolerance for self\((t)\) of type \( \varphi \) is \( \kappa^\varphi_t \equiv \max_{\kappa > t} \{ \kappa | U^\varphi_t (t) < U^\varphi_t (\kappa) \} - t \) and \( \kappa^\varphi_t \equiv 0 \) iff \( \max_{\kappa > t} \{ \kappa | U^\varphi_t (t) < U^\varphi_t (\kappa) \} - t = \emptyset \)

A tolerance for self\((t)\) tells us how many periods at maximum he would be willing to postpone the visit from his perspective. Note here that while \( \kappa^\varphi_t \) can be infinite, it does not necessarily mean that the visit is postponed infinitely. On the other hand, if the tolerance is 0, then the visit is immediate for sure. The next definition provides then a realising visiting period

**Definition 3** *(Revised from OR 1999)* A realising visiting period for a patient who observed the symptoms in period \( t^* \) is \( \kappa^\varphi_{t^*} \equiv \min \{ t | \sigma^\varphi_t = v \}, \varphi \in \{ T, N, S \} \).

\(^{23}\)

It is reasonable to note here that while \( T \)’s and \( N \)’s strategies seem to be the same, \( T \)’s strategy is based on perfectly known future behaviour while for \( N \) perceptions about future behaviour are fallacious.
In our model the intertemporal expected utility takes the following forms

\[
U_t^n (t) \equiv (w - m (\rho_t \gamma + (1 - \rho_t))) H_s + \beta (s, \varphi) \frac{\delta w H_o}{1 - \delta},
\]

\[
U_t^n (t + n) \equiv \sum_{z=1}^{n} P_{t,z} U_t^n \text{(cure}_{t+z}) + R_{t,n} U_t^n \text{(seek}_{t+n}) \text{ for } n \geq 1, \tag{1}
\]

where

\[
P_{t,n} \equiv (1 - \rho_t) \text{ for } n = 1,
\]

\[
P_{t,n} \equiv (1 - \rho_{t+n-1}) \prod_{\tau=0}^{n-2} \rho_{t+\tau} \text{ for } n > 1,
\]

\[
R_{t,n} \equiv \prod_{\tau=0}^{n-1} \rho_{t+\tau},
\]

\[
U_t^n \text{(cure}_{t+n}) \equiv \left(1 + \beta (s, \varphi) \sum_{\tau=1}^{n-1} \delta^\tau\right) w H_s + \beta (s, \varphi) \frac{\delta^n w H_o}{1 - \delta},
\]

\[
U_t^n \text{(seek}_{t+n}) \equiv \left(1 + \beta (s, \varphi) \sum_{\tau=1}^{n-1} \delta^\tau\right) w H_s
\]

\[+ \beta (s, \varphi) \delta^n (w - m (\gamma \rho_{t+n} + (1 - \rho_{t+n}))) H_s + \beta (s, \varphi) \frac{\delta^{n+1} w H_o}{1 - \delta}.
\]

In Eq.(1) the first term on the right hand side is the expected utility from possibilities to get cured without visiting a doctor, and the second term is the expected utility in a case where he does not get cured without the visit before \(t + n\) and visits the doctor in that particular period.\(^{24}\)

To this end we define attractivity, which quantifies the patient’s willingness to visit a doctor by solving the lower limit for the consultation fee for which the agent still would like to postpone the visit at least one period.\(^{25}\)

\(^{24}\)For the derivation of the intertemporal utilities above see Appendix A.

\(^{25}\)For more detailed information how the attractivity has been derived see Appendix B
Definition 4 Given the expected utility, define the attractivity, $A_t^\varphi$, for self(t) of type $\varphi$ as follows

$$U_t^\varphi (t) < U_t^\varphi (t + 1)$$

$$\iff m > \frac{\beta(s, \varphi) \rho_t \delta ws}{(1 - \rho_t (1 - \gamma) - \beta(s, \varphi) \delta \rho_t (1 - \rho_{t+1} (1 - \gamma))) H_s}$$

$$\Rightarrow A_t^\varphi = \frac{\beta(s, \varphi) \rho_t \delta ws}{(1 - \rho_t (1 - \gamma) - \beta(s, \varphi) \delta \rho_t (1 - \rho_{t+1} (1 - \gamma))) H_s}.$$

Note that $A_t^N = A_t^S$, hence we use $A_t^T$ to refer $T$’s attractivity while $A_t^\varphi = A_t$ for $\varphi \in \{N, S\}$. Later on we will see that the attractivity gives handy tools for analysing the patient’s decision problem. For extensive use of tolerance and attractivity we use notation $_t x_n$ and $_t A_n$ to denote self(t)’s perception about self(n)’s tolerance and attractivity, respectively.

Finally, in each period the sole problem for the patient is to maximise $U_t$ by choosing when to visit a doctor given his perception about future selves behaviour.

Costs We assume that the total costs of the public health care system (PHCS) accrue from two different sources: (i) a consultation cost, $\hat{c}$, and (ii) a treatment cost, $c_t$. If a disease is serious the treatment costs start to increase from a base level, $c$, after symptoms have lasted more than $t^* + k^*$ periods.

In other words, when a disease is serious the treatment costs for the health care system are assumed to be the same from the period when symptoms occur for the first time, period $t^*$, until period $t^* + k^*$. From period $t^* + k^*$ onwards, the treatment costs start to grow up. We thus assume that while the time to certify severity of a disease is $k$ periods from observing the symptoms the non-cost-increasing time to visit a doctor is in the interval $[t^*, t^* + k^*]$. This means that if $k < k^*$ and the symptoms were observed in period $t^*$ the patient should first wait $k$ periods to be sure that he really is in need of a doctor and after that there are still $k^* - k$ periods time to actually visit a doctor before the treatment costs start to increase.

For illustrative purposes we use the following treatment cost function that captures the essential and described properties of the treatment cost formation

$$c_z = \begin{cases} 
  c, & z \in \mathbb{N}_{t}^{k^*} \\
  c_{a^z-k^*}, & z \in \mathbb{N}_{k^*}^{\infty} 
\end{cases}$$

26 Note the difference between the consultation costs and the consultation fee. The consultation fee is a slice of the consultation costs and it does not necessarily cover completely those.

27 If $k > k^*$, behaviour that would minimise the expected costs involves waiting $k^*$ periods and then seek for medical help. We, however, found the assumption $k > k^*$ implausible, and hence we concentrate only on the cases where $k \in [t^* - t^* + k^*].$
where $a > 1$ is a constant multiplier, $z$ the period of the visit.

**Financing** To make clear cut findings and emphasise important aspects, the structure of an economy is assumed to be very simple. There are $X_t$ (new born) individuals in each period. In period $t$ it is only the cohort $X_t$ that is responsible to cover costs from PHCS. This is to say that the taxes and the fees levied from $i \in X_t$ are used solely to cover costs that accrue from PHCS in period $t$. For simplicity, the tax rate to cover health care costs defines the net-of-tax income level for all periods.\footnote{One can consider this as an imaginary system of a community where members of age $n$ pay taxes to finance a publicly provided good $g_n$ and the sufficient tax rate for all the publicly provided goods is defined solely by sufficient tax rate to cover medical care.} For convenience we normalise $X_t = 1$ for all $t$.

The government has to finance total costs from PHCS that are denoted here by $K_t$. The financing is through taxation and the fee. The tax, $\tau^g$, the fee, $m$, and the level of deductible, $\gamma$, have to be set so that the following government’s periodical budget constraint will be satisfied.

$$K_t \leq X_t \tau^g w^g + V_t b - I_t (1 - \gamma) m$$

$$= \tau + (V_t - I_t (1 - \gamma)) m,$$

where $\tau = w^g \tau^g$ is the unit tax per individual, $V_t$ are users of health care system and $I_t$ are seriously ill individuals out of $V_t$ in period $t$.

### 3 Patient’s choice

We now turn to analysing the patient’s behaviour. We will show that for almost any given level of the consultation fee and the deductible rate the different types of the patient choose a different visiting time.\footnote{In O’Donoghue and Rabin (1999) it has been shown that naives are keen to procrastinate and sophisticates to prepropagate onerous task when costs are immediate. However, in their setup there does not exist any uncertainty.} This implies that it can be problematic for the government to implement such levels for the consultation fee and deductible rate that those would lead all patients to behave optimally from the government’s perspective, i.e. in the way that minimises the costs. Hence, the fraction of each type in the economy is crucial when choosing the levels for the tax, the consultation fee, and the deductible.
3.1 Analysis

From the perspective of any self with symptoms, postponing the visit increases the anticipated possibility to get cured without visiting a doctor, meaning saved money that can be consumed on other than medical expenses. At the same time postponing the visit means suffering from utility decreasing effect of symptoms. Finally, a delayed visit decreases the expected total payment from the visit. It is then clear that about the total effect from postponing the visit it is hard to say directly anything in terms of expected utility. The following three lemmas shed light on the shape of the expected intertemporal utility function.

We start by stating the basic characteristics of $A_{t}^{T}$.

**Lemma 1** $A_{t}^{T}$ is increasing in (i) $\beta$ and (ii) $t$; decreasing in (iii) $\gamma$, and (iv) in $s$ if

\[\varepsilon^{T} = \frac{H_{0}(1-\rho_{0}(1-\gamma)-\beta s_{0}(1-\rho_{0}(1-\gamma)))}{H_{0}(1-\rho_{0}(1-\gamma))},\]

where $\varepsilon^{T}$ is symptom elasticity of $\beta$, i.e. $\varepsilon^{T} = -\beta^{'}(s)^{\frac{3}{2}}$; (v) $\forall m \exists \gamma \in (0,1]$ such that $\lim_{t \to \infty} A_{t}^{T} \geq m \forall \varphi$; (vi) $\forall \gamma \in (0,1] \exists m$ such that $\lim_{t \to \infty} A_{t}^{T} < m \forall \varphi$

**Proof.** All proofs are consigned to Appendix C

Lemma 1 tells us that the attractivity is bigger for $T$ than for $N$ and $S$, and that visiting a doctor rather immediately than in the subsequent period becomes more attractive when time elapses. Naturally, postponing the visit is the more attractive the bigger is the deductible. An awkward property is that the attractivity is decreasing in symptoms severity. On the other hand, this happens only if the change of symptom level has big enough effect on appreciating the near future. Otherwise, higher symptoms make postponement less attractive. Finally, points (v) and (vi) in the lemma state that for any level of the fee the government can always choose the rate of deductible $\gamma = 0$ to assert that immediate visit is prefered over waiting one extra period for some self($t$). On the other hand, for any positive $\gamma$ the government could set that high fee that the patient would always find waiting for one more period attractive.

Next Lemma 2 simplifies analysis by stating that $U_{t}^{T}(\kappa)$ is single peaked in $\kappa \in \mathbb{N}_{t}^{\infty}$ from a perspective of self($t$), and that if one period waiting is not attractive nor is any longer waiting time. Together with Lemma 1, these findings imply that if the fee is not set too high for a given rate of deductible, the patient will eventually seek for medical help and infinitely long delays do not occur.

**Lemma 2** (i) if $A_{t+n}^{T} < m$ and $A_{t+n+1}^{T} \geq m$, then $U_{t}^{T}(t + k) < U_{t}^{T}(t + k + 1)$ $\forall k \leq n$, $\varphi \in \{T, N, S\}$, and $U_{t}^{T}(t + k) \geq U_{t}^{T}(t + k + 1)$ $\forall k > n$, $\varphi \in \{T, N, S\}$; (ii) if $U_{t}^{T}(t) \geq U_{t}^{T}(t + 1)$, then $U_{t}^{T}(t) \geq U_{t}^{T}(t + k)$ $\forall k \geq 1$.  

13
What is the import of Lemma (2)? It simply tells us that (i) from self(t)'s viewpoint postponing the visit is monotonically utility increasing for all selves and for all types $\varphi \in \{T, N, S\}$ until the period in which $T$ finds for the last time postponing still attractive. This implies one very important point, namely, if period $l$ is such that it maximises $T$'s intertemporal utility, then period $l$ is utility maximising point in time from the perspective of any such self(t) for whom $t \in \mathbb{N}_{l-1}$, of any type. Secondly, the lemma tells us that (ii) if one period postponement is not attractive then the patient considers that there is no reason to wait longer but rather to visit a doctor immediately. In other words, point (ii) in Lemma (2) characterises the first step in the patient’s multistep decision process. For the first step the patient solves for the expected utility from one period waiting and if that is utility increasing he then solves for whether two period waiting would be worthwhile and so on.

The following lemma completes the characteristics of patient’s decision making.

**Lemma 3** (i) If $A_t^T < m$ and $\lim_{t \to \infty} A_t = \lambda, \lambda > m$, then $x_t^T \leq x_t^N \forall t \in \mathbb{N}_\infty$, $\varphi \in \{N, S\}$, and for $x_t^T < \infty$, $x_t^T$ is decreasing in $t$ for all $\varphi \in \{T, N, S\}$ up to $n \in \mathbb{N}_\infty$ such that $x_n^T = 0$;

(ii) if $A_{t-1}^T < m$, $A_t^T \geq m$, and $A_t < m$, then $\forall \varphi \in \{N, S\}$ $\max_{m \in \mathbb{N}_0} \{n | U_t^\varphi (t + n)\} = 1$, and $U_t^\varphi (t + n) - U_t^\varphi (t + n + 1) \geq 0$ for all $n \geq 1$.

Lemma 3 asserts us that until the delay becomes intolerable for $T$, $x_t^T = 0$, tolerance $x_t^T$ is decreasing in time for all the types and always lower (or the same) for $T$ than (as) it is for $\varphi \in \{N, S\}$, and once $x_T^T = 0$ has been reached ($A_T^T \geq m$) the intertemporal utility is always from $N$’s and $S$’s perspective at its maximum level within one period from the ongoing period ($U_t^N (t + n) - U_t^N (t + n + 1) \geq 0 \forall n \geq 1$). It is exactly this property what makes $N$ and $S$ to procrastinate their visits compared to $T$. However, as the next propositions show us, procrastination is not as severe for $S$ as it is for $N$ while the quickest visitor is always $T$.

**Proposition 1** For any pair $(m, \gamma)$ such that $\lim_{t \to \infty} A_t = \lambda, \lambda > m$, there exists

(i) $k_t^T = t^* + p$ such that $A_{t^* + p - 1}^T < m$, $A_{t^* + p} \geq m$;

(ii) $k_t^N = t^* + n$ such that $A_{t^* + n - 1} < m$, $A_{t^* + n} \geq m$

(iii) $k_t^T \leq k_t^N$.

What is $S$’s behaviour relative to $N$ and $T$ then? $S$’s behaviour is different from $N$’s due to $S$’s correct perception about his future preferences. Self(t) of $S$ and $N$ postpones the visit if and only if there exists a self who will certainly visit a doctor in tolerable time from self(t)’s perspective.
However, $S$ knows the true future attractivities and tolerances whereas $N$ has completely fallacious perceptions about those. $S$ takes the future preference reversals into account in the decision making process and chooses his action strategically conditioned on his future behaviour. From the next proposition we see $S$’s realising visiting period relative to $T$ and $N$.

**Proposition 2**  For any pair $(m, \gamma)$ such that $\lim_{t \to \infty} A_t \to \lambda, \lambda > m$, we have $\kappa^T \leq \kappa^S \leq \kappa^N$.

The intuition behind Proposition 2 is the following. $T$ weights the future comparatively more than $S$ or $N$ and in addition has time-consistent intertemporal preferences. Hence, he visits the doctor as soon as it is valuable in terms of the expected subjective intertemporal utility. $S$, from his part, will always complete the visit within periods that are tolerable from the perspective of his first such self for whom the tolerance is finite. Contrary to $S$, $N$ can delay his visits comparatively long since after the period that satisfies $A^T \geq m$, he anticipates in every period that he will visit a doctor in a subsequent period with certainty. What really happens is that he does not visit a doctor before $A^N \geq m$ holds as well.

The most important implication of Proposition 2 is that the set level of the consultation fee and the deductible rate can results in different visiting periods for the different patient types. From the government’s perspective this is problematic. What level they should set for the fee? It is immediately clear that the fractions of different types in the economy become crucial when deciding about the optimal policy. These critical questions will be contemplated in an own section after the following subsection where, by some numerical examples, we still shed more light on the patient’s visiting behaviour.

### 3.2 Numerical examples

Since the realising visiting period for the patient of type $\varphi$ can be formally only characterised while the closed form solutions are in many cases impossible to solve or would not bring any insight about his behaviour, we rather use numerical methods to illustrate the realising visiting periods for different levels of the fee and the deductible.

Throughout the examples in this subsection we set $\tau = 0$, since we are here purely keen to contemplate the effects of $m$ and $\gamma$ on visiting behaviour for different $\varphi$. This is to say that tuples $(m, \gamma, \varphi)$ we use do not necessarily reflect the sufficient levels from the financial perspective which will be studied in the next section. The periodical net-of-tax salary $w$ will be set to unity, the level
of self-control to $\beta = .8$, time-consistent discounting to $\delta = .99$, the full health to $H_o = 1$, the level of positive symptoms to $s = .21$, and subjective probability for persistence to $\rho_{sub}(0|t^*) = .01$.

Finally, we assume that $t^* = 1$, i.e. symptoms occur in the first period. This gives a tuple of fixed parameters: $F \equiv (w, \beta, \delta, H_o, s, \rho, t^*) = (1, .8, .99, 1, .21, .01, 1)$. We investigate realising visiting periods for different types first with fixed deductible rate and varying fee and then with fixed fee and varying deductible.

### 3.2.1 Realising visits for fixed deductible rate and varying fee

Take $F$ as given and set $1 - \gamma = .3$. Solve then for realising visiting periods for the different types with $m \in [0, .9]$, that is, solve for $\kappa^T, \varphi \in \{T, N, S\}$ and $m \in [0, .9]$. That results in Figure 1.

![Figure 1](image)

Figure 1 illustrates clearly how $N$ is delaying his visit the most, and that his behaviour is monotonic in $m$, just like $T$’s behaviour is as well. This follows from the fact that any $N$’s self considers that his future incarnations will behave like $T$. Also the characteristics of $S$’s behaviour are clearly present in the figure. Since self(1) of $S$ never delays over his tolerance it follows that for any level of the fee the delays are always kept below (or equal with) the corresponding tolerance.
level. The non-monotonic behaviour then results in. Small changes in the fee affect S’s strategic behaviour that causes few periods up-and-down differences in the realising visiting period. There exists a point when delays between N and S start to diverge radically, while S’s trend follows the trend with T’s, and those differences are not large. In our calibration this happens when the fee reaches the level of .5. We can say that until a certain level for the fee, behaviour of all the types is pretty much the same, after that threshold level, increases in the fee cause much longer delays for N while for S it does not make big changes compared to T.

3.2.2 Realising visits for fixed fee and varying deductible rate

Take F as given and set first \( m = .3 \) and then \( m = .6 \). Solve then for realising visiting periods for different types with \( \gamma \in [0, .9] \), that is, solve for \( \kappa^*_1, \varphi \in \{T, N, S\} \) and \( \gamma \in [0, .9] \). That results in Figure 2.

From this figure we can easily see how the rate of deductible, \( \gamma \), affects realising visiting period. A direct observation is that decreasing the deductible is more efficient when the fee is higher. Both subfigures also assert that lowering the deductible works as an encouragement to visit a
doctor sooner, and it is most effective on N’s delays when the has been set relatively fee, and has approximately an equal effect on all the different types when the fee has been set relatively low.

4 Government’s task

The sufficiency of the tax rate was out of consideration in previous. We next add up cost formation and a sufficient tax rate into analysis and contemplate the problem from the government’s viewpoint. We will see that in this environment resulting sufficient tax rates have some very interesting properties, and that setting the fee and reimbursement is not an easy task in an economy with the different patient types.

Without loss of generality, we assume from now on that if an individual in $X_t$ is subject to get sick then $t^* = t$.

Consider period $t$ and denote by $P_t$ those individuals in $X_t$ who are having symptoms, and those who are having serious disease by $D_t$. By the assumptions, we have $P_t = \eta X_t = \eta$ and $D_t = \rho P_t = \rho \eta$.

The timing is following. The government first sets the consultation fee, the deductible and the tax rate. Then those individuals who have symptoms make their decision whether to visit a doctor now or later. Finally, the costs accrue from those who visit the doctor during the period and after sorting out the incomes and costs the budget will be in balance, surplus, or deficit.

4.1 Defining a sufficient tax level

Costs In the government’s budget constraint

\[
K_t \leq X_t \tau + (V_t - I_t (1 - \gamma)) m
\]

\[
= \tau + (V_t - I_t (1 - \gamma)) m,
\]

the total costs, $K_t$, and the quality of $V_t$ and $I_t$ depends on behaviour of those who are having symptoms. By the quality we mean how long an incoming patient has possibly waited before

---

30 Recall that the subscript $t$ in $X_t$ refers to a cohort born in period $t$.
31 Recall the notation and normalisation: we have normalised $X_t = 1 \forall t \in \mathbb{N}$, $V_t$ denotes the patients visiting medical center in period $t$ and $I_t$ denotes seriously ill out of $V_t$. 
seeking for medical help. We can write $K_t$ in the following general form

$$K_t = V_t \delta + I_t c_t,$$

where $c_t$ refers to treatment costs that is a function of time waiting time, i.e. how long seriously ill patients, $I_t$s, have delayed their visits. Treatment costs is also a function of the structure of $I_t$s. This structure, from its half, depends on the structure of each $X_t$, that is, on the fractions of $N$, $S$, and $T$ in the economy. Denote the fraction of type $\varphi$ by $\pi_\varphi$. Then, we have a partition of $X_t$ as $\varphi_t = \{\pi_N, \pi_S, \pi_T\}$ where $\pi_T = 1 - \pi_S - \pi_N$.

From the government’s viewpoint there are three different cost-categories in which an incoming patient $p$ can belong. Those are: a category in which (i) the probability for a serious disease is less than 1 and treatment costs are minimised in the case of serious disease; (ii) the probability for a serious disease is 1 and treatment costs are minimised; (iii) the probability for a serious disease is 1 and treatment costs have started to increase. For each described $p$, it holds respectively that: (i) $\kappa_{t^*}^N, \in \mathbb{N}_{t^*+k}^{f_N}$, (ii) $\kappa_{t^*}^S, \in \mathbb{N}_{t^*+k+1}^{f_S}$, (iii) $\kappa_{t^*}^T, \in \mathbb{N}_{t^*+k^*+1}^{f_T}$. For those who are subject to visit before $(k + 1)$th period after $t^*$ there is a $\rho$ probability that they really have a serious disease. On the other hand, out of those who plan to visit later than $k$ periods after the occurrence of symptoms, only the fraction $\rho$ really visits a doctor at some point, since the fraction $1 - \rho$ have cured by itself during periods $t^* + d, d \in \mathbb{N}_1^k$. Within any type group all the patients are identical. We thus get the following three expected total cost groups that are net-of-reimbursement:

$$B_t^\varphi = \begin{cases} 
\eta \pi_{\varphi} (\hat{c} + \rho c - (1 - \rho (1 - \gamma)) m), & \text{if } \kappa_{t^*}^{\varphi} \in \mathbb{N}_{t^*+k}^{f_N}; \\
\eta \pi_{\varphi} (\hat{c} + c - \gamma m), & \text{if } \kappa_{t^*}^{\varphi} \in \mathbb{N}_{t^*+k+1}^{f_S}; \\
\eta \pi_{\varphi} \rho (\hat{c} + ca^{\kappa_{t^*}^{\varphi}-(t^*+k^*)} - \gamma m), & \text{if } \kappa_{t^*}^{\varphi} \in \mathbb{N}_{t^*+k^*+1}^{f_T}.
\end{cases}$$

Hence, the total expected net-of-reimbursement costs caused by cohort $X_t$ during its lifetime are simply the sum of $B_t^\varphi$s over the types $\varphi$. Formally,

$$B_t = \sum_{\varphi} B_t^\varphi. \quad (2)$$
**Tax** For any \( m > 0 \) and \( \gamma \in [0, 1] \), the level of the tax must then satisfy

\[
V_t \hat{c} + I_t c_t \leq \tau + (V_t - I_t (1 - \gamma)) m
\]

\[
\Leftrightarrow \tau = V_t (\hat{c} - m) + I_t (c_t + (1 - \gamma) m).
\]

Finally, since the tax affects patients’ behaviour, the solution for the previous equation will be a fixed point, \( \tau^* \), that satisfies the following equality.

\[
\tau^* = V_t (\tau^*, m, \gamma) (\hat{c} - m) + I_t (\tau^*, m, \gamma) (c_t + (1 - \gamma) m).
\]

(3) Then, due to normalisation \( X_t = 1 \) \( \forall t \) and since all the cohorts are identical, we directly get the sufficient tax level for the steady-state from Eq. (2) by setting \( B_t \) equal to \( \tau^* \).

We have

\[
\tau^* = B_t
\]

From now on, we assume that the government knows that there can be exactly three different individual types, i.e. they know \( \varphi_i \in \{T, N, S\} \) for all \( i \). We also assume that the government knows the exact preferences for each type, i.e. they know the correct \( U_t^\varphi \) for each \( \varphi \). Now, for any pair \( (m, \gamma) \), the sufficient tax level \( \tau^* \) will depend on \( \varphi \). This means that if the government knows \( \varphi \), it will be able to set the equilibrium tax level \( \tau^* \) for any pair \( (m, \gamma) \). If it does not know the correct \( \varphi \), problems will emerge as we will see after the next subsection.

### 4.2 Social planner’s problem

As we are having hyperbolic individuals in our economy, a typical question arises: Whose utility the social planner should maximise? In here, we take the approach discussed in O’Donoghue and Rabin (2003, 2006) and we think that the social planner takes the viewpoint of self(\( t^* - 1 \)) when maximising the expected utility of a cohort \( X_{t^*} \). This is to say that no matter whether the individual in \( X_{t^*} \) is having hyperbolic or exponential preferences taking the viewpoint of self(\( t^* - 1 \)) guarantees that from that perspective the only correct discounting from period \( t^* \) onwards is the exponential

\[\text{As we mentioned } \tau^* \text{ is the tax rate in the steady-state, and in fact, we omit here the problem about financing the transition phase.}\]
one.

The government’s task is then to choose optimally the fee-deductible pair \((m, \gamma)\) resulting in \(\tau^*\) that maximises a weighted sum of expected intertemporal utilities subject to a balanced budget, where \(\pi_\varphi\), are from an assumed partition of \(\varphi\). The social planner then solves

\[
\max_{(m, \gamma)} \left\{ \sum_{\varphi \in \Phi} \pi_\varphi U_\varphi = \sum_{\varphi \in \Phi} \pi_\varphi \left[ (1 - \eta) \frac{W(\tau^*)}{1 - \delta} H_o + \eta G_t \right] \right\}
\]

s.t. \(K_t \leq \tau^* + (V_t - I_t (1 - \gamma)) m\),

where \(G_t\) is an expected intertemporal utility for an individual who meets the symptoms, and it takes the following form.

\[
G_t = \left\{ \begin{array}{l}
\sum_{n=0}^{K^*_t - t^*} \delta^n w(\tau^*) - \delta^{n+1} (1 - \rho (1 - \gamma)) \right\} H_s + \frac{\delta^{K^*_t - t^* + 1} w(\tau^*) H_o}{1 - \delta} & \text{if } \gamma^*_t \in N^{t^*_t + k} \\
\sum_{n=0}^{K^*_t - t^*} \delta^n w(\tau^*) - \delta^{n+1} \gamma m \right\} H_s + \frac{\delta^{K^*_t - t^* + 1} w(\tau^*) H_o}{1 - \delta} & \text{if } \gamma^*_t \in N^{t^*_t + k + 1} \\
\end{array} \right.
\]

Before analysing the problems what the government can meet when pursuing \(\tau^*\), we shortly illustrate numerically how the tax rate depends on set levels of \(m\) and \(\gamma\) for a correct partition of \(\varphi\).

**Sufficient tax level for \((m, \gamma)\)** For numerical illustration of the interaction between the tax rate and different levels of \((m, \gamma)\), we use the same calibration as in Subsection 3.2. We thus use the tuple of fixed parameters: \(F = (w^g, \beta, \delta, H_o, s, \rho, t^* ) = (1, .8, .99, 1, 01, 1)\), but now we have to complete it with partition \(\varphi = \{\pi_N, \pi_S, \pi_T\}\), the consultation cost \(\check{c}\), the base level treatment cost \(c\), and constant \(a\). We set \(\pi_\varphi = \frac{1}{3}\) for all \(\varphi \in \{T, N, S\}\), \(\check{c} = 2\), \(c = 160\), and \(a = 1.1\). Finally, we have to give a values for the time during which the seriousness of the disease is uncertain i.e. \(k\), and for the time when the treatment costs start to increase, i.e. \(k^*\). We set \(k = 10\), and \(k^* = 15\). Redefining \(F\) yields

\[
F = (w^g, \beta, \delta, H_o, s, \rho, t^*, \check{c}, \gamma, c, a, k, k^*)
\]

Let then \(m \in [0, 5]\) while the deductible takes the values \(\gamma \in \{0, .3, .6, .9\}\), and solve for \(\tau^*\).
This results in Figure 3.

From Figure 3 we can see that increasing the deductible rate (lowering the reimbursement rate) has a tendency to increase the tax rate after certain level of the fee. On the other hand, one can notice that increasing the deductible causes actually the patient to delay with lower levels of the fee, and so gives a possibility to lower the tax rate once some part of incoming patients have waited more than $k$ periods after symptoms occurred. Very important finding is that if the fee is set high enough, increasing the fee does not any longer decrease the tax rate but increases it. This is exactly the effect that is caused by unnecessary long delays. Moreover, if also the deductible rate is increased in this region, necessary tax increase will be even greater. If the government did not know partition $\varphi$ the social planner might consider that increasing either $m$ or $\gamma$ or both would lower the tax rate. However, the realisation is completely opposite and the government is actually forced to increase the tax rate. Once this happens the social planner might still consider that levels for $m$ and $\gamma$ are too low leading the government to increase them even more which again would result in even higher tax rate. As it is heavily present in the figure, the lowest tax rate will be attained with
the lowest possible fee if \( \gamma \) is set very high. On the other hand, this tax rate level is kept easiest on the cost of higher fee if the deductible \( \gamma \) is set as low as 0.

We next contemplate the possible problems that the social planner can face if it does not have full information about \( \psi \).

### 4.3 Choosing \((m, \gamma)\) can be problematic

We now turn to analyse the problems that the social planner faces when trying to solve the problem given in Eq. (4). As we mentioned earlier, the social planner solves Eq. (4) with an assumed partition of \( \varphi \). In here we mainly focus on contemplating what happens when this assumption fails to hold.

Basically, the social planner considers first the partition \( \varphi \), and then for all the given parameters he solves for pair \((m, \gamma)\) to maximise Eq. (4). Choosing the pair \((m, \gamma)\) that is based on an assumed \( \varphi \), denoted from now on by \( \varphi^A \), results in an equilibrium tax level \( \tau^{A*} \) that will be set, naturally, at the same time with \((m, \gamma)\). To make discussion easier let us denote this tuple by \((m, \gamma, \tau^{A*})\) where \( \tau^{A*} = \tau(m, \gamma, \varphi^A) \). Once the tuple \((m, \gamma, \tau^{A*})\) has been implemented it then results in a level for total costs that is a correlative of a real \( \psi \). If \( \varphi^A \neq \varphi \) it can happen that the budget deficit, negative or positive, occur.\(^{33}\) An interesting question is then what is the government’s best response on that, i.e. will they change the fee, the deductible, or the tax, or some combination of the previous, and what the chosen response can then affect for its part.

For convenience we assume that true distribution is \( \varphi = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\} \) (= #7) while the false distributions that the social planner is using are as follows\(^{34}\)

\[
\varphi^A = \left\{
\begin{array}{c}
\{1, 0, 0\} \quad \text{"Conventional" \ (#1)} \\
\{0, 1, 0\} \quad \text{"OR" \ (#2)} \\
\{0, 0, 1\} \quad \text{"Laibson" \ (#3)} \\
\{\frac{1}{2}, \frac{1}{2}, 0\} \quad \text{"Near-OR" \ (#4)} \\
\{0, \frac{1}{2}, \frac{1}{2}\} \quad \text{"OR-Laibson" \ (#5)} \\
\{\frac{1}{2}, 0, \frac{1}{2}\} \quad \text{"Near-Laibson" \ (#6)}
\end{array}\right.
\]

where \textit{Conventional} refers to conventional assumption of \( \beta = 1 \), \textit{OR} to O’Donoghue and Ra-

\(^{33}\)Notice that even if by coincidence \( \tau^{A*} = \tau \) for \( \varphi^A \neq \varphi \), the utility is not necessarily at its maximum level, and hence the equilibrium will be only suboptimal.

\(^{34}\)The notation \#n, \( n \in \{1, 2, ..., 7\} \), is just for numbering the assumptions \( \varphi^A \).
bin’s mostly assumed frameworks with assumption of only-naifs with $\beta \neq 1$, Laibson to Laibsons mostly used frameworks of only-sophisticates with $\beta \neq 1$, Near-OR to population with half-half time-consistents and time-inconsistents naifs, OR-Laibson to half-half naifs and sophisticates time-inconsistent, and finally Near-Laibson to half-half time-consistents and time-inconsistent sophisticates. For all other parameters the following calibration holds

$$F \equiv (w^g, \beta, \delta, H_o, s, \rho, \ddot{c}, c, a, k, k^*)$$

$$= (1, .8, .99, 1, .21, .01, 2, 160, 1.1, 10, 15).$$

The following table shows us the solutions for the social planner’s task given the different $\varphi^A$s. It also shows the actual budget balancing tax levels for chosen target $m$ and target $\gamma$, i.e. given target $m$ and target $\gamma$ for assumed $\varphi^A$ the sufficient budget balancing tax level is $\tau^*$ that would locally maximise Eq. (4) and balance the budget at the same time for true $\varphi = \{\frac{1}{5}, \frac{1}{3}, \frac{1}{2}\}$.

<table>
<thead>
<tr>
<th>$\varphi^A$</th>
<th>Target $m$</th>
<th>Target $\gamma$</th>
<th>Target $\tau$</th>
<th>Sufficient $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.5</td>
<td>.9</td>
<td>.0162</td>
<td>.0825</td>
</tr>
<tr>
<td>2</td>
<td>.3</td>
<td>.9</td>
<td>.0162</td>
<td>.0274</td>
</tr>
<tr>
<td>3</td>
<td>.4</td>
<td>.7</td>
<td>.0162</td>
<td>.0215</td>
</tr>
<tr>
<td>4</td>
<td>.9</td>
<td>0</td>
<td>.0216</td>
<td>.0198</td>
</tr>
<tr>
<td>5</td>
<td>.5</td>
<td>.3</td>
<td>.0162</td>
<td>.0211</td>
</tr>
<tr>
<td>6</td>
<td>.9</td>
<td>.2</td>
<td>.0162</td>
<td>.0354</td>
</tr>
<tr>
<td>7</td>
<td>.9</td>
<td>0</td>
<td>.0198</td>
<td>.0198</td>
</tr>
</tbody>
</table>

Table 1: Target levels of $m$, $\gamma$ and $\tau$ and correct balancing level of $\tau$ for different $\varphi^A$

It is remarkable that target $\tau$ is the same for almost all $\varphi^A$ while the fee and the level of deductible varies across $\varphi^A$. The similarity of the target levels is a follow up from the fact that tax rate $\tau = .0162$ is the lowest attainable tax level for the given parametrisation such that if $\varphi^A$ were the true distribution, all the patients with symptoms could be manipulated to wait the optimal time and after that to seek medical help within the time of that would minimise the treatment costs and hence minimise the expected total costs. Note also that due to the construction of the set up, we can immediately see from the differences between the target and sufficient tax rate levels whether the budget will be in balance or not. The biggest deficit is caused when Conventional distribution is assumed and the smallest when Near-OR is assumed. Finally, if the social planner assumed by coincidence $\varphi^A = \varphi$, then $m$ would be relatively high while the suggested deductible rate would
be 0. Hence, the optimal policy in full knowledge case would be to punish those quite hard who preproperate seeking medical help while rewarding generously those who have waited the optimal time when seeking medical help.

Table 1 gives us the starting point to contemplate the government’s attempts to balance the budget. As we now know that the budget will not be in balance with values where the assumption of distribution $\varphi$ is false the government must try to balance it. In here we will consider four possible corrective policies and their consequence: balancing the budget with changing (i) the fee; (ii) the deductible; and (iii) the tax level, and finally (iv) balancing the budget with all three balancing mechanism with different combinations of the fee, the deductible, and the tax, i.e. first with the fee, then if there is still deficit then with the tax, or first with the deductible and if there is still deficit then with the tax, or with all three in order of first with the fee, then with the deductible and finally with the tax. These corrective adjustment policies or treatments are denominated here by (i) $m^a$ or $m - only$, (ii) $\gamma^a$ or $\gamma - only$, (iii) $\tau^a$ or $\tau - only$ and (iv) $s=\tau^a$, and $s=\tau^a$, where $s, z \in \{m, \gamma, \tau\}, s \neq z$ depending on whether there are used one or two adjustment treatments before balancing the budget with the tax.

4.3.1 Balancing the budget

Let us start with an assumption $\varphi^A \neq \varphi$. For our parametrisation the assumption implies immediately,

\[ \tau^{A*} \neq B_i = \sum_{\varphi} B_i^\varphi, \]

and hence the budget will not be in balance. We assume that the government fixes two out of the three following variables $(m, \gamma, \tau^{A*})$ while they try to balance the budget by changing the non-fixed variable, referred here by choice variable. We will consider the following adjustment scheme: if there is deficit after setting the targets then the government increases the choice variable if and only if it resulted in lower deficit in the previous period or if the previous period was the target setting period, otherwise they decrease the variable; if decreasing the variable increases the deficit then the government increases the variable, and vice versa. All the increases and decreases are made by some rule so that the variable converges to the value that (locally) minimises the deficit during
Results  We have collected the main results in the following Table 2 and two figures, Figure 4 and Figure 5. Table 2 describes convergence results for all possible treatments while Figure 4 shows the same information in plotted diagram and Figure 5 the utility comparisons for using the different budget balancing treatments and assumptions about $\varphi$.

<table>
<thead>
<tr>
<th>$\varphi^A$</th>
<th>$m^A$</th>
<th>$m^a$</th>
<th>$\gamma^b$</th>
<th>$\gamma^a$</th>
<th>$s^* \gamma^a$</th>
<th>$s^* \gamma^a$</th>
<th>$s^* \gamma^a$</th>
<th>$s^* \gamma^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.5</td>
<td>.5243</td>
<td>.9</td>
<td>.623</td>
<td>.7239</td>
<td>.0162</td>
<td>.0825</td>
<td>.0434</td>
</tr>
<tr>
<td>2</td>
<td>.3</td>
<td>.3931</td>
<td>.9</td>
<td>1</td>
<td>.9003</td>
<td>.0162</td>
<td>.0274</td>
<td>.0406</td>
</tr>
<tr>
<td>3</td>
<td>.4</td>
<td>.4648</td>
<td>.7</td>
<td>.878</td>
<td>.7001</td>
<td>.0162</td>
<td>.0215</td>
<td>.0404</td>
</tr>
<tr>
<td>4</td>
<td>.9</td>
<td>.81</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.0216</td>
<td>.0198</td>
<td>.0201</td>
</tr>
<tr>
<td>5</td>
<td>.5</td>
<td>.51</td>
<td>.3</td>
<td>.3187</td>
<td>.3011</td>
<td>.0162</td>
<td>.0211</td>
<td>.026</td>
</tr>
<tr>
<td>6</td>
<td>.9</td>
<td>.9213</td>
<td>.2</td>
<td>.1687</td>
<td>.2002</td>
<td>.0162</td>
<td>.0354</td>
<td>.0378</td>
</tr>
<tr>
<td>7</td>
<td>.9</td>
<td>.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.0198</td>
<td>.0198</td>
<td>.0198</td>
</tr>
</tbody>
</table>

Table 2: Target and convergence levels for different adjustment treatments.

From Table 2 and Figure 4 we see that the target values fail to hold almost for all treatments and assumptions about $\varphi$. This is, of course, just a natural consequence from the fact that for all $\varphi^A$ different from $True$ the target values are set by assuming such patient behaviour that will not realise. Hence, fallacious assumption about $\varphi^A$ leads to falsely set $(m, \gamma, \tau)$ which, for its part, causes patients to visit a doctor too soon or too late from the government’s viewpoint. Thus, with the target values, the budget will not be in balance and there will be a need of using some adjustment scheme. From Table 2 we also see whether the single adjustment balances the budget or is there a need to use another adjustment instrument in addition. This can be seen by inspecting whether $s^a$ and $z^a s^a$ yields the same value for $s^a$, where $z \neq s$, $z, s \in \{m, \gamma, \tau\}$, and if it does not yield the same value then the budget can not be balanced by using only a single adjustment treatment.

---

$^{35}$We want to emphasise that due to non-monotonic behaviour of sophisticates there can be locally minimising points where the level of choice value starts to converge. That is, for some level of the fee sophisticates will choose to wait while if the fee is increased they will visit a doctor immediately. Hence, it depends on the initial values where the balancing sequence converges.

It is also reasonable to note here, that it is possible that the convergence is not necessarily complete but there can be some deficit even after balancing scheme is completed.
Figure 4

For the fee adjustment there is only a small difference between the target value and the convergence values if the planner assumes *Conventional*, *OR-Laibson* or *Near-Laibson* distributions. With all fee adjustment schemes and for all distributions, except *Near-OR*, the convergence value is higher than the target value and budget also stays unbalanced. For the deductible adjustment schemes, $\gamma^\alpha$ diverges from the target $\gamma$ for *Conventional*, *OR*, *Laibson* distributions while it follows the target quite accurately for the all other assumed distributions. There is also very clear pattern that either the fee is high and the deductible low or *vice versa*. The fraction of naifs in assumed partition steers the targets so that the more there are naifs assumed relative to other types the lower the fee is set to avoid unnecessary long and expensive delays while at the same time the level of deductible must be set relatively high to prevent too early and unnecessary visits. On the other hand, the more sophisticates and time-consistents are assumed to belong in the partition the harder 'the punishment' from an unnecessary visit becomes, i.e. the fee is substantially high, while the reimbursement for the necessary visits covers almost the whole fee. For the tax adjustment schemes, the target tax is almost always set to lower level than what will be attained after adjustment procedure. For *Conventional* distribution the biggest difference between the target level and the convergence value is with $m \times \tau^\alpha$ procedure that is nearly 900% higher tax rate compared to the

27
target, while the smallest difference is for Near-OR with any tax adjustment treatment which is also the only assumed partition for which the target is actually higher than the resulting tax level after an adjustment scheme. Compared to other adjustment schemes, the tax adjustment results in higher differences from the target values. That is an obvious consequence of possibly unbalanced budget that can be balanced only by using the tax adjustment in addition. It is notable that the tax affects all individuals while the fee and the deductible only patients. We find that \(\gamma \rightarrow \tau\) results in the lowest attainable tax rates that are actually also budget balancing solutions. This means that after choosing the targets, the social planner should keep the fee intact while letting the deductible first balance the budget and if that does not balance the budget completely then complete the balancing by adjusting with the tax. Only for Laibson, keeping the deductible fixed instead of fixing the fee results in lower tax rate after adjusting also with the tax than what \(\gamma \rightarrow \tau\) does. This feature is clearly present in the following Figure 5.

![Figure 5](image)

Figure 5 provides utility comparisons between the different adjustment schemes and the different distribution assumptions. The utility corresponds are calculated by using the True distribution and are based on realising behaviour. The first thing that one might wonder in Figure 5 is that why the assumption of True does not result in the highest expected intertemporal utility but it happens in
the case for which the assumed distribution is *Near-Laibson* and the adjustment is first made with
the deductible and continued with the tax adjustment is used. This peculiarity is not a mistake
but follows from the non-monotonic behaviour of sophisticates that causes local equilibria and
sometimes also invisibility of some 'better' equilibria. The reasoning follows. When the social
planner solves for target $m, \gamma$, and $\tau$ that maximises the expected intertemporal utility given in
Eq. (4) he first makes the assumption $\varphi^A$. After that he calculates the fixed point $\tau^* (m, \gamma)$ for each
pair $(m, \gamma)$ by starting the iteration from the current tax rate that is the sufficient tax rate in the
case where all patients visit a doctor immediately after observing the symptoms. After solving the
tuple $(m, \gamma, \tau^*)$ for each selected pair $(m, \gamma)$, the target $m, \gamma$, and $\tau^* (m, \gamma)$ are set by choosing the
tuple $(m, \gamma, \tau^*)$ that maximises Eq. (4) for an assumed $\varphi^A$. He then implements the chosen $(m, \gamma, \tau^*)$
and after the first period observes whether the budget is in balance or not. If it is not in balance he
chooses the treatment to balance the budget. But now, the initial values for the adjustment schemes
are the originally implemented values, i.e. the target values, that can depend on $\varphi^A$. Then, the
same adjustment scheme with different initial values can actually lead to different budget balancing
equilibria that can be such that they were not reached when selecting the target values. This is
exactly what happens in our calibration. The all-visit-tax-rate is now such that when assumed
distribution is *True*, the target values, which are now also the equilibrium values since $\varphi^A = \varphi$,
that are supposed to maximise the expected intertemporal utility do not maximise the expected
intertemporal utility globally but rather locally. Hence, it is a pure coincidence that assuming
(incorrectly) *Near-Laibson* instead of *True* results in such target values that when adjusting the
implemented fee, the deductible and the tax rate to balance the budget the iteration leads to a
'better' equilibrium that would not have been found if *True* had been assumed when iterating the
target values.

Assuming distribution *Conventional* results in, in general, lower expected intertemporal utilities
than what assuming other distributions would do. The adjustment with the tax only and first with
the deductible and then with the tax results in the highest utilities. It is also apparent from the
figure that when the true distribution includes naifs but the choice is made by assuming that there
are no naifs at all, i.e. by assuming distributions *Conventional, Laibson, and Near-Laibson*, the
intertemporal utilities are always lower than when the existence of naifs is assumed. This happens
since not assuming the existence of naifs leads to relatively high tax rates due to unexpected costs
from naifs’ delayed visits. High tax rates do then hurt not only patients but all individual, and
hence decreases the expected intertemporal utility. On the other hand, assuming that only naïfs exist while dropping out either the existence of sophisticates or time-consistents or both, lead always to higher utility level than what it is in the case where the existence of naïfs is dropped out. The reasoning is the same as in previous but now its the opposite, the budget balancing tax rates are now relatively lower and hence the expected intertemporal utility higher. We can then say that assuming the existence of naïfs never affects the targets so that when balancing the budget the resulting equilibrium does not cause time-consistents or sophisticates to behave harmfully for the system, while if naïfs are not assumed then the budget balancing equilibrium hurts naïfs by letting them to delay their visits to doctor and hence by the higher tax rate those delays can be seen as a negative externality for the whole system.

5 Conclusion

In this paper we have concentrated on studying the patient behaviour of hyperbolic individuals. While a time-consistent patient always seeks medical help in reasonable time we have seen that when also hyperbolic patients exist the optimal level of a consultation fee, deductible for it and a budget balancing tax rate can be very problematic to set. The problems follow from the fact that the different patient types choose different times to visit a doctor. If the difference between visiting times differs allot across the patient types then treatment costs and their formation become important. As the treatment cost can increase very fast in time after certain threshold, it is crucial for the public health care system that unnecessary long delays do not exist before seeking medical care. On the other hand, unnecessary visits to a medical centre can, as well, cause an increase in total costs of the system. Hence, the level of the fee, the deductible and the tax rate should be such that all the types wait long enough but not too long, when the costs from the system are financed by the fee and the tax.

In this paper the consultation fee makes an individual who meets symptoms to wait, while the deductible should make an individual willing to visit a doctor when he is sure that he really is in need of a doctor. The patient’s mechanism to certify himself about severity of a disease follows from a subjective probability that increases along the passage of time or then the symptoms vanish away which drops the probability to zero level.

The results establish that fairly small increases (decreases) in the fee (the deductible) can cause
substantial changes in patient’s behaviour. Naifs are affected the most and actually their delays can become so long that a decrease (increase) instead of an increase (decrease) in the fee (the deductible) would decrease the total costs from the system and hence the tax rate resulting in higher expected intertemporal utility from social planner’s viewpoint. For sophisticates and time-consistents it is more important to make them wait long enough with a high fee and then reward them with low deductible than to certify them to visit a doctor. For naifs, instead, it is more important to make them to visit a doctor soon enough with low fee and reminding them about unnecessary visit with high level of the deductible than to certify them to wait long enough. Finally, when all the types of patients do exist, assuming the existence of naifs is more important than assuming some other patient’s type existence, as assuming the absence of naifs can hurt sophisticates, time-consistents and all individuals with unnecessary high tax rates while assuming the absence of time-consistents or sophisticates does not hurt naifs in similar way and tax rates stay comparably low resulting in comparably higher expected intertemporal utility.

What is the true partition of naifs, sophisticates and time-consistents in the real world is naturally impossible to know. However, as it is plausible to assume that there are to a certain extent all those types, in the light of the results from this study, we should always conduct the policies for the public health care system so that naifs play the leading role. We should thus consider the possible behavioural effects of using fiscal instruments in health care more deeply than what it is currently done, since an assumption of the conventional partition for the individuals in the economy can actually lead to more expensive and worse situation than to what an assumption of the inconventional partition would lead us.

Appendix

A Expected intertemporal utilities

Take the viewpoint of self\((t)\). A subjective probability that he is still sick in period \(t + n - 1\) is

\[
R_{t,n-1} \equiv \rho_t \rho_{t+1} \cdots \rho_{t+n-2},
\]
Then, a subjective probability that he cures without visiting a doctor between periods $t + n - 1$ and $t + n$ is

$$1 - \rho_{t+n-1}.$$ 

Hence, the probability for being sick until $t + n - 1$ and then getting cured without the visit is

$$P_{t,n} \equiv (\rho_{t}\rho_{t+1}\ldots\rho_{t+n-2})(1 - \rho_{t+n-1}).$$

With this scenario the patient has intertemporal utility that is

$$U^\varphi_t (\text{cure}_{t+n}) \equiv \left(1 + \beta(s, \varphi) \sum_{\tau=1}^{n-1} \delta^\tau\right) wH_s + \beta(s, \varphi) \frac{\delta^n wH_o}{1 - \delta}.$$ 

In a scenario where the patient seeks for medical help in period $t + n$, he is sick until that period producing intertemporal utility of

$$\left(1 + \beta(s, \varphi) \sum_{\tau=1}^{n-1} \delta^\tau\right) wH_s.$$ 

while his expected instantaneous utility from that period is

$$(w - m (\gamma \rho_{t+n} + (1 - \rho_{t+n})))(H_s),$$ 

and from that period onwards the utility is given by

$$\beta(s, \varphi) \frac{\delta^{n+1} wH_o}{1 - \delta}.$$ 

Putting these utilities together we get an intertemporal utility for self($t$)

$$U^\varphi_t (\text{seek}_{t+n}) \equiv \left(1 + \beta(s, \varphi) \sum_{\tau=1}^{n-1} \delta^\tau\right) wH_s$$

$$+ \beta(s, \varphi) \frac{\delta^n wH_o}{1 - \delta}.$$ 

Naturally, the probability assigned to this scenario is $R_{t,n}$.

We are then ready to write out the ‘full’ expected intertemporal utility for self($t$) in a case where he considers to visit a doctor in period $t + n$ and takes it into account that he might get cured before that period. This utility is denoted by $U^\varphi_t (t + n)$, where superscript refers to the type, subscript to the viewpoint and $t + n$ to the period of the planned visit. We get

$$U^\varphi_t (t + n) \equiv \sum_{z=1}^{n} P_{t,z} U^\varphi_t (\text{cure}_{t+z}) + R_{t,n} U_t (\text{seek}_{t+n}),$$

where the first term on the right hand side is the expected utility from the cases where the would not be needed while the second term represents the expected utility from the visit in period $t + n$ conditional on having the symptoms still present.
B Derivation of attractivity

Self\(t\) would like to postpone the visit at least one period if

\[ U_t^S \lessdot U_t^S (t + 1) \]

\[ \iff \quad (w - (1 - \rho_t) m - \rho_t \gamma m) H_s + \beta \sum_{\tau = 1}^{\infty} \delta^\tau w H_o < \]

\[ wH_s + \beta \left[ \rho_t \left[ \delta ((w - (\rho_{t+1} \gamma + (1 - \rho_{t+1})) m) H_s + \sum_{\tau = 2}^{\infty} \delta^\tau w H_o \right] + (1 - \rho_t) \sum_{\tau = 1}^{\infty} \delta^\tau w H_o \right] \]

\[ \iff \quad wH_s - m ((1 - \rho_t) + \rho_t \gamma) H_s + \beta \sum_{\tau = 1}^{\infty} \delta^\tau w H_o < \]

\[ wH_s + \beta \rho_t \delta w H_s - \beta \rho_t \delta (\rho_{t+1} \gamma + (1 - \rho_{t+1})) m H_s \]

\[ + \beta \rho_t \sum_{\tau = 2}^{\infty} \delta^\tau w H_o + (1 - \rho_t) \beta \sum_{\tau = 1}^{\infty} \delta^\tau w H_o \]

\[ \iff \quad -m ((1 - \rho_t) + \rho_t \gamma) H_s < \beta \rho_t \delta w H_s - \beta \rho_t \delta (\rho_{t+1} \gamma + (1 - \rho_{t+1})) m H_s \]

\[ \iff \quad m ((1 - \rho_t) + \rho_t \gamma) H_s > \beta \rho_t \delta w s + \beta \rho_t \delta (\rho_{t+1} \gamma + (1 - \rho_{t+1})) m H_s \]

\[ \iff \quad m > \frac{\beta \rho_t \delta w s}{(1 - \rho_t (1 - \gamma) - \beta \delta \rho_t (1 - \rho_{t+1} (1 - \gamma))) H_s} \]

\[ \iff \quad m > \frac{\beta \rho_t \delta w s}{(1 - \rho_t (1 - \gamma) - \beta \delta \rho_t (1 - \rho_{t+1} (1 - \gamma))) H_s} \equiv A_t^S \]

C Proofs

C.1 Lemma (1)

Proof. (i) Derivate \(A_t^S\) w.r.t. \(\beta\).

(ii) We have to show that \(A_t\) is increasing in \(t\). To do this we have to show that \(\frac{\partial A_t}{\partial t} > 0\), where \(A_t\) is defined by the following equation

\[ A_t = \frac{\beta \rho_t \delta w s}{(1 - \rho_t (1 - \gamma) - \beta \delta \rho_t (1 - \rho_{t+1} (1 - \gamma))) H_s}. \]

Assume that symptoms occur in period \(t\) for the first time. Recall then that \(\rho_s = \rho^{t+s}\) is increasing in \(n\), and that \(n\) is the time distance from \(t^*\). For all \(s > 0\) \(s\) and \(H_s\) are redundant for the derivative with respect to \(\rho\), hence set \(H_s = 1\) and rewrite

\[ A_t = \frac{\beta \rho_t \delta w}{(1 - \rho_t (1 - \gamma) - \beta \delta \rho_t (1 - \rho_{t+1} (1 - \gamma)))}. \]

Now it is clear that when \(t\) increases the numerator increases. We are thus done if we are able to show, as a sufficient condition, that the denominator decreases in \(t\). Elaborate the denominator in
as follows

\[
(1 - \rho_t (1 - \gamma) - \beta \delta \rho_t (1 - \rho_{t+1} (1 - \gamma)))
\]

\[
= 1 - \rho_t (1 - \gamma) - \beta \delta \rho_t + \beta \delta \rho_t \rho_{t+1} (1 - \gamma) \equiv B_t.
\]

Note that

\[
\frac{\partial \rho(t+1)^{-1}}{\partial t} = -\frac{1}{(t + 1)^2} \rho(t+1)^{-1} \ln \rho.
\]

Now,

\[
\frac{\partial B_t}{\partial t} = (1 - \gamma) \frac{1}{(t + 1)^2} \rho_t \ln \rho + \beta \delta \frac{1}{(t + 1)^2} \rho_t \ln \rho
\]

\[-\beta \delta (1 - \gamma) \rho_{t+1} \frac{1}{(t + 1)^2} \rho_t \ln \rho - \beta \delta (1 - \gamma) \rho_t \frac{1}{(t + 2)^2} \rho_{t+1} \ln \rho
\]

\[= \rho_t \left( (1 - \gamma) \frac{1}{(t + 1)^2} + \beta \delta \frac{1}{(t + 1)^2} - \beta \delta (1 - \gamma) \rho_{t+1} \frac{1}{(t + 1)^2} - \beta \delta (1 - \gamma) \frac{1}{(t + 2)^2} \rho_{t+1} \right) \ln \rho
\]

\[= \rho_t \begin{pmatrix}
(1 - \gamma) \frac{1}{(t + 1)^2} - \beta \delta (1 - \gamma) \rho_{t+1} \frac{1}{(t + 1)^2}
+ \beta \delta \frac{1}{(t + 1)^2} - \beta \delta (1 - \gamma) \frac{1}{(t + 2)^2} \rho_{t+1}
\end{pmatrix} \ln \rho
\]

\[< 0,
\]

which completes the proof.

(iii) Derivate \( A_t^\gamma \) w.r.t. \( \gamma \).

(iv) Denote

\[
\frac{\beta \rho_t \delta ws}{B(t - \rho_t (1 - \gamma) - \beta \delta \rho_t (1 - \rho_{t+1} (1 - \gamma)))} H_s \equiv A_t = B_t \frac{s}{H_s},
\]

then the effect of the level of symptoms is given by the following derivative
\[
\frac{\partial A_t}{\partial s} = B_t \frac{H_o}{(H_o - s)^2} + \frac{s}{H_s} \left( \beta' \rho \delta w D + \beta \rho \frac{\delta w}{D^2} \left( 1 - \rho_{t+1} (1 - \gamma) \right) \right) \\
= B_t \frac{H_o}{(H_o - s)^2} + \frac{s}{H_s} \left( \beta' \frac{\beta}{D} \right) + \frac{\beta' \delta \rho_t}{D} \left( 1 - \rho_{t+1} (1 - \gamma) \right) \\
= B_t \left( \frac{H_o}{(H_o - s)^2} + \frac{s}{H_s} \left( \beta' + \frac{\beta' \delta \rho_t}{D} \left( 1 - \rho_{t+1} (1 - \gamma) \right) \right) \right) \\
= B_t \left( \frac{H_o}{H_o - s} + \frac{s}{H_s} \left( \frac{D + \beta \delta \rho_t}{D \beta} \left( 1 - \rho_{t+1} (1 - \gamma) \right) \right) \right) \\
= B_t \left( \frac{H_o}{H_o - s} + \frac{s}{H_s} \left( \frac{1 - \rho_t (1 - \gamma)}{D \beta} \right) \right) \\
= B_t \left( \frac{H_o}{H_o - s} + \frac{s}{H_s} \left( \frac{1 - \rho_t (1 - \gamma)}{D} \right) \right) \\
= B_t \left( \frac{H_o}{H_o - s} - \frac{s}{H_s} \left( 1 - \rho_t (1 - \gamma) \right) \right),
\]

where \( \varepsilon = -\frac{s \beta'}{D} \) is the \( \beta \)'s symptom elasticity. The derivative \( \frac{\partial A_t}{\partial s} \) is then positive if

\[
\varepsilon \beta < \frac{H_o D}{(H_o - s) (1 - \rho_t (1 - \gamma))}
\]

and otherwise negative.

(v) Notice and denote \( \lim_{t \to \infty} A_t^\gamma = \frac{\beta(\gamma)}{\beta \lambda(\beta, \gamma)} = \lambda(\beta, \gamma) \), since by points (i) and (iii) \( A_t^\gamma \) is increasing in \( \beta \) while decreasing in \( \gamma \), let \( \beta(\gamma) < 1 \). Then, for any \( \beta(\gamma) \in (0, 1) \) we have \( \lim_{\gamma \to 0} \lambda(\beta(\gamma), \gamma) = \infty \).

(vi) Fix \( \gamma = \delta, \delta' \in (0, 1) \), then \( \lambda(\beta(\gamma), \delta') = \frac{\beta(\gamma)}{\beta \lambda(\beta(\gamma), \delta')} < \infty \) for any \( \beta(\gamma) \in (0, 1) \). Choose \( m > \lambda(\beta(\gamma), \delta') \) to complete. ■

C.2 Lemma (2)

**Proof.** We first prove (ii) and then (i).

(ii) Let us prove even stronger claim: if \( U_t^f(t) \geq U_t^f(t+1) \), then \( U_t(t+n) - U_t(t+n+1) \geq 0 \) \( \forall n \geq 1 \). By transitivity then also \( U_t^f(t) \geq U_t^f(t+k) \forall k \geq 1, k \in \mathbb{N} \), and we are done. For notational ease let us denote \( U_t^f(\cdot) = U_t(\cdot) \) for \( \varphi \in \{N, S\} \). Assume that \( U_t^f(\cdot) \geq U_t^f(t+1), \) and \( \varphi \in \{N, S\} \). This implies \( U_t(\text{seek}_k) \geq \rho_t U_t(\text{cure}_{t+1}) + (1 - \rho_t) U_t(\text{seek}_{t+1}) \). By Lemma (1) \( U_t(t) - U_t(t+1) \) is increasing in \( t \), hence \( 0 \leq U_t(t) - U_t(t+1) < U_{t+k}(t+k) - U_{t+k}(t+k+1) \)
∀k ≥ 1. Choose an arbitrary n ≥ 1.

\[
U_t(t + n) - U_t(t + n + 1) = R_{t,n} U_t(\text{seek}_{t+n}) - P_{t,n+1} U_t(\text{cure}_{t+n+1}) - R_{t,n+1} U_t(\text{seek}_{t+n+1})
\]

\[
= R_{t,n} \left( U_t(\text{seek}_{t+n}) - \frac{P_{t,n+1} U_t(\text{cure}_{t+n+1})}{R_{t,n}} - R_{t,n+1} U_t(\text{seek}_{t+n+1}) \right)
\]

\[
= R_{t,n} \left( U_t(\text{seek}_{t+n}) - (1 - \rho_{t+n}) U_t(\text{cure}_{t+n+1}) - \rho_{t+n} U_t(\text{seek}_{t+n+1}) \right)
\]

\[
= R_{t,n} \left( U_t(\text{seek}_{t+n}) - ((1 - \rho_{t+n}) U_t(\text{cure}_{t+n+1}) + \rho_{t+n} U_t(\text{seek}_{t+n+1})) \right)
\]

\[
= R_{t,n} \delta^n \left( \frac{U_t(\text{seek}_{t+n})}{\delta^n} - \frac{(1 - \rho_{t+n}) U_t(\text{cure}_{t+n+1}) + \rho_{t+n} U_t(\text{seek}_{t+n+1})}{\delta^n} \right)
\]

\[
= R_{t,n} \delta^n (L - M),
\]

where

\[
U_t(\text{cure}_{t+n+1}) = \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s + \beta w H_s + \beta \frac{\delta w H_s}{1 - \delta}
\]

\[
= \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s + \beta w H_s + \frac{\delta w H_s}{1 - \delta} + w H_s - w H_s
\]

\[
= \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s - w H_s (1 - \beta) + U_{t+n}(\text{cure}_{t+n+1})
\]

\[
= \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s + \beta U_{t+n}(\text{cure}_{t+n+1}),
\]

\[
U_t(\text{seek}_{t+n+1}) = \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s + \beta w H_s
\]

\[
+ \beta \delta (w - m (\gamma \rho_{t+n+1} + (1 - \rho_{t+n+1}))) H_s + \beta \frac{\delta^2 w H_s}{1 - \delta}
\]

\[
= \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s + \beta w H_s
\]

\[
+ \beta \delta (w - m (\gamma \rho_{t+n+1} + (1 - \rho_{t+n+1}))) H_s + \beta \frac{\delta^2 w H_s}{1 - \delta} + w H_s - w H_s
\]

\[
= \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s - w H_s (1 - \beta) + U_{t+n}(\text{seek}_{t+n+1})
\]

\[
= \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} w H_s + \beta U_{t+n}(\text{seek}_{t+n+1}).
\]
Thus,

\[
\frac{((1 - \rho_{t+n}) U_t (\text{cure}_{t+n+1}) + \rho_{t+n} U_t (\text{seek}_{t+n+1}))}{\delta^n} = \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s + (1 - \rho_{t+n}) \left( \beta w H_s + \beta \frac{\delta w H_o}{1 - \delta} \right) + \rho_{t+n} \left( \beta w H_s + \beta \delta (w - m (\gamma \rho_{t+n+1} + (1 - \rho_{t+n+1}))) H_s + \beta \frac{\delta^2 w H_o}{1 - \delta} \right)
\]

\[
= \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s - wH_s (1 - \beta) + (1 - \rho_{t+n}) U_{t+n} (\text{cure}_{t+n+1}) + \rho_{t+n} U_{t+n} (\text{seek}_{t+n+1})
\]

\[
= \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s - wH_s (1 - \beta) + U_{t+n} (t + n + 1)
\]

\[
= \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s + \beta U_{t+n}^T (t + n + 1).
\]

On the other hand, we have

\[
U_t (\text{seek}_{t+n}) \frac{\delta^n}{\delta^n} = \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s + \beta (w - m (\gamma \rho_{t+n} + (1 - \rho_{t+n}))) H_s + \beta \frac{\delta w H_o}{1 - \delta}
\]

\[
= \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s + \beta \left( (w - m (\gamma \rho_{t+n} + (1 - \rho_{t+n}))) H_s + \frac{\delta w H_o}{1 - \delta} \right)
\]

\[
= \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s + \beta U_{t+n}^T (t + n).
\]

Hence,

\[
U_t (t + n) - U_t (t + n + 1) = R_{t,n} \delta^n (L - M)
\]

\[
= R_{t,n} \delta^n \left( \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s + \beta U_{t+n}^T (t + n) - \left( \frac{1 + \beta \sum_{\tau=1}^{n-1} \delta^\tau}{\delta^n} \right) wH_s - \beta U_{t+n}^T (t + n + 1) \right)
\]

\[
= R_{t,n} \delta^n \beta \left( U_{t+n}^T (t + n) - U_{t+n}^T (t + n + 1) \right) \geq 0,
\]

which completes the proof, since \(n\) was chosen arbitrarily.

(i) Assume \(A_{t+n}^T < m\) and \(A_{t+n+1}^T \geq m\). Notice that assumption implies \(U_{t+n}^T (t + n) < U_{t+n}^T (t + n + 1)\) and \(U_{t+n+1}^T (t + n + 1) \geq U_{t+n+1}^T (t + n + 2)\). Choose an arbitrary \(k \in \{1,...,n\}\). From the proof of (ii) we get

\[
U_t (t + k) - U_t (t + k + 1) = R_{t,k} \delta^k \beta \left( U_{t+k}^T (t + k) - U_{t+k}^T (t + k + 1) \right) < 0,
\]

\[
37
\]
since $A_i^p$ is increasing in $l$, i.e. $U_i^p(l) - U_i^p(l+1)$ is increasing in $l$, and we assumed that $U_{i+1}^p(t + n) - U_{i+1}^p(t + n + 1) < 0$. For the case $k > n$ the proof is analogous. Again, since $k$ was chosen arbitrarily the proof is completed. ■

C.3 Lemma (3)

**Proof.** (i) Assume $A^T_i < m$ and $\lim_{t \to \infty} A^T_i \to \lambda, \lambda > m$, and let $t = \max_{n \in \mathbb{N}_0} \{U^T_i(t + n)\} > t$. Denote $\delta_{t,n} = U^T_i(t + n) - U^T_i(t + n + 1)$ and remark that from the proof of Lemma (2) we get $\delta_{t,n} = R_{t,n}\delta^\alpha(s, \varphi) (U^T_{i+1}(t + n) - U^T_{i+1}(t + n + 1)) = R_{t,n}\delta^\alpha(s, \varphi) \Delta^T_{i+1,n}$ for $n \geq 1$. Define $R_{t,0} \equiv 1$. Notice that $|\delta_{t,n}|$ is decreasing in $n$ when $t + n < l$ and increasing in $t$ when $t + n \geq l$. Hence, $U_i(t + n)$ is concave in $n$ when $t + n < l$. Notice then that $|\delta_{t,n}|$ is decreasing in $t$ when $t < l$. We thus have $|U_i(t) - U_i(l)|$ is decreasing in $t \leq l$. Define then $S_{t,p} \equiv U_i(t) - U_i(t + p)$ and notice $S_{t,p} = \sum_{n=0}^{p} \delta_{t,n}$. By the definition of tolerance we have $\sigma^T_i = \max\{p|S_{t,p} < 0\}$. Without loss of generality, let $k$ be for self($t$) such that $S_{t,k} = 0$. Thus, $\sigma_i = k - 1$. Now, since $|\delta_{t,n}|$ is in decreasing in $t$ when $t + n \geq l$ it immediately follows that $\sigma_i^T = \sigma_i^T < \sigma_i^T$, and further, since $S_{t,p}$ is increasing in $t$, $\sigma_i + 1 < \sigma_i + n \forall n \in \mathbb{N}_{t-1}$. Finally, since $S_{t,p}$ is increasing also in $\beta$ it then follows that $\sigma_i^T \leq \sigma_i^T$ for $\beta \in \{N, S\}$ and for all t.

(ii) Notice that $\Delta^T_{i,n} < 0 \forall n \geq 1$ for $\varphi \in \{N, S\}$ if $\Delta^T_{i+1,n} > 0$. On the other hand $\Delta^T_{i+1,n} > 0$ if $A^T_i < m$. $A^T_i < m$ implies $A^T_i < 0$ since $A_i$ is increasing in $t$. Finally, if $\Delta^T_i < 0$ and $\Delta^T_{i,0} > 0$ then $\Delta^T_{i,0} < 0$ and $\Delta^T_{i,n} \geq 0 \forall n \geq 1$, and hence $\max_{n \in \mathbb{N}_0} \{U^T_i(t + n)\} = 1$. ■

C.4 Proposition (1)

**Proof.** (i) Let $p$ be such that $A^T_{i+1,p} < m$, $A^T_{i+1,p} \geq m$, $A^T_i \geq m$ implies $\sigma_i^T = 0 \forall n \in \mathbb{N}_0$. Without loss of generality, assume that $A^T_{i+1,p} = m$, then by using Lemma 2 and 3 it is easy to show that $\mu_i \equiv \min\{\sigma_i^T\} = t^* + p - t \forall t \in \mathbb{N}_{i+1}^{i+1}$ is strictly decreasing in $t$. Then clearly $\sigma_i^T = d$ for all $n \in \mathbb{N}_0$ and $\sigma_i^T = v$ for $n = p$. Thus $\sigma_i^T = t^* + p$;

(ii) Let $p$ be such that $A^T_{i+1,p} < m$, $A^T_{i+1,p} \geq m$. Since $A_i \geq m$, $A_i^N \geq 1$, and by Lemma 3 $\sigma_i^N = \sigma_i^N$. Since $\sigma_i^N = \sigma_i^N$ for all $i < j$ and $\mu_i$ is strictly decreasing in $t$ and we have $\sigma_i^N = d$ for all $n \in \mathbb{N}_0$. In addition, $\nu^T_{i+1,p+n} = 0 \forall n \in \mathbb{N}_0$ which implies that also $\sigma_i^N = d$ and $\sigma_i^N = d$. Thus, $\nu_i^N \geq t^* + p + 1$. Finally, $\sigma_i^N = v$ if and only if $A^N_{i,h} = m$ for some $h$, implying $\nu_i^N = 0$. Since $A_i$ is increasing in $\beta h \geq p$. Such $h$ exists since $\lim_{n \to \infty} A_i \to \lambda, \lambda > m$ by the assumption, and $A_i^N$ is increasing in $j$. To complete, denote $h$ by $n$, hence $\nu_i^N = t^* + n$. If $A^T_{i+1,p} < m$, $A^T_{i+1,p} > m$, and $A^T_{i+1,p} \geq m$, we have $\nu_i^N = \nu_i^N$.

(iii) The points (i) and (ii) $\Rightarrow \nu_i^N = t^* + n = t^* + h \geq t^* + p = \nu_i^T$. ■

C.5 Proposition (2)

**Proof.** Let $p$ and $n$ be such that $A^T_{i+1,p} < m$, $A^T_{i+1,p} \geq m$ and $A_{i+1,n-1} \geq m$, $A_{i+1,n} \geq m$, $n \geq p$. Proposition 1 (iii) gives us $\nu_i^N \leq \nu_i^N$. Now, $\nu_i^N = \nu_i^N \forall j \in \mathbb{N}$. By the definition $\sigma_j^N = v$ if and only if there does not exist a self($i > j$) in tolerable time for whom $\sigma_j^S = v$, while by the proof of 1 we have $\sigma_j^N = v$ if and only if $A_j^N \geq m$. Due to $\nu_i^N = \nu_j^N$ for all $j \in \mathbb{N}$, we have $\nu_i^N = \nu_i^N$ and only if $\nu_i^N \geq t^* + n$ if and only if $\nu_i^N \geq t^* + n \forall i \in \mathbb{N}_0$. By the proof of Lemma 3 it is easy to show that if $\nu_i^N = 1$
then necessarily \( \kappa_{t^*+n-1} = 1 \). From this it follows that self(\( t^* + n - 2 \)) completes the visit since self(\( t^* + n - 1 \)) would postpone it due to \( \kappa_{t^*+n} = 0 \). We have thus shown \( \sigma_{t^*+n-2}^N = v \) while \( \sigma_{t^*+n-2}^N = d \), which implies \( \kappa_{t^*+n-2}^N \leq \kappa_{t^*+n-2}^N \) since clearly \( \sigma_{t^*+n}^N = d \forall j \in \mathbb{N}_3^2, \forall \varphi \{N, S\} \).

To prove \( \kappa_{t^*}^T \leq \kappa_{t^*}^S \) we use Lemma 3(i). (I) Consider first the case \( n = p \). Then, \( \kappa_{t^*}^S = 0 \), and hence \( \sigma_{t^*}^p = v \forall \varphi \in \{T, S\} \). By Lemma 3(i) we have \( \kappa_{t^*}^S \geq \kappa_{t^*}^T \forall j \in \mathbb{N}_1^{p-t} \), in addition \( \kappa_{t^*}^S \) is strictly decreasing in \( t \) until \( t = p \) and hence \( \kappa_{t^*}^S > \kappa_{t^*}^T \forall i > j \). Thus, \( \sigma_{t^*}^p = d \forall j \in \mathbb{N}_1^{p-t} \), \( \varphi \in \{T, S\} \).

Hence, \( \kappa_{t^*}^S = p \forall \varphi \in \{T, S\} \). (II) Consider then the case \( n > p \). Then, \( \kappa_{t^*}^S = 0 \), and hence \( \sigma_{t^*}^T = v \).

Since \( \kappa_{t^*}^T \) is strictly decreasing in \( t \) until \( t = p \) and for all \( \varphi \in \{T, S\} \), we have \( \sigma_{t^*}^p = d \forall j \in \mathbb{N}_1^{p-t} \), which implies \( \kappa_{t^*}^T = p \). If then \( \kappa_{t^*}^S = 1 \), we know that \( \max_n \{n|U_p - U_p (p + n) < 0\} = 1 \). Hence, it must be that \( \max_n \{n|U_{p+1} (p+1) - U_{p-1} (p+1 + n)\} \geq \max_n \{n|U_p (p) - U_p (p + n) < 0\} + 1 = 2 \) resulting in \( \sigma_{t^*}^S = d \) and \( \sigma_{t^*}^T = v, d \). Thus, \( \kappa_{t^*}^T \geq p \) and \( \kappa_{t^*}^S \leq n \). Together (I) and (II) imply that \( \kappa_{t^*}^T \leq \kappa_{t^*}^S \). ■

References


