Incomplete draft. Please do not cite or quote without premission

Physician Information Acquisition In a Dynamic Setting

Rud Faden
University of Copenhagen
rud.faden@econ.ku.dk

March 16, 2015

Abstract

In this project, I examine provider and patient demand for information in a dynamic model where the diagnostic precision is assumed to be related to physician effort, and effort is non-contractible. In each period where the patient and physician interact, the physician gathers information about the patient, and the diagnostic precision is increased. Therefore, optimal physician effort decreases as the physician and patient tie increases. As the physician is unobserved, the insurer compensates the physician by the average effort in the physician population and physician will not provide an optimal level of diagnostic precision in the in the first encounters with a new patient. Therefore the switching cost of the patient increases as the tie with the physician lengthens. This model explains (i) why the cost is negatively related with patient, physician ties and (ii) also introduces the concept of an "information trap", where competition is deceasing in the patient physician tie as switching cost increases. Increases.

1 INTRODUCTION

It is a popular view among doctors that long term relationship between doctor and patient is a vital to good primary health care and promote patient satisfaction, and cost-effectiveness. Long term relationships are thought to increase in value because

> the practitioners come to know patients over time, and patients come to know the practitioners. The benefits of this knowledge [can]

> be expected to accrue in a variety of ways. For example, patients should make fewer visits be- cause many problems can be managed on the phone. Fewer hospitalizations should also result, since practitioners are more likely to be able to ascertain whether or not the problem [can] be managed at home (Starfield 1993, p. 41–42).

and

> it … takes time for physicians to understand and empathize with patients' values and feelings and to be able to help patients identify and utilize health care services that are appropriate for their condition and life situation…Decisions to adopt healthy habits, to stop smoking to spare a child from passive smoke …are more likely to be made if recommended by a trusted physician in the context of an ongoing relationship (Emanuel and Dubler 1995, p. 324–235).

find more recent quotation

This has to some degree been backed by empirical evidence. Using insurance data Weiss and Blustein (1996) find that long-term relations are associated with more preventive care, less hospitalization, and lower costs. However Weiss and Blustein does not attempt to separate matching effects from the patient tie effect, which makes the results hard to interpret. In another study on the physician knowledge about his patient, Hjortdahl and Borchgrevink (1991) finds that physicians with more knowledge about a patient is less likely to order lab tests, more likely to adopt a "wait-and-see" approach, and more likely to prescribe drugs and refer patients to specialists.

Despite this evidence, models that incorporate physician knowledge about their patients are scares, and no models exists which describes how physician knowledge is obtained.

In my model, information about a patient is seen as an economics good, which can be produced at a cost c by the physician. Further it attempts to describe the process of how information is accumulated by the physician and how this may both decrease cost and competition among physicians at the same time.

Very general models of learning have been developed by Grossman et al. (1977). However such a model quickly becomes very complex and as a results applications are not easy. Instead I limit myself to a class of monotone decision problems, which greatly simplifies the analysis.

2 THE PATIENTS UTILITY

Following Rochaix (1989), the patient has a utility function

$$U = u(t,s) : T \times S \longrightarrow \mathbb{R}$$

where t is treatment, s is disease variable classified by its severity of illness, where both t and s traverse a real line. It is assumed that the utility function has both increasing and decreasing parts to capture the negative effects of both under and over treatment. It is further assumed that the dis-utility of moving away from the optimum is increasing in s such that the patient is more *risk-sensitive* for higher values of s

Definition 2.1. Assuming that the decision problem has a unique solution $t^*(s)$ and for $s' > s$, $u(t^*(s), s) = u(t^*(s'), s')$. Then for $t^*(s) > t_1$, $t^*(s') > t_2$, $t_2 \geq t_1$, and $t^*(s') - t_2 = t^*(s) - t_1 = \Delta t$ the *risk-sensitivity* is monotonically increasing in s if

$$0 \leq u(t^*(s), s) - u(t_1, s) \leq u(t^*(s'), s') - u(t_2, s')$$

for all s and t and I write that $u(t_1, s) \leq u(t_2, s')$

The notion of *risk-sensitivity* is illustrated in fig. 1.

Theorem 2.1. *If the risk sensitivity is increasing in s then for $s' > s$, $u(t, s)$ has a single crossing property in (s, t)*

Proof. Let $t' = t^*(s')$ and $t = t^*(s)$. Given definition 2.1, it is clear that $t' \geq t_2 > t \geq t_1$. Therefore I can rewrite definition 2.1 as

$$u(t', s') - u(t_2, s') \geq u(t, s) - u(t_1, s) \geq 0 \tag{1}$$

eq. (1) clearly have increasing differences in s and thereby satisfies the single crossing property.

□

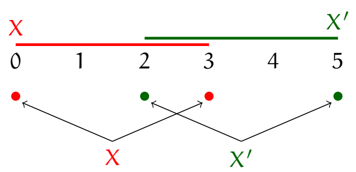
The intuition behind theorem 2.1, is that the marginal change $u(t', \cdot) - u(t, \cdot)$ is larger, when s is larger. When $u(t, s)$ is differentiable and concave as in fig. 1, one might note $\partial u(t, s)/\partial t \geq 0$ for $t \leq t^*(s)$ and that for $\partial u(t, s)/\partial t \leq 0$ for $t \geq t^*(s)$. Thereby $\partial u(t, s')/\partial t > \partial u(t, s)/\partial t$ for $t \leq t^*(s)$ and $\partial u(t, s')/\partial t < \partial u(t, s)/\partial t$ for $t \geq t^*(s)$. One might also note that $\partial u(t, s')/\partial t$ crosses $\partial u(t, s)/\partial t$ at most once, and only from below.

Proposition 2.2. *Given that $u(t', s) - u(t, s)$ has a single crossing propositionerty in s and that both S and T are well ordered sets (in the strong set order)*[1]*, then*

$$t^*(s) = \arg\max_{s \in S} u(t, s)$$

is increasing in s.[2]

[1] If $X' \geq X$ in the strong set order, then $\max(x', x) \in X'$ and $\min(x', x) \in X$. E.g. $[2, 5] \geq [0, 3]$ in the strong set order, while $\{2, 5\}$ and $\{0, 3\}$ is not.



[2] It should be noted that the assumption of quasi-supermodularity is not needed as the choice space is well ordered (e.i. a chain).

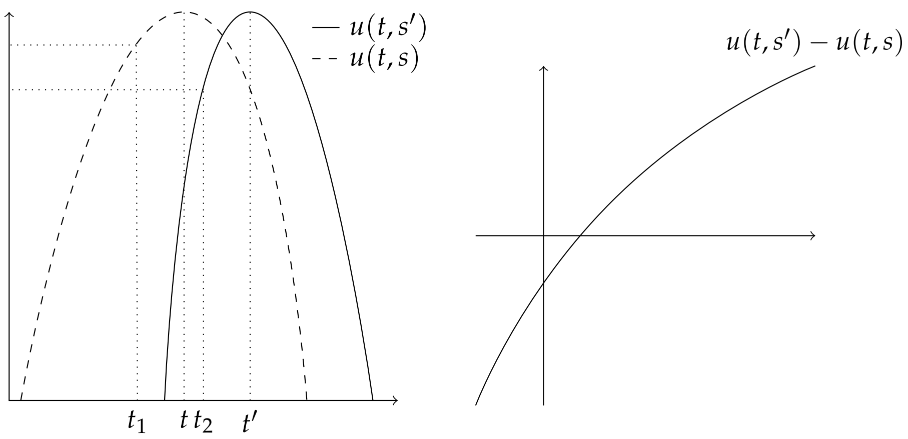


Figure 1: The patients utility functions. If for the same change in t the loss in utility is smaller for s than for s' when $s < s'$ then I says that risk sensitivity is increasing in s

Proof. As both T and S are real lines, then it follows trivially that they are well ordered. For the rest of the proof see Milgrom and Shannon (1994) or section A.3

□

Example 1. A function with the properties defined in theorem 2.1 and definition 2.1 and have the form as in fig. 1 is

$$u(t, s) = c - s(s - t)^2$$

where $t, s \in \mathbb{R}^+$. Assuming that example 1 is continuous and twice differential in t, s the derivative $\partial^2 u(t, s) / \partial t \partial s = 2s^2 \geq 0$ and example 1 has increasing differences and thereby also a single crossing property in (t, s)

∘

3 UNCERTAINTY WITH PERFECT AGENCY

In reality however, s is never observed. The level of severity for the patients is a random variable represented by S characterized by a subjective CDF. $F(s)$ with density $f(s)$ where s is a realization of S. The expected value of choosing an admissible treatment intensity t is given by

$$U(t, S) = E[u(t, s)] = \max_t \int_S u(t, s) \, dF(s)$$

It is however possible to acquire costly information about s through medical diagnostics and physician effort. However, for two experiment X, Y on S it is not a priori certain that one experiment X is necessary more *informative* about s than the experiment Y where *informative* is to be understood in the way the posterior decision induced by the experiment X insures greater expected utility than the decision induced by the experiment Y. Therefore we must introduce an order of information.

3.1 *Information ordering*

Definition 3.1. (Milgrom and Weber 1982) For a family of density functions, let $x \vee s$ denote the component wise maximum and $x \wedge s$ the component wise minimum. Then x and s are affiliated if for all s and x

$$f(s \vee x)\, f(s \wedge x) \geq f(s)\, f(x)$$

Affiliation of two random variables are equivalent to the monotone likelihood ratio property, and the intuition behind definition 3.1 is that higher signal realization of x makes the probability that s is large, higher. Similarly small signal realization of x makes the probability of a small s more likely.[3]

Athey (2002) shows that solution given in proposition 2.2 is robust to uncertainty, such that

$$t^*(x) = \arg\max_{x \in X} \int_S u(t, s)\, \mathrm{d}G^\eta(s \mid x)$$

is increasing in x whenever x and s are affiliated.[4]

Definition 3.2. (Persico 2000) Given two signals (experiments) X^η and $X^{\eta'}$, $X^{\eta'}$ is more accurate than X^η if

$$T_{\eta,s}(x) = F^{\eta'^{-1}}(F^\eta(x \mid s) \mid s) \tag{2}$$

is non decreasing in s for all x.[5] Let E be a real line. A family of signals $\{X^\eta\}_{\eta \in E}$, with support $X := \bigcup_{x \in E} X^\eta$, is accuracy ordered (A-ordered) if a signal with higher index is more accurate than a signal with lower index.

To understand the concept of accuracy, it can be noted that

$$T_{\eta,s}(X^\eta \mid s) \sim X^{\eta'} \mid s$$

Thus a more accurate signal can be obtained from a less accurate signal, by the transformation $T_{\eta,s}(X)$ For a better understanding of the accuracy concept, see examples 1 and 2.

Example 1. Let $\eta \in [0, \infty)$ and let S be distributed according to any CDF and let X_η be uniformly distributed on $[s - 1/\eta, s + 1/\eta]$ and let $f^\eta(x \mid s) = \eta/2$ on X^η. Then

[3] See section A.2 for more details

[4] Note that here we, unlike in theorem 3.3, consider the problem of deciding a treatment t before the state s is know, but *after* observing the signal x

[5] Note that eq. (2) can also be written as $F^{\eta'}(T_{\eta,s}(x) \mid s) = F^\eta(x \mid s)$ which implies that $T_{\eta,s}(x)$, solves $f^{\eta'}(x \mid s) = f^\eta(T^{-1}(y) \mid s) T'^{-1}(y)$

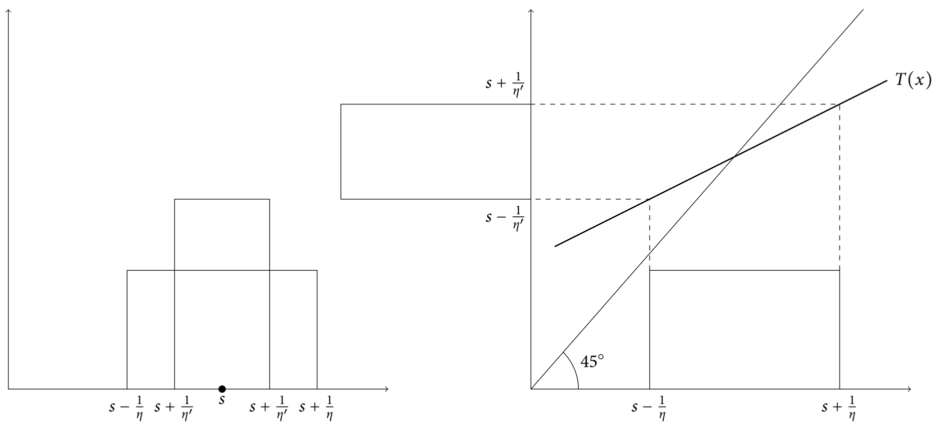


Figure 2: The $T_{\eta,s}(x)$ transformation

for $\eta' > \eta$

$$\frac{\eta'}{2} = \frac{\eta}{2}\left(T_{\eta,s}^{-1}\right)'(x) \Leftrightarrow T_{\eta,s}(x) = \frac{\eta}{\eta'}(x-s)+s$$

To see why, note that

$$T_{\eta,s}^{-1}(y) = \frac{\eta'}{\eta}(y-s)+s \Rightarrow$$
$$\left(T_{\eta,s}^{-1}\right)'(y) = \frac{\eta'}{\eta}$$

Further, one might also note that $T_{\eta,s}(x)$ is an increasing function by taking the derivative

$$\frac{\partial^2}{\partial\eta'\partial s}T_{\eta,s}(x) = \frac{\eta}{\eta'^2} > 0$$

Thus, $T_{\eta,s}(x)$ transforms X^{η} into $X^{\eta'}$ when $T_{\eta,s}(x)$ is increasing (Persico 1996).

From fig. 2, the ting to note is that $T_{\eta,s}(x)$ has a slope less than 1 and crosses the 45° line at s. This means that $T_{\eta,s}(x)$ contracts mass around s. Further, observe that $T_{\eta,s}(x)$ is a straight line passing through (s,s). Thus increasing s to s' will course $T_{\eta,s}(x)$ to shift up, so the line passes through (s',s'). Hence the notation of "more accurate signal" can be interpreted as one signal $X^{\eta'}$ being more correlated with the random state s than another signal X^{η}, and the function $T_{\eta,s}(x)$ imposes this additional correlation.

∘

An illustrative example can also be given by applying eq. (2) to hypothesis testing.

Example 2. Consider the case where s can take two values $s_1 < s_2$. Let X^{η} be a information structure affiliated with S. The optimal test based on X^{η} is given by the rejection region $X^\eta > x^*$ such that s_1 is rejected in favor of s_2 when $X^\eta > x^*$. The probability of a type I error is then $\mathbb{P}(X^\eta \leq x^*) = F(x^* \mid s_2)$ and the probability of a type II error is $1 - F(x^* \mid s_1)$. Now given that $X^{\eta'}$ is more accurate than X^η is is possible to design a test with the same probability of type I error, by choosing x^{**} such that $F(x^{**} \mid s_2) = F(x^* \mid s_2)$ (e.i. accept s_2 if $X^\eta \geq x^{**}$. However, since $X^{\eta'}$ is more accurate than X^η then $x^{**} \geq F^{\eta'^{-1}}(F^\eta(x \mid s_1) \mid s_1)$. As x^{**} lies on or, to the right of x^* then the test based on $X^{\eta'}$ is a least as powerful as the test based on X^η (Lehmann 1988; Persico 2000).[6]

∘

3.2 *Demand for information*

Given that we now know, when a test can be considered more informative than another, I can know turn to the problem of informativeness. To use the notation of informativeness I make two assumptions

Assumption 3.1. *The utility function is differential in t and the optimal solution $t^*(x)$ is differentiable in in η and x*

Assumption 3.2. *For all pairs of signals and states (s, x) the CDF $G^\eta(s \mid x)$ is differentiable in η on E and is continuous in s*

Theorem 3.3. *(Persico 2000) Suppose that X^η and $X^{\eta'}$ are affiliated with S and that $\{X^\eta\}_{\eta\in E}$ is A-ordered, such that $X^{\eta'}$ is more accurate than X^η. Then for all utility functions with a single crossing property, $X^{\eta'}$ is more informative than X^η*

Proof. See Lehmann (1988) section 4 and Karlin and Rubin (1956) Lemma 3–4 and theorem 1.

□

It follows directly from theorem 3.3 that when $X^{\eta'}$ is more informative than X^η then

$$\begin{gathered}\int_X \int_S u(t,s)\,\mathrm{d}G^{\eta'}(s \mid x)\,\mathrm{d}F^\eta(x) \geq \int_X \int_S u(t,s)\,\mathrm{d}G^\eta(s \mid x)\,\mathrm{d}F^h(x) \Leftrightarrow \\ U(t,s;\eta') \geq U(s,t;\eta)\end{gathered} \tag{3}$$

From the above equation, it is clear that the patient will always prefer a more accurate signal to a less accurate signal.

Persico (2000) goes on to show that when $u(t, s)$ is risk-sensitive increasing in s, then the marginal value of information

$$MR(\eta) = \frac{\partial}{\partial \eta} \int_X \int_S u(t,s)\,\mathrm{d}G^\eta(s \mid x)\,\mathrm{d}F^\eta(x) \tag{4}$$

[6] For a graphical example, see section A.1

is increasing in s.

From eq. (4) it follows straight forward that the optimal level of accuracies defined by

$$MR(\eta) - C(\eta) = 0 \tag{5}$$

is increasing in s, where $C(\eta)$ is the cost of obtaining the signal X^{η}. In all of this paper it will be assumed that the cost of signal acquisition is increasing in η, such that more accurate signals are more costly. The intuition is that when the risk sensitivity increases, information becomes more valuable.

3.3 *Information Aggregation and Ordering Over Time*

i now turn to the problem of aggregating information over time. The question is, if the physician receives a signal x_1 in period 1 and x_2 in period 2, when will the aggregation of the two signals overt time then be better than just one of them? Stated more formally, when will $MR(\eta(i))$ be increasing in time i?

Lemma 3.4. *if $g(\cdot)$ and $h(\cdot)$ are affiliated and non-negative, then $f(\cdot) = g(\cdot)h(\cdot)$ is also affiliated.*

Proof. See Milgrom and Weber (1982) or Topkis (1998, Collary 2.6.3)

□

1A consequence of lemma 3.4 is that the posterior of a density will be affiliated if both the prior and the likelihood function are affiliated.

Example 3. Assume that in the first period, the agent receives a signal x_1 and forms the posterior $h(s \mid x_1)$ where x_1, s is affiliated. In the second period the agent receives a new signal x_2. Lemma 3.4 then says that the posterior $f(s \mid x_2, x_1) \propto h(s \mid x_1)g(x_2 \mid x_1, s)$ is affiliated, if x_2, s is also affiliated.

∘

Lemma 3.4 is needed to guarantee that the optimal treatment is still an increasing function of the signal, even after updating, such that $t^*(x_i)$ is increasing in x_i.

I now turn to the problem of ordering information over time. Let $\eta(i)$ be the information order at time i after receiving $X^{\eta(i)} = X^1, \dots, X^i$ signals. It is still assumed that $\{X^{\eta(i)}\}_{\eta(i) \in E}$ and that higher $\eta(i)$ implies higher accuracy.

The question is then, under which conditions the expected utility of the patient is increasing in time. But first I introduce a technical lemma by Persico (1996) and one assumption that will be used in the proof

Lemma 3.5. *Let $[a, b]$ be in interval on $\mathbb{R}$. Let $J(\cdot)$ be a nondecreasing real function and $H(\cdot)$ be a quasi monotone increasing real function. Assume that for some measure $\mu \in \mathbb{R}$ it holds that*

$$\int_a^b H(s)\, d\mu(s) = 0 \tag{6}$$

then

$$\int_a^b H(s)J(s)\,\mathrm{d}\mu(s) \geq 0$$

Proof. Because of eq. (6) and the fact the $H(\cdot)$ is quasi-monotone there must be a part on $s_0 \in [a,b]$ before which H is non-positive and after which is it non-negative. Let $\bar{J} = J(s) - J(s_0)$: because J is non-decreasing, $\bar{J}$ non-positive before s_0 and non-negative after. I can then write

$$\int_a^b H(s)J(s)\,d\mu(s) = \int_a^b H(s)\bar{J}(s)\,\mathrm{d}\mu(s)$$

because of eq. (6). Rewriting the rhs the above to

$$\int_a^{s_0} H(s)\bar{J}(s)\,\mathrm{d}\mu(s) + \int_{s_0}^b H(s)\bar{J}(s)\,\mathrm{d}\mu(s) \geq 0$$

because of the fact that $H(\cdot)$ and $\bar{J}(\cdot)$ have the same sign on $[a, s_0]$ and $[s_0, b]$

□

Assumption 3.6. *i is a continues real variable and $\eta(i)$ is continuous and differential in i*

Theorem 3.7. *For a A-ordered family of signals the $MR(i) \geq 0$ iff (i) the T-transformation function $T_{\eta(i),s}$ is increasing in i (ii) a decision rule $t^{\eta(i)}(x)$ is quasi-monotone in the accuracy-order, such that higher A-order cannot lead to a decreasing $t^{\eta(i)}(x)$*

Proof. I want to show that

$$\frac{\partial}{\partial i}\int_X \int_S u(s, t^{\eta(i)}(x))\,\mathrm{d}G^{\eta(i)}(s \mid x)\,\mathrm{d}F^{\eta}(x) \geq 0$$

Given the definition of the T-function, I can write the above as

$$\frac{\partial}{\partial i}\int_X \int_S u(s, t^{\eta(i)}(T_{\eta(i),s}(x)))(s \mid x)\,\mathrm{d}G^{\eta(i-1)}\,\mathrm{d}F^{\eta}(x)$$

as the T-function transform the signal from the previous period in to the signal in this period.

The envelope theorem states that I can ignore the differentiation of the optimal treatment wrt. i. The inner integral can the be written as

$$\int_S \left[\frac{\partial}{\partial t} u(t^{\eta(i)}(T_{\eta(i),s}(x)))\right] \frac{\partial}{\partial \eta(i)} t^{\eta(i)}(x) \frac{\partial}{\partial i} T_{\eta(i),s}(x)\,\mathrm{d}G^{\eta(i-1)}$$

However the first-order-condition says that

$$\int_S \left[\frac{\partial}{\partial t} u(t^{\eta(i)}(T_{\eta(i),s}(x)))\right] \frac{\partial}{\partial \eta(i)} t^{\eta(i)}(x)\,\mathrm{d}G^{\eta(i-1)} = 0$$

The term $\left[\frac{\partial}{\partial t} u(t^{\eta(i)}(T_{\eta(i),s}(x)))\right]$ is quasi-monotone by the fact that it has a single crossing property. Quah and Strulovici (2009) discuss in proposition 9, under which circumstances and increasing decision rule under the signal η is also increasing under η'. If this is the case, then the term $\frac{\partial}{\partial \eta(i)} t^{\eta(i)}(x)$ is also quasi-monotone. Thus the first part of the integral is quasi-monotone. If $\frac{\partial}{\partial i} T_{\eta(i),s}(x)$ is a non-decreasing function of i then $MR(i) \geq 0$ by lemma 3.5.

□

The intuition behind theorem 3.7, is that the T-function transforms add correlation between the state s and the signal x only when $T_{\eta(i),s}$ is an increasing function.

In the remainder of this paper, I will assume (unless otherwise stated) that $\mathrm{d}/\mathrm{d}i > 0 \Rightarrow \mathrm{d}/\mathrm{d}\eta > 0 => \mathrm{d}/\mathrm{d}T_{\eta(i)} > 0$ when the $T_{\eta(i),s}$ is increasing.

4 THE PATIENTS PROBLEM

In this section I describe the patients decision problem in more detail. My main assumption about patient behavior is that the patient has a optimal level information $(\tilde{\eta})$, which is the solution to

$$\int_X \int_S u(t,s) dG^{\eta}(s \mid x) dF^{\eta}(x) = C(\eta)$$

One might think of this level of information, as the level of information, that the patient would chooses, if he were fully informed and payed the full cost. Another way to think of $(\tilde{\eta})$ is as the appropriate level of information, set by medical protocol or norms.

It is further assumed that the physician can store knowledge about patients, such that $\eta(i+1) \geq \eta(i)$ always hold.

5 PATIENT SEARCH IN A DYNAMIC SETTING

In this section, I will describe the equilibrium of the game. The games is as follows.

1. The patient consults a random physician
2. The physician chooses a effort level η and the receives a signal x and chooses $t^{\eta(i)}(x)$
3. The patient receives utility of $u(t^{\eta(i)}(x), s)$
4. The patient decides to stay or leave his physician

5.1 *The patients problem*

Let the the minimum effort level of the physicians be given by 0 and let the maximum level of effort be given by $\bar{\eta}$.[7] Assume that the physicians has some ethical

[7] Here a minimum level of effort equal to 0 can be interpreted as the minimum level of effort that is required by professional norms

constrain on how little effort they will provide and let $Q(\eta)$ be the distribution of this minimum effort. Then a patient that consults a random physician would expect to receive effort $E\eta = \int_0^{\bar{\eta}} \eta \, \mathrm{d}Q(\eta)$.

To keep things simple, I first analyze the game in a two period setting, $i = 1, 2$. If the patient stays, he will receive

$$u(t^{\eta(1)}(x), s) + u(t^{\eta(2)}(x), s)$$

and if he leaves he will receive

$$u(t^{\eta(1)}(x), s) + u(t^{E\eta}(x), s)$$

So the physician will only stay if

$$u(t^{\eta(2)}(x), s) > u(t^{E\eta}(x), s)$$

which is only true if $\eta(2) > E\eta$.

5.2 *The physicians problem*

The physician has a utility function given by

$$V = v(M, \eta)$$

where the physicians payment, M is given by

$$M = \delta + (p - \omega)t$$

where δ is a fixed payment, p is the physician payment per unit of treatment, t is quantity of treatment.

I also assume that the physician can increase the quality of treatment by increasing the diagnosis accuracy through effort, such that η increases proportionally to the effort applied. Thus η is the both level of accuracy in X^η and the physicians effort.

The cost of accuracy is given by a continues, increasing and convex cost function C. Assuming a separable utility function, the physician problem can be expressed as

$$v(M, \eta) = M - C(\eta)$$

As the subject of this paper, is the acquisition of information, through physician effort, I ignore the treatment decision and assume for simplicity that the payment for treatment p can be set exactly equal to the cost ω. The problem then simplifies to

$$v(\delta, \eta) = \delta - C(\eta) \tag{7}$$

.

The physician will not apply effort, when the value function is negative. Therefor the maximum effort that he will apply is

$$\delta = C(\eta)$$

which implies that

$$C^{-1}(\delta) = \bar{\eta}$$

Looking at the physician problem, the physician receives

$$\delta - C(\eta(1)) + \delta - C(\eta(2))$$

if the patient stays and

$$\delta - C(\eta(1)) + 0$$

if the patient leaves. Thus the physician will only apply effort

$$\delta - C(\eta(2)) > 0$$

However, the above equation is only positive if the patient stays. So the physician must apply more effort than the expected outside option. Therefore we have that

$$\delta > C(\eta(2)) > C(E\eta)$$

Finding the inverse we have that

$$\begin{aligned}
\bar{\eta} > \eta(2) > E\eta \\
\bar{\eta} > \eta(1) + \Delta\eta(2) > E\eta \\
\bar{\eta} - \Delta\eta(2) > \eta(1) > E\eta - \Delta\eta(2)
\end{aligned}$$

Thus in this 2 period setting we can conclude that (i) no physician will apply effort above $E\eta$, as doing so will result in a decreased effort level. (ii) physicians with $C(\eta(2)) > \delta$ will leave the market.

Expanding this to a i periods we can formulate the physicians constraint as

$$\sum_{i=2}^{n+1} \delta - C(\eta(i)) > 0$$

In order to calculate this stream of expected efforts that the patient will receive when leaving physician in period $j > i$, the patient must consider all possible exit strategies. That is, he must consider all exit strategies of the form: leave the current physician if the effort provided by this physician is below a given level assuming that he would randomly draw from the pool of physicians if he were to leave. let $\eta^{exp}(j)$ denote this optimal exit strategy. If we again take the inverse we can state the physicians n period problem as

$$\sum_{i=2}^{n} \bar{\eta} > \sum_{i=2}^{n} \eta(i) > \sum_{j=i+1}^{n+m} \eta^{exp}(j)$$

$$\sum_{i=2}^{n} \bar{\eta} > \eta(1) + \sum_{i=2}^{n} \Delta\eta(i) > \sum_{j=i+1}^{n+m} \eta^{exp}(j)$$

$$\sum_{i=2}^{n} \bar{\eta} - \Delta\eta(i) > \eta(1) > \sum_{i=2}^{n} \sum_{j=n+1}^{n+m} \eta^{exp}(j) - \Delta\eta(i)$$

From the above it should be clear that as long as the game is long enough, the physician will apply effort equal to 0. This might become more clear if I (unrealistically) assume that the physician applies equal amounts of effort in each period. If this is the case, I can rewrite the above as

$$n\bar{\eta} > \eta(1) + n\Delta\eta > m\eta^{exp} \tag{8}$$

$$\bar{\eta} > \frac{\eta(1)}{n} + \Delta\eta > \frac{m}{n}\eta^{exp} \tag{9}$$

The interpretation of eq. (9) is that: The longer the game (lifespan) the less the initial level of effort will matter. This is given by the term $\frac{\eta(1)}{n}$. Further, outside option for the patient is increasing in the time left in the game (e.i. $(m - n)$). This is given by the term $\frac{m}{n}\eta^{exp}$. Intuitively, the less time there is left in the game, the less time the patient has the recuperate the information that he looses from switching physician.

Proposition 5.1. *(i) No physician will ever apply effort above $E\eta$ and (ii) If the the game is long enough, there is a pooling equilibrium where all physician will end up applying 0 effort.*

REFERENCES

Amir, Rabah. 2005. "Supermodularity and Complementarity in Economics: An Elementary Survey". 71 (January 2004): 636–660. (Cit. on p. 16).

Athey, Susan. 2002. "Monotone Comparative Statics Under Uncertainty", *Oxford University Press* 117 (1): 187–223. (Cit. on p. 5).

Emanuel, E J, and N N Dubler. 1995. "Preserving the physician-patient relationship in the era of managed care." 273, no. 4 (): 323–329. (Cit. on p. 2).

Grossman, S, et al. 1977. "Bayesian approach to the production of information and learning by doing, Rev". 44 (3): 533–547. (Cit. on p. 2).

Hjortdahl, P, and C F Borchgrevink. 1991. "Continuity of care: influence of general practitioners' knowledge about their patients on use of resources in consultations." 303, no. 6811 (): 1181–1184. (Cit. on p. 2).

Karlin, Samuel, and Herman Rubin. 1956. "The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio". 27 (2): 272–299. (Cit. on p. 7).

Lehmann, E. L. 1988. "Comparing Location Experiments". 16 (2): 521–533. (Cit. on p. 7).

Milgrom, Paul, and Chris Shannon. 1994. "Monotone Comparative Statics". 62 (1): 157–180. (Cit. on p. 4).

Milgrom, Paul R, and Robert J Weber. 1982. "A Theory of Auctions and Competitive Bidding". 50, no. 5 (): 1089–1122. (Cit. on pp. 5, 8).

Persico, Nicola. 1996. *Information acquisition in affiliated decision problems*. (Cit. on pp. 6, 8).

— . 2000. "Information Acquisition in Auctions". 68 (1): 135–148. (Cit. on pp. 5, 7).

Quah, John K.-H., and Bruno Strulovici. 2009. "Comparative Statics, Informativeness, and the Interval Dominance Order". 77 (6): 1949–1992. (Cit. on pp. 10, 16).

Rochaix, L. 1989. "Information asymmetry and search in the market for physicians' services." 8 (1): 53–84. (Cit. on p. 2).

Starfield, Barbara. 1993. *Primary Care: Concept, Evaluation, and Policy*. 14:500. 4. (Cit. on p. 2).

Topkis, DM. 1998. *Supermodularity and complementarity*. (Cit. on pp. 8, 16).

Weiss, L. J., and J. Blustein. 1996. "Faithful patients: The effect of long-term physician-patient relationships on the costs and use of health care by older Americans". 86, no. 12 (): 1742–1747. (Cit. on p. 2).

A APPENDIX

A.1 *More informative & least as powerful test*

See fig. 3

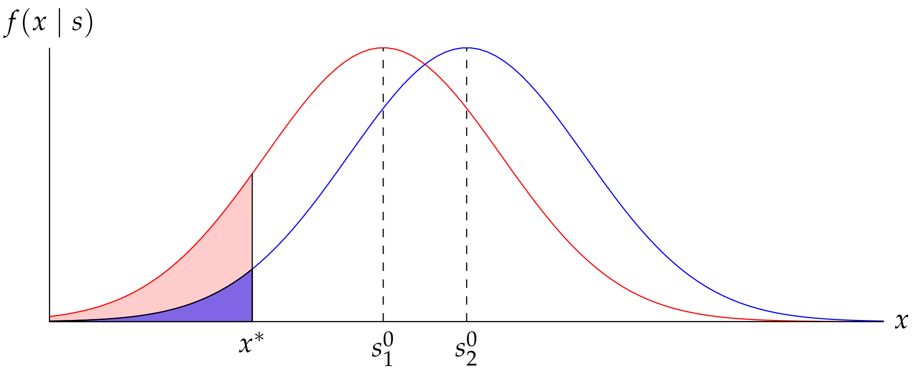
A.2 *Stochastic Affiliation*

Definition A.1. The random variables $X_1, \ldots, X_n$ are said to be affiliated if their joint PDF $f(\mathbf{x})$ is log-supermodular, meaning that for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_n b$, we have

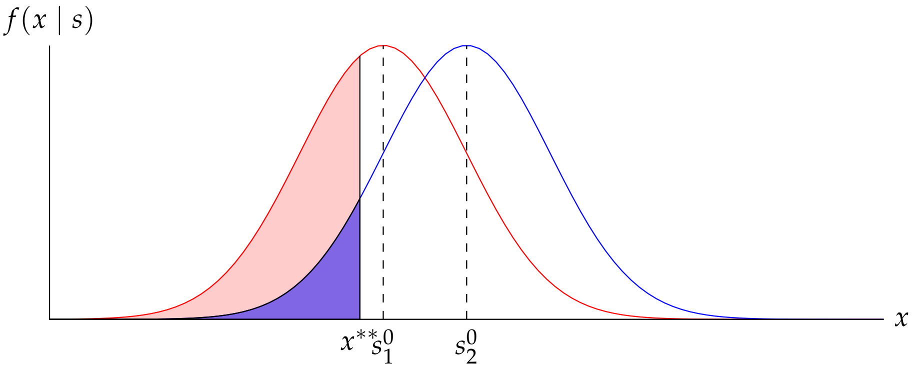
$$f(\mathbf{x} \wedge \mathbf{x}') f(\mathbf{x} \vee \mathbf{x}') \geq f(\mathbf{x}) f(\mathbf{x}')$$

When f is twice continuously differentiable, then equivalently $bX_1, \ldots, X_n$ are affiliated iff for all $i \neq j$

$$\frac{\partial^2}{\partial x_i \partial x_j} \ln f \geq 0$$



(a)



(b)

Figure 3: The blue shaded area is the probability of a type I error, and the red shaded area is the test power, while the non-shaded area under the red curve is the probability of a type II error. Figures 3a and 3b have the same probability of a type I error, but the power of the test in fig. 3b is higher than in fig. 3b, as the probability of a type II error is smaller

Consider two joint random variables X and S. Affiliation then tells us about the conditional distribution of $f(S \mid x)$. If $x \leq x'$, $y \leq y'$, then affiliation implies that

$$\frac{f(x, y')}{f(x, y)} \leq \frac{f(x', y')}{f(x', y)}$$

which we can rewrite as

$$\frac{f(y' \mid x) f_X(x)}{f(y \mid x) f_X(x)} = \frac{f(y' \mid x)}{f(y \mid x)} \leq \frac{f(y' \mid x')}{f(y \mid x')} = \frac{f(y' \mid x') f_X(x')}{f(y \mid x') f_X(x')}$$

which tells us the that likelihood ratio $f(\cdot \mid x) / f(\cdot \mid x)$ is increasing in x and thereby displays the monotone likelihood ratio property, which implies first-order (and hence second-order) stochastic dominance. Quah and Strulovici 2009, in proposition 8 and 10, proofs that the monotone likelihood ratio property is required for the family of functions with the single-crossing property to have an increasing optimal solution in the signal x.

A further property of affiliation is that it is preserved under Bayesian updating, such that if the prior $f(s')/f(s)$, $s' \geq s$ has a monotone likelihood ratio property, then the posterior $f(s' \mid x)/f(s \mid x)$ also have the monotone likelihood ratio property for any likelihood function $f(x \mid \cdot)$.

At last, affiliation is also preserved under multiplication, such that if $f(\mathbf{x}) = h(\mathbf{x})g(\mathbf{x})$, where g and h is nonnegative and affiliated, then f is also affiliated.

A.3 *A simplified proof of Topkis's Monotonicity Theorem*

Theorem A.1. *Consider the problem*

$$t^*(s) = \arg\max_{s \in S} u(t, s)$$

where $T, S \in \mathbb{R}$ and $T_s \subset T$ is the correspondence from S to T with T_s being the set of feasible treatments, when the diseases is s. Assume also that

(i) u has increasing differences in (t, s) and

(ii) $T_s = [g(s), h(s)]$ where $h, g : S \to \mathbb{R}$ are increasing functions with $g \leq h$

Then the maximal and minimal selection of $t^*(s)$, $\bar{t}(s)$ and $\underline{t}(s)$ is an increasing functions.

Proof. The proof is done by contradiction. Assume that $\bar{t}(s)$ is not increasing. Then for some $s' > s$ $\bar{t}(s') < \bar{t}(s)$. Then using assumption (ii) and the fact that $\bar{t}(s) \in T_s$ and $\bar{t}(s') \in T_{s'}$ it follows that $g(s) \leq g(s') \leq \bar{t}(s') < \bar{t}(s) \leq h(s) \leq h(s')$ so that $\bar{t}(s) \in T_{s'}$ and $\bar{t}(s') \in T_s$. Using the latter facts along with $\bar{t}(s) \in t^*(s)$ and $\bar{t}(s') \in t^*(s')$ we have

$$\begin{aligned}
0 &\geq u[s', \bar{t}(s)] - u[s', \bar{t}(s')] && \text{By optimality of } \bar{t}(s') \\
&\geq u[s, \bar{t}(s)] - u[s, \bar{t}(s')] && \text{by increasing differences *} \\
&\geq 0 && \text{by optimality of } \bar{t}(s)
\end{aligned}$$

which holds throughout. Hence it follows that $u(s', \bar{t}(s)) = u(s', \bar{t}(s'))$, such that $\bar{t}(s) \in t^*(s')$, which is a contradiction to the fact that $\bar{t}(s') = \max\{t^*(s')\}$ in the view of $\bar{t}(s') < \bar{t}(s)$. Hence, $\bar{t}(s)$ is an increasing function. The proff for $\underline{t}(s)$ is symmetric (Amir 2005).[8]

□

[8] In this proof, it is assumed that $T_s \cap T_{s'} \neq \varnothing$. If this was the case, one would have that $\sup T_s < \inf T_{s'}$ and it follows trivially have that $t^*(s) < t^*(s')$

A.4 *The inverse of a strictly convex function is concave*

Theorem A.2. *Let f be a real function which is convex on the open interval I, and let $J = f(I)$. Then if f is strictly increasing on I, then f^{-1} is concave on J.*

Proof. Let $X = f(x) \in J$ and $Y = f(y) \in J$. From the definition of a convexity, it follows that $\forall a, b \in \mathbb{R}_{++}, a + b = 1 : f(ax + by) \leq af(x) + bf(y)$.

Let f be increasing on I. Then f^{-1} is increasing on J, as the inverse of a monotone increasing function is increasing. Thus, $af^{-1}(X) + bf^{-1}(Y) = ax + by \leq f^{-1}(aX + bY)$, and hence f^{-1} is concave on J

□