PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

Thesis Draft
Master in Economics

León Tagle Le Blanc

as of 04-nov-2024

Contents

1 Introduction

The question of sickness at work presents us with a principal-agent problem: though in cases of true sickness it is best for the employee's personal wellbeing and indeed even firm productivity Chunyu et al. (2024) that they should stay home, it would also be possible to call in sick without any affliction, skipping work undetected.

The common solution to this problem is third-party certification, a specialized professional to attest to the condition and allow for or even demand of the employer that sick leave be granted. And who else to carry out this evaluation than physicians themselves—knowledgeable on the subject as they are—making diagnosis, treatment referral and certification of sickness a one-stop-shop for the patients.

The problem with this oversized, multi-faceted role physicians play in healthcare systems is that it causes conflict of interests. In cases where sick leave calls (and insurance claims) call for state-subsidied payments to be issued, we can speak of a second principal-agent problem between the physician, exercising the role of "gatekeeper", and the taxpayer, bankrolling public healthcare and insurance.

There is evidence to the fact that even in the best of cases physicians have no incentive to place doubt on their patients claim to unobservable symptoms (Carlsen et al.; 2020), and in murkier scenarios they could be motivated to gain the reputation of being "lenient" to enjoy increased demand for their services (Markussen and Røed; 2017).

In Chile, the country of our sample, the healthcare system is bi-modal: citizens may opt for one of many private health insurers (ISAPRES) or the sole public option (FONASA), and 82% of them are affiliates of the latter. In their case, the authority behind sick leave granting is COMPIN, with the power to oversee, investigate and even sanction physicians that fall under it. A worker granted sickness absence is entitled to a work incapacity benefit (SIL), periodic payments paid in function of sick leave length.

Anecdotal evidence speaks to the presence of fraud in the system. A survey by the U. Andrés Bello Public Health Institute (2024) has it that 51% of those surveyed know someone who was granted sick leave without any sickness, 62% believe that physicians "frequently" create irregular businesses for the sale of sick leave certificates, and 56% think it would be "easy" to purchase one.

The presence of fraud is not unknown to the overseeing authority. In September 2021, COMPIN sanctioned 188 physicians who were not able to justify their unusually high issuance rate. Oteíza (2023) finds that this one time intervention had a spillover on non-sanctioned doctors, who on average reduced their issuance rate by at least 5.22%, on conservative estimates. This is further proof that sick leave granting *is* influenced by profit-seeking, and readjusted expectations of the probability of being met with punishment does influence physicians' willingness to provide them to patients.

This paper seeks to develop a framework within which physicians' behavior may be rationalized and modeled. It presents physician in a dual role as both medical caretakers and issuers of sick leave, and patients value both aspects separately according to population distributions. Patient strategies are the probabilities with which they'll visit any given physician, and physician strategies is their choice of "leniency", how willing they are to grant sick leave according to patient characteristics.

A generalized framework of patient search is laid out and contrasted with a non-search baseline, i.e. a particular kind of equilibrium in which patient strategies are set as randomized over all physicians with no distinction, that is to say, in which they're assigned a physician with no say on the matter but whether to visit. Once patient 'search' comes into play, physicians take into consideration the behavior of the rest of their colleagues—and its implication on their own market share—when opting for a given level of strictness in sick leave issuance, meaning there's a *strategic* element now involved, where physicians will be wary to lose out on too many patients to more *lenient* colleagues.

We then present two particular specifications of the 'search' framework: an "implicit" search model, where patient behavior takes on the form of a modified McFadden Logit,[1] and an "explicit" search model, akin to sequential job search models, in which patients visit physicians which render them services above their "reserve utility".

The dual sources of patient utility—services and sick leave—combined with the strategic nature of the game, where each physician's strategy responds to the choices of all other physicians, render a closed form equilibrium equation infeasible, requiring numerical methods. We develop best-response algorithms to compute the various equilibrium strategies and market aggregates.

Finally, using data on sick leave issuance from Chile's national public insurer, we fit our model through a GMM exercise and run a counterfactual scenario. Being impossible for a policymaker to reliably police physician issuance to patients who don't require sick leave, as for conditions of low severity there's little scrutiny to their diagnosis, we compute the magnitude of the fine required so that, at the very least, *aggregate* sick leave issuance be at the level it would be if direct surveilance were possible.

Our result is that if the perceived probability of a fine or its magnitude were to increase around $1.64x$ at each level of sick leaves issued, physicians would endogenously adjust their strictness so that aggregate sick leave granting in the physician-patient market correspond to the level it would have if the policymaker could specifically prevent issuance to the 32% of patients of least severity, , including those with minimal or no symptomatology (where sick leave issuance constitutes fraud).

The empirical element of this paper, estimation, is somewhat hamstrung by the data available, chiefly the lack of information on overall visits to physicians rather than just the sick leaves they issued, and more observables on physician that could allow for an independent assessment of their "quality". Possible, or indeed needed extensions to this work if it should prove useful to policymaking, should begin with refining this

[1] See McFadden (1972).

aspect. Fitting the model thorugh GMM would greatly benefit from more varied and reliable data moments, as well as having more of the parameters involved be retreived from the data itself.

The rest of the paper proceeds as follows. Section 2 presents a brief review of the literature concerning physician markets, particularly those involving *prescriptions*. Section 3 presents the general theoretical framework of our model as well as two specifications. Section 4 addresses the numerical computation of equilibrium in these models. Our available data on sick leave issuance is covered in Section 5, which will be used in Section 6 to fit our model, calibrating the required parameters through GMM. Section 7 concludes with a simulated counterfactual, increased penalization for physicians, and some final comments regarding this whole endeavor.

2 Literature Review

It is a well established factum that taking sick leave is subject to an economic calculus on part of the workers, rather than being an orthogonal, merely health-concerned matter. Johansson and Palme (2005) begin their article with a quote by Nobel Laureate Ragnar Frisch: "Regarding the high absence rate at the Department: Acquiring minor diseases, such as colds or flu, is an act of choice". Their paper is among many others—Paola et al. (2014), Markussen et al. (2012), Stearns and White (2018), Henrekson and Persson (2004)—which give empirical evidence of such a choice being driven by economic incentives, through an event study on exogenous institutional regime changes in the sampled nation's public insurance system. This line of research, however, is concerned with the actions of workers themselves and their subsequent effect on macroeconomic employment variables, whereas our main focus shall be the role played by physicians.

As for the literature on physicians themselves, it has become commonplace to regard them as agents with two sources of utility: their income and their "altruistic" interest on their patients' well-being. Empirical support for the latter can be found in experimental evidence both on medical students (Brosig-Koch et al. (2017), Hennig-Schmidt and Wiesen (2014)) as well as doctors themselves (Kesternich et al. (2015), Brosig-Koch et al. (2016)). Crea (2019) finds no evidence for this, whereas Godager and Wiesen (2013) do, and explore its heterogeneity across physicians. The fact that physicians are also concerned with revenue, rather than being purely altruistic, is also well evidenced, see Clemens and Gottlieb (2014), Hennig-Schmidt et al. (2011), Autor et al. (2014), and also Robertson et al. (2012) for a review on the matter. Therein lies the dilemma with giving physicians the status of gatekeepers for different services and certifications, like disability insurance (as in autor). As Markussen and Røed (2017, p. 1) put it: "In essence, the GPs [general practitioners] have been assigned the task of protecting the public (or private) insurer's purse against the customers who form the basis for their own livelihood".

Both factors being well established in the literature on physicians, one strand of it would seek to design an optimal contract for medical care in the presence of such an economic calculus, see Choné and Ma (2011) or Gaynor et al. (2023); another, more in line with our approach, would evaluate the effects increased competition among doctors has on their rendered services. In general, Currie et al. (2023) propose that increased competition would lead to physicians offering more services that please the clients yet relatively hurt their own utility (like drug prescriptions), and less services which bring them, physicians, more utility, at the expense of patient utility (like unwanted, expensive surgical procedures). Iversen and Lurås (2000) and Iversen (2004) provide empirical evidence that somewhat supports it: physicians with a shortage of customers will provide more services, thus obtaining more income *per customer*. This line of work is more in line with our own approach, as the altruistic motivations of physicians are set momentarily aside and they're modeled as purely profit-seeking. This reasoning applies to scenarios like ours, where the issuance of sick leaves—particularly for short-term absences, as is the case in most situations[2]—does not have a significant impact on client health.

To our knowledge, the only article dealing specifically with sick leave certicate granting as a function of competition among physicians is Markussen and Røed (2017). Carlsen et al. (2020) deal with sick leave as well, but make only the narrower point that in a Bayesian context doctors have almost no incentive to distrust patients' self-reported, unverifiable symptoms. Markussen and Røed's methodology is similar to our own: after performing "raw" regression analysis, they set-up a model of patient choice as a McFadden Logit over observables X_i, including physician leniency (assumed to be observable for prospective patients), and as such can estimate the role leniency plays in demand for their services. They then perform a series of exercises, some of their findings include: physicians with variable wage (i.e. dependent on clientele) certify 7% more absence days per month than fixed wage physicians; half of this difference has a causal interpretation, as observed from within-physician responses to market conditions; in general, more lenient gatekeeping gives the GP (physician) more customers, and more customers make the GP less lenient. Most of these stylized facts are taken up in our equilibrium models, in which patients take into account both sick leave "leniency" ($\bar{\kappa_j}$) as well as physician quality (V_j), such that physicians which offer better medical attention and thus enjoy higher demand can afford to be "stricter", and those who don't will have the incentive to gain clientele through leniency. Not captured is the physician fixed effect, the idiosyncratic motive to leniency: we assume leniency as merely a strategic choice based on expected demand, where prior to choosing leniency physicians only differ in "quality" (V_j) and visit revenue/cost of visit to the patient (r_j, τ_j).

Despite the different subject matter, the main source of inspiration for this paper is Schnell (2022), such that ours can be seen as an attempt to replicate her model and framework, intially devised for opioid markets, to the market for sick leaves. Schnell seeks to model the market scenario for the opioid crisis, with a primary market composed of physician prescriptions and a secondary black market. The presence of the latter, she concludes, makes unilateral interventions ineffective: curbing both

[2] For cases where sickness leave is medically required we introduced the upper limit $\bar{\kappa}_{\max}$, such that physicians would never be so strict as to be negligent.

excessive physician prescriptions as well as black market sales is required to make more than a dent on the number of opioids consumed in America.

Her paper includes four benchmark for the patient-physician market, building up to her main model including patient search, such that patients with a hire taste for opioids are assortatively matched with physicians more willing to prescribe them, and a secondary market. In our model we keep the former but not the latter, as the "black market" for sick leaves falls *within* the primary market of physicians, composed of those willing to knowingly issue fraudulent sickness certificates. Search was re-purposed to fit a two-dimensional frame of physicians, characterized both by their strictness as well as their service quality. What we call the "explicit" search model is in the vein of Schnell's sequential patient search, though more fleshed out in its dynamic programming framework for the reason just mentioned. Then there's the "implicit" model, which defines patient behavior according to a modified McFadden Logit. We show in the Appendix A.2 that such a framework wouldn't have altered Schnell's main conclusions.

Our model differs from Schnell's chiefly in the fact that physicians take into account *other* physicians' behavior when selecting their own strategy, such that market equilibrium requires a Nash equilibrium in physician strategies. We discuss in the Appendix A.1 that the source of this feature, not present in Schnell's model, is the lack of additive separability across patients in the physician's utility function, such that her optimal behavior takes into account *aggregate* patient demand as well as *marginal.*

3 Model

3.1 The Physician-Patient Market

We now present the model itself before proceeding with further discussion. We consider $i = 1, ..., I$ patients and $j = 1, ..., J$ physicians.[3]

Patients are characterized by their "medical necessity" $\kappa_i \in \mathbb{R}^+$ and their "taste for sick leave" $\gamma_i \in \mathbb{R}$, respectively following the ex-ante cumulative distributions $F(\kappa)$ and $G(\gamma)$, which are public knowledge. Physicians are described by their "service quality" $V_j \in \mathbb{R}^+$, also known both to all patients and other physicians.

A patient i can visit a physician for treatment and may be also granted sick leave. After the patient is assigned a physician j, his utility function—implicitly dependent on his characteristic (κ_i, γ_i) tuple—is defined piecewise as follows:

$$U_i(V_j) = \begin{cases} \gamma_i + V_j\kappa_i - \tau_j & \text{if he's granted sick leave,} \\ V_j\kappa_i - \tau_j & \text{if he only visits the physician,} \\ 0 & \text{if no visit takes place} \end{cases}$$

[3]Note on language: For ease of reference, we refer to any patient as "he", and to any physician as "she" (as we have done already). At this point it shall be noted that the first person plural (we) employed over the course of this article is to be read as including the reader or perhaps as a royal *we*.

As we see, there are three components to patient utility: an interaction between the patient's medical need κ_i and the physician's service quality V_j which implies their complimentarity, his "taste" for sick leave γ_i in the case he's granted one, and τ_j, which we define as the cost of visit for physician j. As well as being complementary, κ_j as multiplied by V_j would give it the interpretation of being the marginal willigness to pay for physician quality.[4]

Whereas patients may visit at most one physician, a physician may see several patients. We define Q_j as the expected number of patients for physician j, the demand for her services. We say 'expected' because, as we'll see later on, patients may opt for a mixed strategy, assigning a certain *probability* to visiting j, and we define Q_j as the sum of the ex-ante probability of visit of all patients, not their ex-post realization.

As physician j has the option to issue sick leave to a given patient i which visits her, we likewise define X_j as the *expected* number of sick leaves granted by physician j, given her ex-ante patient demand and how many of them would be granted one.

We define the physician's utility function as follows:

$$U_j(Q_j, X_j) = R_j(Q_j) - P(X_j)$$

where $R_j(\cdot)$ is an individual, concave *revenue* function defined over expected total clients, and $P(\cdot)$ is a convex *punishment* function on X_j, composed of the probability of being fined for a certain number of sick leaves issued and the magnitude of the fine. The implication is that after a given number of patients, the disutility of an additional (expected) sick leave issued would outweigh physician j's financial incentive for further clientele. The j subindex indicates that we will allow $R_j(\cdot)$ (revenue by visit) to vary across physicians. We assume both $R_j(\cdot)$ and $P(\cdot)$ to be twicely differentiable.

We stress again that we define physician utility in terms of the *expected* realization of patient demand and granted sick leaves, as befits the logic of this game, where doctors take action before patients (we will specify the timing of our game below).

Following Schnell (2022), we focus on *threshold equilibria*, wherein each physician's strategy is the choice of a value $\bar{\kappa_j} \in \mathbb{R}_0^+$, such that of the patients who visit j, those with a κ_i value above or at that threshold will be granted sick leave, and those strictly under it won't. As a result of this, each physician will be identified by their given 'service quality', revenue function, cost of visit to the patient and choice of threshold: $\{V_j, R_j(\cdot), \tau_j, \bar{\kappa_j}\}$. The set $\{\bar{\kappa_j}\}_{j=1}^J$ will be known to patients when deciding on their behavior strategy.

Our models are different iterations of a general game with the following timing:

- First stage: Physicians simultaneously choose $\bar{\kappa_j}$.
- Second stage: Observing the set $\{V_j, \tau_j, \bar{\kappa_j}\}_{j=1}^J$, each patient chooses (or is assigned to) some doctor j.
- Third stage: Each patient can choose to see their physician and incur a visit cost τ_j, or refrain from doing so.

[4]Conceptually, willingness to pay would also depend on patient i's wealth level w_i. To steer away from discussions on wealth disparity, we interpret κ_i as the *average* willingness to pay for V_J.

Conditional upon his visit (as not to visit renders null utility), the utility of patient i from seeing physician j is:

$$U_i(V_j, \bar{\kappa_j}) = \begin{cases} \gamma_i + V_j\kappa_i - \tau_j & \text{if } \kappa_i \geq \bar{\kappa_j}, \\ V_j\kappa_i - \tau_j & \text{if } \kappa_i < \bar{\kappa_j} \end{cases}$$

We will usually use u_{ij} as a shorthand.

3.2 Non-Search Equilibrium

For illustration purposes, we first devote attention to a non-search baseline, where each patient is randomly assigned to a physician, with an equal probability of being matched to any of the J physicians. Their only say in the matter is whether they'll then visit physician j, which is to say, the second stage of the game is out of their hands, and they only make choices in the third stage after assignment.

A patient won't visit his assigned physician if his expected utility from such a visit is negative, we call this the *free disposal* requirement. As such, a physician j's expected patient demand, as a function of her threshold $\bar{\kappa_j}$ (and given the parameter V_j) will be the following:

$$Q_j(\bar{\kappa_j}) \; = \; \frac{I}{J}\left[\int_{\tau_j/V_j}^{\infty} dF(\kappa) + \int_{\min\{\bar{\kappa_j}, \tau_j/V_j\}}^{\tau_j/V_j} \int_{\tilde{\gamma}(\kappa)}^{\infty} dG(\gamma)\, dF(\kappa)\right] \tag{1}$$

where the left term consists of the mass of patients who just by virtue of physician j's service quality V_j would be willing to pay a visit (i.e. $\kappa_i \geq \tau_j/V_j$), and the right term would be patients who only see physician j solely out of the expectation of getting sick leave ($\kappa_i \geq \bar{\kappa_j}$ & $\gamma_i \geq \tau_j - V_j\kappa_i$), and wouldn't visit otherwise. We define $\tilde{\gamma}(\kappa) := \tau_j - V_j\kappa$ as the lower limit of the inner integral over γ.

Given that each patient with a κ_i higher or equal to $\bar{\kappa_j}$ is granted sick leave, the expected total number of such certificates granted by j, as a function of $\bar{\kappa_j}$, is:

$$X_j(\bar{\kappa_j}) \; = \; \frac{I}{J}\int_{\bar{\kappa_j}}^{\infty} \int_{\tilde{\gamma}(k)}^{\infty} dG(\gamma)\, dF(k) \tag{2}$$

which includes in theory patients from both the left and right term of equation (1).

For given formulations of the functions $Q_j(\bar{\kappa_j})$ and $X_j(\bar{\kappa_j})$, each physician solves for the following constrained optimization:

$$\bar{\kappa_j}^* \; \equiv \; \arg\max_{\bar{\kappa_j}} R_j(Q_j) - P(X_j) \qquad \text{s.a.} \quad 0 \leq \bar{\kappa_j} \leq \bar{\kappa}_{\max} \tag{3}$$

where $\bar{\kappa}_{\max} < \infty$ is the maximum value the choice of threshold of any physician may take, as we mentioned earlier.[2]

An inner solution to problem (3) takes on the following form because of the first order condition:

$$R_j'(Q_j)\frac{\partial Q_j}{\partial \bar{\kappa_j}} = P'(X_j)\frac{\partial X_j}{\partial \bar{\kappa_j}} \tag{4}$$

Proposition 3.1. *In the non-search model as described thus far, when the solution $\bar{\kappa_j}^*$ of problem (3) $\in (0, \frac{\tau_j}{V_j}]$, this $\bar{\kappa_j}^*$ is described by:*

$$R_j'(Q_j(\bar{\kappa_j})) = P'(X_j(\bar{\kappa_j}))$$

Proof. For $Q_j(\bar{\kappa_j})$ and $X_j(\bar{\kappa_j})$ as defined by (1) and (2), when $\bar{\kappa_j} \in (0, \frac{\tau_j}{V_j}]$:

$$\frac{\partial Q_j}{\partial \bar{\kappa_j}} = -\frac{I}{J}\int_{\tilde{\gamma}(\bar{\kappa_j})}^{\infty} dG(\gamma)\, f(\bar{\kappa_j}) = \frac{\partial X_j}{\partial \bar{\kappa_j}}$$

such that both derivatives are cancelled in equation (4). As a result, if the optimal $\bar{\kappa_j}^* \in (0, \frac{\tau_j}{V_j}]$, the proposition follows. □

This proposition implies that when for low enough values of κ_i, any new patients gained by physician j are those she entices through the expectation of getting sick leave, meaning that in the vicinity of $\bar{\kappa_j} \in (0, \frac{\tau_j}{V_j}]$ the change in expected patients $\Delta Q_j/\Delta\bar{\kappa_j}$ through a variation in $\bar{\kappa_j}$ is the same as the change in expected sick leaves issued $\Delta X_j/\Delta\bar{\kappa_j}$.

We can see that (3) offers three possibilities by way of solution: two corner solutions at 0 and $\bar{\kappa}_{\max}$, and an inner solution fulfilling equation (4). The proposition below proves such an inner solution can't hold for $\bar{\kappa_j} \in (\frac{\tau_j}{V_j}, \bar{\kappa}_{\max})$.

Proposition 3.2. *In the non-search model as described thus far, no $\bar{\kappa_j}$ in $(\frac{\tau_j}{V_j}, \bar{\kappa}_{\max})$ can be the solution of problem (3).*

Proof. For $\bar{\kappa_j} \in (\frac{\tau_j}{V_j}, \bar{\kappa}_{\max})$,

$$Q_j(\bar{\kappa_j}) \;=\; \frac{I}{J}\left[\int_{\tau_j/V_j}^{\infty} dF(\kappa)\right] \quad \text{and thus} \quad \frac{\partial Q_j}{\partial \bar{\kappa_j}} \;=\; 0$$

whereas

$$\frac{\partial X_j}{\partial \bar{\kappa_j}} = -\frac{I}{J}\int_{\tilde{\gamma}(\bar{\kappa_j})}^{\infty} dG(\gamma)\, f(\bar{\kappa_j}) \;<\; 0$$

so equation (4) can never hold, which would be required of $\bar{\kappa_j}$ as an inner solution to problem (3). □

The intuition for this proposition is that $\kappa_i = \tau_j/V_j$ defines the threshold above which patients are willing to visit their assigned physician j regardless of whether they get sick leave, so choosing such a $\bar{\kappa_j}$ is always either too strict or inefficiently lenient. We can speak of a "captured clientele". A similar behavior will be observed in the models with patient search, only that, since patients freely choose physicians, a "captured" patient won't be one which always visits j, rather one that always asigns *positive probability* to visiting physician j, no matter her choice of threshold $\bar{\kappa_j}$.

Equilibrium in the non-search model. *Given a physician-patient market specified by $\{\{(\kappa_i, \gamma_i)\}_{i=1}^{I}, \{(V_j, R_j(\cdot), \tau_j)\}_{j=1}^{J}, P(\cdot)\}$, we define equilibrium as a set of physician thresholds $\{\bar{\kappa_j}\}_{j=1}^{J}$ satisfying problem (3) and their corresponding equilibrium (expected) patient demand and issued sick leaves $\{(Q_j, X_j)\}_{j=1}^{J}$ as described by (1) and (2), respectively.*

This non-search baseline model, based off Schnell (2022), already presents some elements which will be present in our more complex models later, but completely neglects patient behavior, who are reduced to being assigned a physician and choosing to visit her or not. Patient search will allow us to formalize their free choice of physician as a function of the model parameters and physician strategies.

3.3 General Search Framework

We introduce patient 'search' as a general framework in which patients can choose freely among all physicians.

We define for each patient i a vector $S_i \in \Delta(\mathcal{J})$, where $\mathcal{J}$ is the the J-dimensional vector composed of all $1, ..., J$ physicians. S_i represents the *strategy* of patient i in this game, specifying the probability of visiting each physician j. Each component $s_{i1}, \ldots, s_{iJ}$ within S_i denotes the probability that patient i will visit physicians $1, \ldots, J$, respectively.

In order to play the role of a proper subprobability measure, for each patient i the components $s_{ij} \in S_i$ must fulfill the following criteria:

$$\forall j,\ s_{ij} \geq 0 \tag{i}$$

$$\sum_{j=1}^{J} s_{ij} \leq 1 \tag{ii}$$

We will allow the sum of all components for each patient i to be less than one (hence *sub*-probability), implying the presence of an *outside option* for patients, that is, to not visit any physicians. To have this option is important, as patient rationality in our models entails a "*free disposal*" property, meaning that a patient will never visit a physician if his expected utility from such a visit is less than 0, that is, $s_{ij} = 0$ if $U_i(V_j, \bar{\kappa_j}) = 0$. This makes the third stage of the game as described in Section 3.1 trivial, as under free choice patient i will only be "assigned" with positive probability to physicians he will be willing to visit.

We can re-interpret the non-search model as each patient being made to play by the strategy $\{S_i : s_{i1} = s_{i2} = ... = s_{iJ} = \frac{1}{J}\}$, and it is the lack of a free choice which makes the third stage non-trivial.

We shall further specify the form S_i will take. Consider now the J-dimensional vector U_i, where each component $u_{i1}, ..., u_{iJ}$ indicates the utility patient i expects from a visit to physicians $1, ..., J$ respectively. As a shorthand, we shall write $u_{i,-j}$ to indicate the $J - 1$ components of U_i excluding u_{ij}.

In the models considered, each component s_{ij} of S_i will be defined as:

$$s_{ij} \equiv g_i(u_{ij}, u_{i,-j})$$

where $g_i(\cdot)$ is a continuous function weakly increasing in the first argument u_{ij}, and weakly decreasing in the remaining arguments given by $u_{i,-j}$. Our model specifications will consist in giving this function $g_i(u_{ij}, u_{i,-j})$ a specific form.

Defining s_{ij} over u_{ij} makes it dependent upon patient i's characteristic tuple (κ_i, γ_i). We will use the $s_{ij}(\kappa_i, \gamma_i)$ formulation to define physician expected demand and sick leaves issued as functions over patient strategies:

$$Q_j(\bar{\kappa_j}, \bar{\kappa}_{-j}) \;=\; \int_0^\infty \int_0^\infty s_{ij}(\kappa, \gamma)\, dG(\gamma)\, dF(\kappa) \tag{5}$$

$$X_j(\bar{\kappa_j}, \bar{\kappa}_{-j}) \;=\; \int_{\bar{\kappa_j}}^\infty \int_0^\infty s_{ij}(\kappa, \gamma)\, dG(\gamma)\, dF(\kappa) \tag{6}$$

At face value, these definitions appear less informative than their non-search equivalents in (1) and (2). This is because s_{ij} is more general than the patient 'strategy' of the non-search model (of being assigned to any physician with equal probability), so we can't yet make claims about e.g. sections in the distribution of γ where s_{ij} is strictly 0. The two models we will introduce, the *implicit* and *explicit* search models, will build upon the general framework we've delineated and we will highlight their different qualitative implications.

For now though, notice that present we Q_j and X_j not only in terms of the threshold $\bar{\kappa_j}$ chosen by physician j, but also in terms of the thresholds of the other $J - 1$ physicians, which we abbreviate as $\bar{\kappa}_{-j}$. Whereas before in 'non-search' physicians were simply alloted a given number of patients, now they will *compete* for them, as patients' S_i behavioral strategy will consider the whole of $\{\bar{\kappa_j}\}_{j=1}^J$ when considering which physician(s) they will visit with positive probability.

As a consequence of physician competition for patients, equilibrium $\{\bar{\kappa_j}^*\}_{j=1}^J$ will need to take on the form of a Nash equilibrium, as physicians take into consideration the strategies of the other $J - 1$ physicians when choosing their own, such that they adapt to the influence they expect other physicians will have on their own "client" base, which we call the *strategic effect*. We explore in the Appendix A.1 the reasons for the presence of this effect, which is absent from Schnell's models, the basis for our own. We attribute it to the lack of *additive separability* across patients in the physician's utility function when one of $R_j(\cdot)$ or $P(\cdot)$ is *strictly* convex/concave.

Before moving on to our first model specification, we note that this general formulation of s_{ij} is sufficient to make the next proposition.

Proposition 3.3. *In the patient search framework as described, for two physicians j, l, such that $\bar{\kappa_j} \neq \bar{\kappa_l}$, $\bar{\kappa_j}$ and $\bar{\kappa_l}$ are strategic complements.*

Proof. The proof is in the Appendix A.3. □

3.4 The "Implicit" Search Model

The "implicit" search model is a McFadden Logit choice model modified to give null probability visits to physicians which afford patient i non-positive utility, so that the free disposal requirement is fulfilled. We also call it the Logit model.

The reason for calling it the "implicit" search model is that strictly speaking patient search is not formally included, yet one arrives at result qualitatively similar to specifications that do (like the "explicit" search model). The choice strategy noisily assigns positive probability to physicians who render i high utility, and the higher this u_{ij} expected utility is, the higher the probability of visit.

Instead of *explicitly* defining search and subsequent choice, probability of visit depends upon expected utility from physician j, u_{ij}, and a weighing parameter $\lambda \in \mathbb{R}$. The lower the value of λ, the noisier patient "search" is: they give high assingment probability to sub-optimal—yet feasible—physician choices. On the other hand, as $\lambda \to \infty$, patient choice is concentrated with probability approaching 1 at the physician with the highest V_j.

We define the components s_{ij} of the patient's strategy vector S_i as follows:

$$s_{ij} = \frac{\alpha_{ij}}{\sum_{k=1}^{J} \alpha_{ik}}, \quad \text{where} \quad \alpha_{ij} = \begin{cases} e^{\lambda u_{ij}}, & \text{if} \quad u_{ij} > 0 \\ 0, & \text{if} \quad u_{ij} = 0 \end{cases} \tag{7}$$

Unlike the "explicit" search model, the probability that patient i visits physician j is *strictly* growing in κ_i, rather than being a piecewise constant function with different levels. There is a discrete jump in probability at $\bar{\kappa_j}$, but elsewhere above 0 the function is smoothly increasing, rewarding physicians with high V_j.

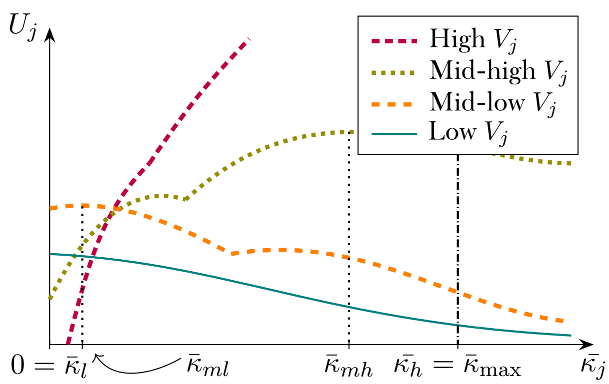


Figure 1: Physician utility U_j in the Logit model for different levels of V_j, as it varies across their choice of $\bar{\kappa_j}$

Figure 1 is a stylized representation of four "types" of physician, as ordered by their level of service quality V_j. The 'low' and 'high' types are driven to a corner solution at 0 and $\bar{\kappa}_{\max}$, respectively. We differentiate between two 'middle' types who choose inner solutions of $\bar{\kappa_j}$, as many physicians will present two local maxima across their curve, and their type is determined by which of the two is highest.

The reason curves present this piecewise behavior, as though composed of two individual concave curves joint together, is because for each physician j in this model there exists the threshold $\kappa_i = \tau_j / V_j$, as in the non-search model, above which patients always give positive probability to visiting j. Visiting chances may be *improved* upon by also granting sick leave to patients above that threshold (in terms of their κ_i), but taking it away doesn't drive their s_{ij} towards 0, as it would with patients whose $\kappa_i < \tau_j / V_j$.

This difference in incremental patient loss through a raise in $\bar{\kappa_j}$ between the $[0, \frac{\tau_j}{V_j}]$ and $[\frac{\tau_j}{V_j}, \bar{\kappa}_{\max}]$ sections of the curve accounts for the presence of two peaks for some values of V_j. We call 'mid-low' those physicians whose first peak is higher than the second, the others we call 'mid-high' (as they'd have a higher V_j for the parametrizations we use later in this paper).

The fact that the breaking point of the curve occurs earlier the higher the V_j results from the fact that τ_j grows at a lower rate than 1 for increases in V_j in our simulations in Section 4.2 on which we based this graph, but this need not be the case. Notice also that $\bar{\kappa_j}$ is growing in V_j, which isn't an analytical property from our model formulation, but an ex-post proclivity, which comes from using *well-behaved* (read: single-peaked, continuous) distributions for $F(\kappa)$ and $G(\gamma)$.

Equilibrium for the Logit model is formalized below.

Equilibrium in the "implicit" search model. *Given a physician-patient market specified by* $\{\{(\kappa_i, \gamma_i)\}_{i=1}^{I}, \{(V_j, R_j(\cdot), \tau_j)\}_{j=1}^{J}, P(\cdot), \lambda\}$*, we define equilibrium as a set of patient strategies* $\{S_i : s_{i1} \dots s_{iJ}\}_{i=1}^{I}$ *as defined by (7) and physician thresholds* $\{\bar{\kappa_j}\}_{j=1}^{J}$ *satisfying problem (3) with their corresponding equilibrium (expected) patient demand and issued sick leaves* $\{(Q_j, X_j)\}_{j=1}^{J}$ *as described by (5) and (6), respectively.*

3.5 The "Explicit" Search Model

The "explicit" search model, which we also call the sequential model, has patient strategies S_i be the *explicit* result of formalized physician search on part of the patient. This specification is more in line with Schnell (2022), although made trickier by physician heterogeneity in *quality* (V_j) as well as "prescription" *leniency* — in our case, of sick leave issuance.

Given the two-fold source of patient utility from physicians, we will speak in general of patients' 'reserve utility'. In a similar vain to sequential job search models, we envision patients as searching for a physician in discrete time, and when assigned a physician j, they have the option to go visit them, which is equivalent to recieving utility u_{ij} in perpetuity, or look for a physician again, "drawing" another physician (with replacement) next period. We assume a temporal discount factor $\beta \in (0, 1)$, and no-recall, meaning a patient may only visit the current physician assigned.

This modelation can be defined in terms of dynamic programming. For each patient i, the solution to this problem will be an optimal threshold $\bar{U_i}$, henceforth their "reserve utility", such that each patient is only willing to visit some physician j if the utility he would obtain from that visit, u_{ij}, is greater or equal to $\bar{U_i}$. We demonstrate in the Appendix A.4 that the optimal characteristic equation for $\bar{U}_i$ is the following:

$$\bar{U}_i = \frac{\beta}{1-\beta} \sum_{j=1}^{J} \left\{ \frac{\mathbb{1}[u_{ij} \geq \bar{U}_i] \cdot (u_{ij} - \bar{U}_i)}{\mathbb{1}[u_{ij} \geq \bar{U}_i]} \right\} \tag{8}$$

The components s_{ij} of each patient's strategy vector S_i are then defined in terms of patient i's 'reserve utility' $\bar{U}_i$:

$$s_{ij} = \frac{\mathbb{1}[u_{ij} \geq \bar{U}_i]}{\sum_{k=1}^{J} \mathbb{1}[u_{ik} \geq \bar{U}_i]} \tag{9}$$

Equation (A.4) makes the rather simplifying but reasonable assumption that patient i 'draws' physicians randomly and thus has equal probability to visit any of the physicians who give him utility above his threshold $\bar{U}_i$.

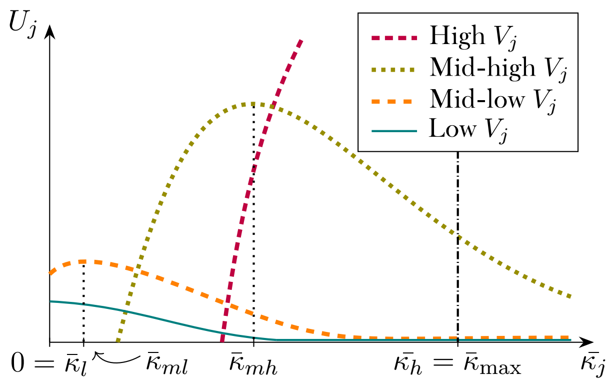


Figure 2: Physician utility U_j in the sequential model for different levels of V_j, as it varies across their choice of $\bar{\kappa_j}$

Figure 2 presents a the stylized utility functions of physicians of different V_j in the manner of Figure 1 in the previous section. The curves are more *well-behaved* (read: quasi-concavity) than in the "implicit" search model for the simulations which form the basis of this figure, and could perhaps analitically be proven to have a unique pure strategy equilibrium under certain conditions.

On thing of note is that more often than in the Logit model, for (relatively) low enough values of V_j, physician utility will be strictly 0 for high choices of $\bar{\kappa_j}$, and for high values of V_j, utility will be *negative* at low thresholds $\bar{\kappa_j}$. This effect is more pronounced the higher the β parameter is, which plays the same role as λ in the previous model: the higher it is, the more patient demand will be concentrated at high quality physicians. Patient's 'patience' β determines "how long" they're willing to look for the ideal physician, i.e. of highest V_j. As $\beta \to 1$, demand for the physician with the highest V_j tends to be I, and for anyone else to be 0.

As before, we formalize the notion of equilibrium in the sequential model.

Equilibrium in the "explicit" search model. *Given a physician-patient market specified by* $\{\{(\kappa_i, \gamma_i)\}_{i=1}^{I}, \{(V_j, R_j(\cdot), \tau_j)\}_{j=1}^{J}, P(\cdot), \beta\}$*, we define equilibrium as a set of patient thresholds* $\{\bar{U}_i\}_{i=1}^{I}$ *fulfilling (8), patient strategies* $\{S_i : s_{i1} \ldots s_{iJ}\}_{i=1}^{I}$ *as defined by (9) and physician thresholds* $\{\bar{\kappa_j}\}_{j=1}^{J}$ *satisfying problem (3) with their corresponding equilibrium (expected) patient demand and issued sick leaves* $\{(Q_j, X_j)\}_{j=1}^{J}$ *as described by (5) and (6), respectively.*

4 Computation

4.1 Algorithms for Model Computation

Given a set of physicians $\{(V_j, R_j(\cdot), \tau_j)\}_{j=1}^{J}$, punishment function $P(\cdot)$, distribution functions for patient parameters $F(\kappa)$ and $G(\gamma)$, and one search model parameter z, which is λ in the Logit model, β in the sequential model, we develop an algorithm to compute market equilibrium.

Equilibrium aggregates are computed on the basis of Monte-Carlo matrix calculus. Set J as the number of physicians,[5] I as the size of the sample drawn randomly from $F(\kappa)$ and $G(\gamma)$. For both models we define a class which can output a matrix S where each column is a patient's strategy vector S_i following that model, when given as input the arrayed set of physicians' quality and visit cost $\{(V_j, \tau_j)\}_{i=1}^{J}$, an arrayed set of patients $\{(\kappa_i, \gamma_i)\}_{i=1}^{I}$, the model parameter z and a given vector of physician strategies $\{\bar{\kappa_j}\}_{j=1}^{J}$.

We input as "patients" our I samples from $F(\kappa)$ and $G(\gamma)$, then a $J \times I$ matrix U is computed, where each component u_{ji} corresponds to the utility the sampled patient i would get from a visit to physician j.[6] This step is the same for both classes.

What differs between both models is the computation of the matrix S of patient strategis out of the utility matrix U:

- **Implicit search model (Logit):** First, an 'α-matrix' is calculated over matrix U, where each α_{ji} is $e^{\lambda u_{ji}}$ if $u_{ji} > 0$ and 0 if not. Then, for each patient i, that is, for each column, each component s_{ji} of the S matrix takes on the values $s_{ji} = \alpha_{ji} / \sum_{k=1}^{J} \alpha_{ki}$.
- **Explicit search model (Sequential):** Recall Equation (8) characterizing patient thresholds. We first compute the I-dimensional vector of the sampled patients' respective $\bar{U}_i$. Define U_i as the set $\{u_{ji}\}_{j=1}^{J}$ of utility patient i recieves from a visit to each physician. In matrix terms U_i would be the ith column of the $J \times I$ matrix U. The operation to compute each $\bar{U}_i$ is the following:

[5] Or, as we'll later interpret it, as the number of bins, where each bin j is a unique combination of $(V_j, \tau_j, R_j(\cdot))$.

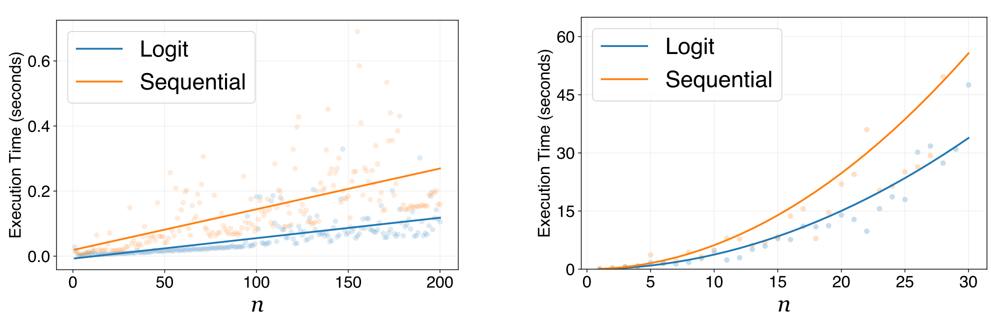
[6] As we have defined our matrices $J \times I$ for ease of visualization, we will refer to matrix components as x_{ji} in this section rather than x_{ij} as we do elsewhere in the paper.

$$\bar{U}_i \equiv \arg\min_{x \in U_i} \left\| x - \frac{\beta}{1-\beta} \sum_{j=1}^{J} \left\{ \frac{\mathbb{1}[u_{ji} \geq x] \cdot (u_{ji} - x)}{\mathbb{1}[u_{ji} \geq x]} \right\} \right\|$$

where the norm $\|\cdot\|$ is defined in $\mathbb{R}$ as simply the absolute value $|\cdot|$. This is to say, for each sampled patient i we evaluate x for each u_{ji} in U_i. In plain words, if patient i where to say: "I will only visit physicians which grant me at least as much utility as physician j", the optimal choice of j and its respective u_{ji} would be the minimal in U_i for the evaluation of the absolute value in the left-hand side of (4.1). This is computationally less intensive than seeking to compute the exact root of Equation (8), which would be a redundant exercise, because there's a discrete number of physicians above that mark, and selecting instead to use "the lowest u_{ji} above the root of (8)" as threshold instead of the root proper would result in the same vector of strategies S_i for each patient.[7]

Once having computed our I-dimensional vector of patients' $\bar{U}_i$, we evaluate column-wise the binary operator $\mathbb{1}\{u_{ji} \geq \bar{U}_i\}$ over matrix U, and the $J \times I$ matrix S of patient strategies takes on the values $s_{ji} = \mathbb{1}\{u_{ji} \geq \bar{U}_i\} / \sum_{k=1}^{J} \mathbb{1}\{u_{ki} \geq \bar{U}_i\}$.

The S matrix in both cases is computed in linear time, $\mathcal{O}(n)$ in big-O notation, where input size n is the number of physicians J. Both consist of a series of ufunc or vectorized operations, where the inclusion of one more physician introduces a new row with I elements on which operations are performed at each stage. The additional step of calculating the $\bar{U}_i$ vector makes the execution of the 'explicit' search model relatively slower (see Panel Figure 3a).



(a) Linear polynomial fit of execution time for S matrix in both models

(b) Quadratic polynomial fit of execution time for $\bar{\kappa}^*$ vector in both models

Figure 3: Execution time comparison between the *implicit* and *explicit* search models

The steps following the computation of the S matrix are the same for both models. The J-dimensional vector Q of each physician's expected demand Q_j is achieved

[7]Granted, this is not strictly true, as in our formulation a 'u_{ji}' may be selected as threshold which is actually below the root proper in $\mathbb{R}$, but closer to it than the nearest one above it. The effect of this "estimation noise" on our overall results is negligible.

through the row-wise summation of all patient's demand for j, i.e. performing $\sum_{i=1}^{I} s_{ji}$ for each j. The vector X of each physician's expected sick leaves issued X_j performs that same summation, conditional on the patients' κ_i being above or at physician j's chosen threshold $\bar{\kappa}_j$, that is, $\sum_{i=1}^{I} \mathbb{1}[\kappa_i \geq \bar{\kappa}_j]\, s_{ji}$. Both sums are normalized to fit the actual number of patients in the physician-patient market.

For a given vector $\bar{\kappa}$, computing patient strategies, physician aggregates and then physicians' utility—defined as $\{R_j(Q_j) - P(X_j)\}$—entails a sequence of calculations done in linear time. However, the calculation of the *equilibrium* vector $\bar{\kappa}^*$ is performed in quadratic time $\mathcal{O}(n^2)$, as it requires each of the aforementioned steps to be done some x number of times *per* physician, such that the inclusion of an additional physician increases the number of operations required *per* physician as well as the amount of physicians whose $\bar{\kappa}_j$ needs to be computed. See Panel Figure 3b.

The algorithm to find the equilibrium J-vector $\bar{\kappa}^*$ of physician strategies is as follows:

i Input an initial guess $\bar{\kappa}^0$ for physician thresholds.

ii For each physician, we fix the the value of $\bar{\kappa}_{-j}$, the threshold values of the other $J-1$ physicians, and compute equilibrium aggregates and physician j's utility for different choices of $\bar{\kappa}_j$ across a grid. In particular, $U_j(\bar{\kappa}_j, \bar{\kappa}_{-j})$ is computed for each decimal step between 0 and $\bar{\kappa}_{\max}$, the maximum threshold physicians may choose.

iii The choice of threshold k_j which rendered j the most utility in the previous step is selected, and a grid is set-up spanning the 19 centesimal values in $[k_j - 0.05, k_j + 0.05]$,[8] each value therein input as physician j's threshold $\bar{\kappa}_j$ to compute U_j again, and the value within the grid which maximizes utility is chosen as $\bar{\kappa}_j^1$.

iv Having performed the previous step for all J physicians, we input as a new guess the vector $\bar{\kappa}^1 = \{\bar{\kappa}_1^1, ..., \bar{\kappa}_J^1\}$ to run steps ii and iii again. This defines an equilibrium-searching loop $\bar{\kappa}^n = \Phi(\bar{\kappa}^{n-1})$, and the loop is concluded when a fixed point is found, that is, the vector $\bar{\kappa}^*$ such that $\bar{\kappa}^* = \Phi(\bar{\kappa}^*)$.

v Optionally, having found a two decimals fixed point $\bar{\kappa}^*$, there's an algorithm in place to find an x-decimal fixed point $\bar{\kappa}^{*x}$. It starts by running a modified version of step iii on the two-decimal solution, setting up a grid of the 19 *millesimal* values in $[\bar{\kappa}_j^* - 0.005, \bar{\kappa}_j^* + 0.005]$ for each physician j,[9] selecting that which renders highest utility for each, and inputting this new vector as guess to run this step again, and so on until the fixed point $\bar{\kappa}^{*3} = \Phi(\bar{\kappa}^{*3})$ is found.

  This goes on like this, calculating the t decimals solution out of the $t-1$ decimals solution by setting up $10^{-(t-1)}$-sized grids for each physician in each iteration, until the specified amount of decimals wanted from the solution vector $\bar{\kappa}^{*x}$ is reached.

[8] If k_j is 0 (or $\bar{\kappa}_{\max}$), only the 10 centesimal values above (below) and at k_j are computed.

[9] For $\bar{\kappa}_j^* = 0$ or $\bar{\kappa}_{\max}$ a similar logic to the footnote above follows.

On the one hand, having to perform several operations *per* physician is quite computationally intensive and runs the cost of quadratic execution time. On the other hand, altough each iteration takes some time, for the right parameters the algorithm is quite efficient in the number of iterations needed for convergence, usually taking between 2 and 5. The option to find fixed points to $n+1$ decimal places incurs progressively smaller execution time costs compared to n decimals, as the initial guess for $n+1$ decimals becomes increasingly accurate. See Figure 4.

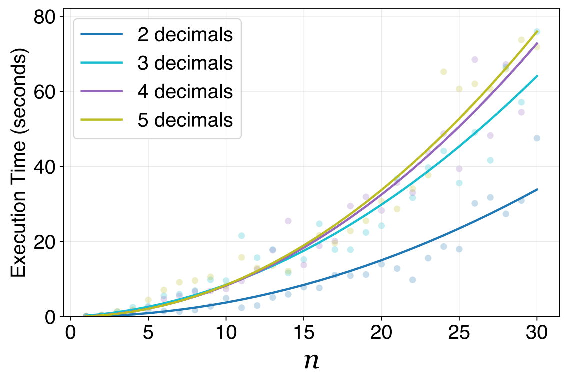


Figure 4: Quadratic polynomial fit of execution time of $\bar{\kappa_j}^*$
for different decimal approximations in Logit model

Once the equilibrium vector $\bar{\kappa_j}^*$ is calculated, it may be used to compute the equilibrium aggregates Q_j and X_j for all physicians, as well as physician utility.

4.2 Illustrative Examples

We use the following parametrization for illustration purposes:

$F(\kappa)$	=	$U[0,1]$	$R(x)$	=	$1.5x$
$G(\gamma)$	=	$U[0,1]$	$P(x)$	=	$\frac{1}{100}x^2$
V_j	$\sim$	$U[0,10]$	τ	=	1

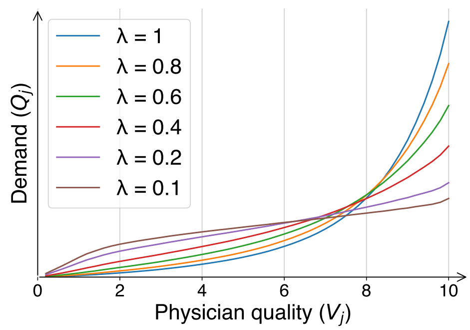
In particular, we consider $J = 50$ physicians whose V_j is evenly spaced in increments of 0.2 between 0.2 and 10. They share the same $R_j(\cdot)$ and τ_j.

For the Logit model, we present some results for different parametrizations of λ in Figure 5. Observe first Panel 5a, graphing patient demand Q_j when all physicians choose $\bar{\kappa_j} = 0$: when $\lambda = 0.1$, patient demand is more evenly distributed across physicians of different quality, and as it increases towards 1 it becomes more concentrated at physicians of higher quality.

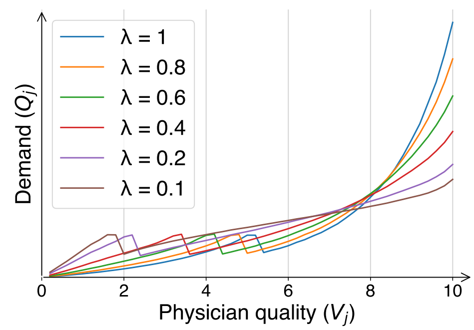
Panel 5b shows patient demand at the equilibrium $\bar{\kappa_j}*$ of each physician. For each level of λ, there is more patient demand at lower levels of V_j and then a drop-off, followed by a subsequent rise. The drop-off represents the transition from a $V_{mid-low}$

equilibrium to $V_{mid-high}$ (recall Figure 1), when physicians start opting for a higher $\bar{\kappa_j}$ in an inner solution. Before the drop-off, low V_j physicians opt for a threshold equal to 0, maximizing their possible patient demand, and then briefly slightly above 0. After a certain level of V_j, physicians find it better to be strict and decrease their patient demand slightly but scaling down their issued sick leaves by a larger measure, fulfilling the FOC condition (4).

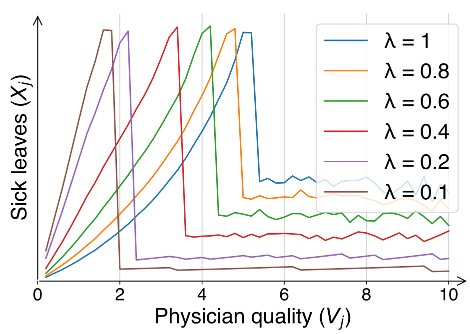
We can observe this behavior in Panel 5c. At each λ there's a significant drop-off in sick leaves at a certain V_j once physicians start preferring the stricter inner solution. Notice how after the drop, sick leaves remain flat, because physicians share $R_j(\cdot)$ and $P(\cdot)$ in our example, and in particular $R_j(\cdot) = r_j$, so as patient demand Q_j increases with V_j, all physicians must do in order to remain at the FOC physicians is increase their threshold $\bar{\kappa_j}$ to remain at the same level of X_j. We can see how $\bar{\kappa_j}^*$ increasies in V_j in Panel 5d.



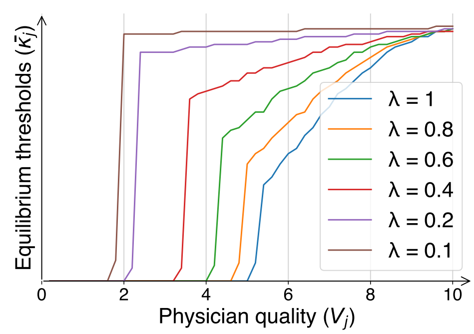
(a) Patient demand Q_j
by physician quality, $\bar{\kappa} = \vec{0}$



(b) Patient demand Q_j
by physician quality, equilibrium $\bar{\kappa}^*$



(c) Sick leaves issued X_j
by physician quality, equilibrium $\bar{\kappa}^*$



(d) Equilibrium $\bar{\kappa_j}^*$ by physician quality

Figure 5: Physician aggregates and strategies for different values of λ
in the *implicit* search model

Let's now look at the sequential model with the same parametrization, for different values of β (Figure 6). Something to not is that β is restricted to values within $(0, 1)$. As $\beta \to 1$, Q_j behaves similarly to when $\lambda \to \infty$ in the Logit model, which is why the concentration of patients at high V_j's with $\beta = 0.99$ is much more pronounced in

Panel 6a than it is with $\lambda = 1$ in the previous Figure.

The shape of the demand curve points to two disimilarities from the Logit model:

- For physicians of low enough V_j demand is strictly 0 rather than just low, as they don't afford to any patient utility over their threshold $\bar{U}_i$.
- At high values of V_j the demand curve is concave rather than convex like before. This is because the formulation of the 'explicit' search model doesn't continuously (and exponentially) reward higher values of V_j, it only registers whether V_j is enough to have u_{ij} over patients' reserve utility $\bar{U}_i$. Past a certain point, most patients' $\bar{U}_i$ (which can be seen in Panel 6e) lay below the u_{ij} a certain physician j affords, and incremental gains to even higher V_j are low, hence the concavity.

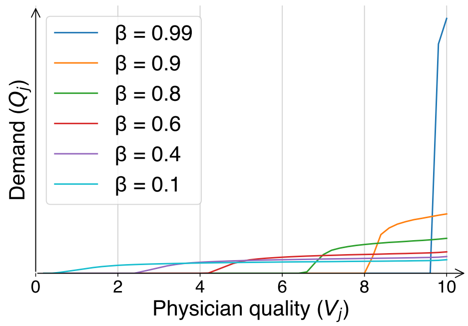
Panel 6b is not particullarly illuminating, but Panel 6c is, and can show the difference between 6b and 6a. Panel 6c displays sick leaves granted X_j as a function of physician quality V_j for different values of the model parameter β, at an equilibrium physician strategies vector $\bar{\kappa}^*$. Graphically, it exhibits two peaks in X_j, which grow more pronounced and further apart from each other and the origin as β increases.

The first peak in each curve is composed of physicians who manage to get positive clientele from leaving their threshold $\bar{\kappa}_j$ at 0. The drop-off from the top signals the transition to an inner solution, where the patient base of physician j decreases for the sake of more starkly decreasing sick leaves issued (X_j), fulfilling the FOC condition (4). The second peak has a smoother lead-up and descent from it, in which physicians make the trade-off of gaining patients through lower $\bar{\kappa}_j$ but requiring sick leave lenience. The peak itself represents the tipping point, after which physicians gain a significant patient base just by virtue of their high V_j and can afford to cut back on sick leave issuance (i.e. increase strictness) with little downside to their overall Q_j. Then there's a final peak seen specifically in $\beta = 0.99$, which results out of the high concentration of patients around the maximum V_j, such that patients whose $\kappa_i \geq \bar{\kappa}_{\max}$, previously more spread out across physicians, visit the highest V_j's exclusively.

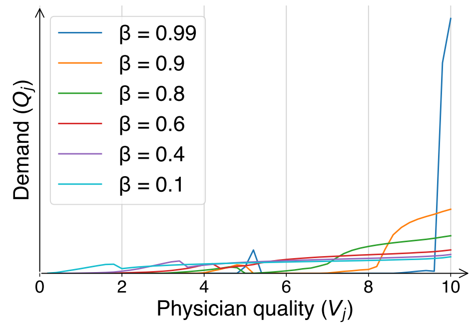
As β increases, so too the patients' reserve utility $\bar{U}_i$ as κ_i grows, as can be seen in Panel 6e. These higher values concentrate demand at higher values of V_j and also shorten the window in which patient demand may be gained through sick leave issuance (as opposed to just by virtue of a high V_j), which is why the peaks become "slimmer".

Despite these idiosyncracies, we can see in Panel 6d that physician thresholds $\bar{\kappa}_j$ grow with V_j in a very similar fashion to how they did in the previous model.

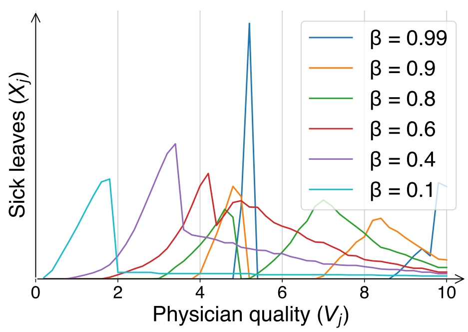
Some more figures from these simulations can be seen in the Appendix A.5. They show how physician utility is strictly increasing in V_j (meaning the drop-offs in patient demand after transitioning to inner solutions of $\bar{\kappa}_j$ are strategic and not a detriment to utility), and also highlight the difference in Q_j and X_j between the equilibrium values of $\bar{\kappa}^*$ and other physician strategy vectors.



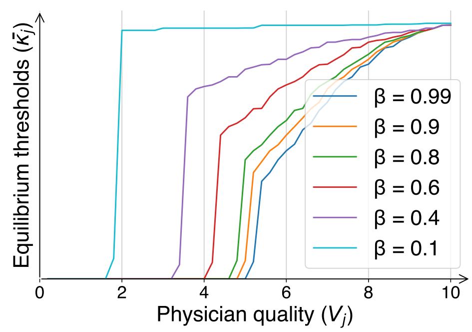
(a) Patient demand Q_j
by physician quality, $\bar{\kappa} = \vec{0}$



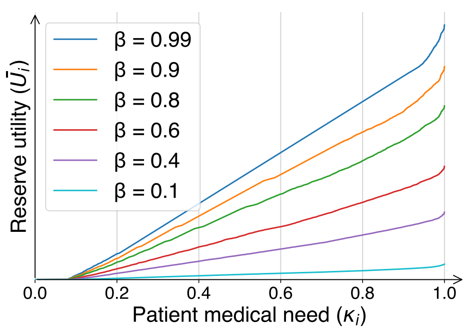
(b) Patient demand Q_j
by physician quality, equilibrium $\bar{\kappa}^*$



(c) Sick leaves issued X_j
by physician quality, equilibrium $\bar{\kappa}^*$



(d) Equilibrium $\bar{\kappa_j}^*$ by physician quality



(e) Patient reserve utility $\bar{U}_i$
by medical necessity (κ_i), equilibrium $\bar{\kappa}^*$

Figure 6: Physician aggregates and strategies for different values of β
in the *explicit* search model

5 Data

We now turn to the empirical section of this paper. To estimate our models, we will use administrative data provided by FONASA from all sick leaves issued from January 2018 to October 2022 in the public health system. We consider authorized sick leaves only. Our sample includes 48,611 different physicians and 1,445,696 different patients. The dataset contains limited information: each observations consists of a single issued sick leave including encrypted ID's for the physician and patient incolved (as well as the patient's employer), duration of sick leave and *cie10* classification of the condition, and some data on issuance and payment methods.

Table 1 presents descriptive statistics on sick leaves by physician and patients, i.e. on the empirical distributions of X_j and X_i. One first indication of the skewness of both distributions is the difference between mean and median: whereas the median physician issued 24 sick leaves over the sampled period, the *average* physician issued 88. The right tail is long, the 90th percentile physician issued 216 sick leaves, which quickly goes up to 1,007 at the 99th percentile, and then the most issued by a single physician is more than eight times that amount at 8,743.

The story is similar with regard to patients, as a considerable share of the sample appear only once and then the median is just *two* sick leaves received over the sampled period, goes up to 7 and 14 at the 90th and 99th percentiles, respectively, but then the single worker with the most sick leaves authorized is over five times the 99th percentile at 73.

Statistic	Physicians	Patients
Mean	88	3
S.D.	199	3
Min	1	1
Max	8,743	73
Percentiles		
10th	2	1
50th	24	2
90th	216	7
99th	1,007	14

Table 1: Total sick leaves issued *by* physicians *to* patients
Summary statistics over the sampled period

Long tails are not only present in both distributions, but interrelated. Oteíza (2023) argues that there is a meaningful degree of assortative matching in this sample, where physicians willing to grant many sick leaves are matched with—found by—patients with a high demand for them.

This effect is captured by our framework: patients with serious conditions prioritize higher "quality" physicians, who, in turn, tend to treat patients with greater medical need (κ_i) due to their endogenously determined higher strictness threshold ($\bar{\kappa_j}$). Conversely, patients with lower κ_i, but a strong desire for sick leave (γ_i), naturally gravitate toward more lenient physicians with lower strictness thresholds ($\bar{\kappa_j}$).

We don't put the $cie10$ classification to use to estimate the distribution of κ, partly because κ_i's interpretation as willingness-to-pay isn't one to one with strict medical severity, but mainly because the very subject of this paper is fraud, and in that regard the health status of patients reported by physicians are unreliable, especially for the most common conditions attested to by physicians (see Table 3 in the Appendix A.7). In fact, these routine conditions are in many cases hard to gage even for well-intentioned physicians. Carlsen et al. (2020) establish that physicians in these scenarios have no incentive to place doubt on the patients' self-reported symptoms. This is part of the reason we choose to fit a structural model rather than opting for more straightforward, data-oriented methods of estimation.[10]

6 Estimation

6.1 Data Moments

Having explored both our theoretical framework and the data we have on sick leaves, we turn to calibrating our "implicit" search model using that data. Our main blind spot is the fact that we have no information on overall patient visits by each physician (Q_j), only the sick leaves they were authorized to issue in the sampled period (X_j). We get the necessary variation to be able to identify and estimate our parameters from two sources: by assigning each of the 48,611 physicians to one of ten equally sized "quality" bins (or deciles) and by making a distinction between overall sick leaves granted and those of 30 days or more.

A qualitative consequence of both our proposed models is that physician strategies $\bar{\kappa_j}$ were growing on physician "quality" V_j, or alternatively put, "strictness" decreasing on V_j, meaning that as physicians expect less and less demand from their own reputation as medical professionals, they would adjust by acquiring a reputation as lenient sick leave issuers. This behavior, not an analytical given from our framework, holds for our proposed distributions for $F(\kappa)$ and $G(\gamma)$.

Being that we lack data on observable qualities from physicians, including individual salary and educational background, our approach to grouping them into quality bins was to make use of that observed relationship and its corollary: higher quality physicians will raise their threshold $\bar{\kappa_j}$ and thus see patients of higher κ_i on average, and in particular, a higher proportion of their patient base will be composed of patients satisfying $\kappa_i > \bar{\kappa}_{\max}$.

In this regard we can make use of the insensive margin of the sick leaves in our data,

[10] As the severity of the ailment increases, and the duration of sick leave accordingly, higher institutional and medical scrutiny is placed on diagnosis and issuance. We will discuss this in the next section.

namely that we know of how many days they were. We argue that the population equivalent of patients whose $\kappa_i > \bar{\kappa}_{\max}$ is patients who received sick leaves of 30 or more days. Though in a sense an arbitrary threshold, it so happens that starting at 30 days sick leaves are subject to higher institutional scrutiny and are automatically sent to the overseeing authority, subject to their authorization, requiring some medical proof—a test result—of the patient's ailment. At that point, the possibility of outright fraud can be safely discarded.

From a medical standpoint, a condition which would require a leave of absence of 30+ days is serious enough that no competent medical professional would wave off entirely the necessity for rest. This is to say, for a given authorized sick leave of 30+ days in our sample, we could reasonably picture that a "stricter" physician would have signed for less days off, say, just two weeks, but it is out of the question that, agreeing on diagnosis, what one physician considers as requiring over a month off work the other would not even consider as warranting any sick leave at all.

Figure 7 illustrates the empirical cumulative distribution of sick leave durations across all observations in our dataset.[11] Sick leaves lasting 30 days or more account for 20.05% of issuances during the sample period. In particular, sick leaves of exactly 30 days represent 15.64% of the total, a significant discrete jump in the curve which afterwards becomes almost flat. The highest duration of any single sick leave within our sample is 728 days.

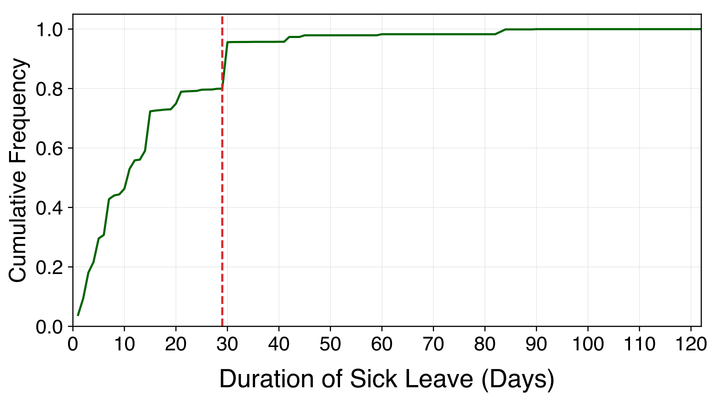


Figure 7: Empirical cumulative distribution of sick leave durations.
A red dashed line is placed at 29 days.

We first make use of this distinction to set up our physician quality bins. We design a *quality score* variable by physician based on the three following criteria:

i Average duration of sick leaves issued.

ii Percentage of sick leaves of 30+ days over total in sampled period.

iii Amount of sick leaves of 30+ days issued.

[11] Since this probability distribution assumes integer values only, the cumulative probability of a sick leave lasting 30 or more days is calculated as $1 - \text{CDF}(29)$.

The quality bin to which a physician is assigned corresponds to the decile rank of *quality score* to which they belong. See Appendix A.7 for a breakdown of the construction of the score variable and the physician bins.

This procedure does perhaps amount to an over-exhertion of our data sample, recognizing that by taking qualitative results from our model as a presumed basis we somewhat pre-fit our data to the model even before calibration, but it is by no means an econometrically unsanctioned manouver, as the population moments with which we calibrate our parameters contain, in part, different information to that with which we set up the bins themselves, so that identification is in fact possible.

It would indeed be best to assign a quality score based on a set of observables wholly separate from those which play the role of endogenous outcomes in our model, but our course of action is sufficient for our purpose, which isn't to *falsify* our framework—built upon observations and findings established in the relevant literature—but to showcase it and present an approach for its estimation.

Our second use of the distinction between overall sick leaves and the '30+ days' subsample is that we use both as data moments to calibrate our parameters. This means we have two moments *per* quality bin, 20 moments in total to fit our model, as we won't consider variation across the r_j / τ_j axes.

To define both moments formally within our framework, each bin corresponds to a representative physician j, where $J = 10$, such that we have two J-dimensional vectors, X and X^{30}, where the jth component of each corresponds respectively to the two following moments for representative physician j:

$$X_j = \int_{\bar{\kappa}_j}^{\infty} \int_0^{\infty} s_{ij}(\kappa, \gamma)\, dG(\gamma)\, dF(\kappa) \tag{6}$$

$$X_j^{30} = \int_{\bar{\kappa}_{\max}}^{\infty} \int_0^{\infty} s_{ij}(\kappa, \gamma)\, dG(\gamma)\, dF(\kappa) \tag{10}$$

The advantage of using this second defined moment, X^{30}, lies in its independence from the choice of strategy κ_j and its *near* independence from the distribution $G(\gamma)$. Any patient satisfying $\kappa_i > \bar{\kappa}_{\max}$ will, by definition, be granted sick leave by any physician. The implication of this on the probability s_{ij} that such a patient visits any physician j can be seen as follows:

$$s_{ij} = \frac{\alpha_{ij}}{\sum_{k=1}^{J} \alpha_{ik}} = \frac{e^{\lambda u_{ij}}}{\sum\limits_{k:\, u_{ik} \geq 0} e^{\lambda u_{ik}}} = \frac{e^{\lambda\{V_j \kappa_i + \gamma_i - \tau_j\}}}{\sum\limits_{k:\, u_{ik} \geq 0} e^{\lambda\{V_k \kappa_i + \gamma_i - \tau_k\}}}$$

$$= \frac{e^{\lambda\{V_j \kappa_i - \tau_j\}} \cancel{e^{\lambda \gamma_i}}}{\sum\limits_{k:\, u_{ik} \geq 0} e^{\lambda\{V_k \kappa_i - \tau_k\}} \cancel{e^{\lambda \gamma_i}}} = \frac{e^{\lambda\{V_j \kappa_i - \tau_j\}}}{\sum\limits_{k:\, u_{ik} \geq 0} e^{\lambda\{V_k \kappa_i - \tau_k\}}}$$

We see that γ_i can be factored out of the defining fraction. It still plays a role, however, in the condition of the summation in the denominator. Physicians afforded a strictly positive probability of visit by patient i must meet the $u_{ij} \geq 0$ condition implied by our *free disposal* patient rationality. This in turn requires of one such physician j to fulfill $V_j\kappa_i + \gamma_i \geq \tau_j$. So the scale of γ_i still has implications for X^{30} in this regard, only insofar as it may take a physician completely out of rotation for a patient's strategy vector S_i. We can see in the Appendix A.6 that the moment outcomes of our model fit are much less responsive to variations in the parameters for $G(\cdot)$ in the case of X^{30} as in overall X.

6.2 Parameters and Estimation Results

There are several elements in this model that require specification and then calibration. Starting with the two population distributions $F(\kappa)$ and $G(\gamma)$ describing patients, in keeping with their analogues in Schnell (2022), we choose an exponential distribution for κ with a λ_F scale parameter, such that as "medical necessity" ('pain severity' in Schnell) grows larger , it becomes increasingly rare — whereas for γ, the "taste for sick leave" (for prescription drugs in Schnell), we choose a normal distribution, parametrized by a mean μ and variance σ.

Both μ and σ we will fit through calibration, but for λ_F we simply normalize it to 2, in order to have the average κ_i in the population be 1/2, on aesthetic grounds. We normalize λ_F because it couldn't otherwise be correctly identified, as κ_i only ever appears in the patients' utility function accompanied by a multiplying V_j. Being the case that we will calibrate the values of V_1 through V_{10}, we need to fix a value for the distribution of κ, else identification will fail. Fortunately for us, and not coincidentally, what matters really is the shape of the distribution. Given the *scaling property* of the exponential distribution, if some value k^* is the true population parameter for λ_F, rather than 2, then the true values for the components of the physician quality vector $\vec{V}$ would follow $V_j^* = 2/k^*\ V_j$. But then the accompanying expect value of κ_i would itself be scaled by $k^*/2$, such that for these *true* values the level of each $V_j\kappa_i$ would remain as before. Given that only this single term $V_j\kappa_i$ is relevant to inform agent behavior, not each product individually, it makes no difference in the end our choice of λ_F.

As for physicians, as we mentioned, we will estimate the ten components of the $\vec{V}$ vector. With respect to their utility function, the form we give to the revenue and punishment functions, $R_j(\cdot)$ and $P(\cdot)$, is a linear function r_jQ_j and a strictly convex quadratic function $p/2 \cdot (X_j)^2$. For convenience, we will disregard the j subscript in the revenue function like we have done all along for $P(\cdot)$. The parameter p will be subject to GMM estimation, whereas r—and for that matter the cost of visit τ_j, which shall also be the same τ for all physicians—we obtain from the data directly, doing a weighted averaged of listed prices for doctor visits by FONASA matched with the 10 most common conditions granted sick leave. The resulting values for r and τ are 21.182 and 8.944, respectively. See Appendix A.7 for more detail on the calculation.

The choice of a linear function for $R_j(\cdot)$ is natural and precedented, the parameter r_j (r) has a direct interpretation as the revenue by visit for physician j. The choice of $P(\cdot)$ is a little more subtle. At this point we would like to point out, it is not selfevident that our *punishment* function should take total sick leaves granted by physician (X_j) as its argument. The reasoning behind it is the fact that the patients' κ_i would be unobservable to the institutional overseer, as is the case in the real-life setting, where reports on the patients' conditions come in most cases from the very same agent that issues sick leaves, the physician, making fraud and misdiagnosis be undetectable in these instances.

As a crutch, institutional discipline relies on what variables *are* observable, namely, the amount of sick leaves issued within a given period by physician. As chronicled by Oteíza (2023), in 2021 the regulatory authority COMPIN sanctioned a subsample of physicians who were eggregious outliers in their volume of sick leaves issued and were unable to justify it. The 'expected punishment' which faces some physician j could be formalized as:

$$\text{fine\_\$} \times \Pr[\text{fined} \mid X_j]$$

The probability of being find would be near 0 until X_j gets to be above around the 90th percentile, at which point starts steadily increasing. If we allowed the magnitude of the fine to increase in proportion to the level of X_j, the 'punishment function' would be more convex still. We choose $p/2 \cdot (X_j)^2$ as a reasonable polynomial approximation of the function implied by our description. Convexity would not be nearly as steep as before, but it has the advantage of parsimony. The parameter p then takes on the interpretation of both magnitude and probability of fine, anything that impacts the expected punishment physicians envisage for any given level of X_j.

Parameter	Value	Source
Physician quality bins		
$V_1 \ldots V_{10}$	$\vec{V}_{1\times 10}$	GMM
Punishment function $P(\cdot)$		
p	1.953	GMM
Medical need distribution $F(\kappa)$		
λ_F	2	Normalization
Taste distribution $G(\gamma)$		
μ	73.997	GMM
σ	18.492	GMM
Logit model parameter		
λ_s	0.842	GMM
Maximum threshold level		
$\bar{\kappa}_{\max}$	0.763	GMM
Revenue/cost of visit		
r	21.182	Data
τ	8.944	Data

Table 2: Model parameters

Finally, we must also estimate λ_s, the parameter specific for the Logit model, characterizing the "effectiveness" of search, and $\kappa_{\max}$, the upper bound for physician strategies, regulating the aggregate level of X^{30} in the market. In its estimated level, 21.75% of patients will fall above this threshold, near the 20.05% of sick leaves of 30+ days in our sample.

All the parameters in our model and their function within it are summarized in Table 2, alongside they value they take on and its source. Two parameters are directly retreived from the data (r and τ), one we select/normalize ourselves (λ_F), the rest are subject to GMM estimation. That makes 15 parameters to be evaluated out of 20 data moments.

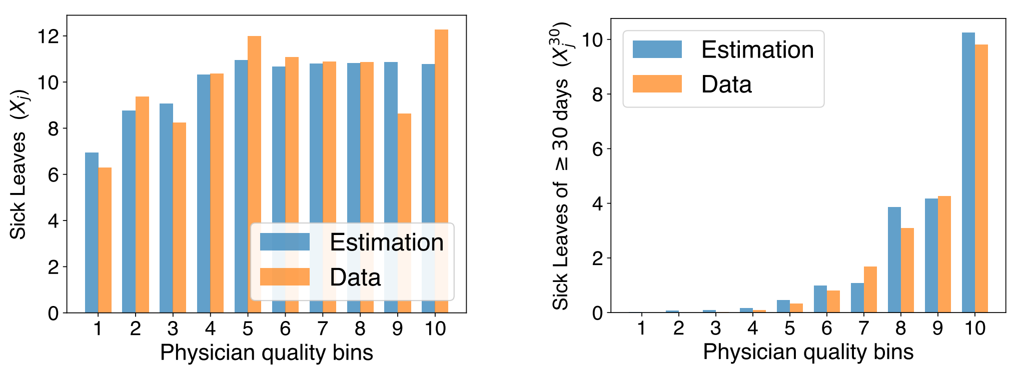
Define $\theta = \{V_1, ..., V_{10}, p, \mu, \sigma, \lambda_s, \bar{\kappa}_{\max}\}$ and $m = \{X_1, ..., X_{10}, X_1^{30}, ..., X_{10}^{30}\}$. The generalized method of moments (GMM) defines the optimal set of parameters as:

$$\hat{\theta}_{GMM} = \min_{\theta} \parallel m(x) - m(x|\theta) \parallel \tag{11}$$

where $m(x)$ are is the moment vector from real data x, and $m(x|\theta)$ the vector of moments from the model that correspond to the real-world moment vector $m(x)$. The optimal $\hat{\theta}_{GMM}$ minimizes the distance between these two vectors.

Minimization is achieved through a series of global and local optimization algorithms, using the '*dual_annealing*' and '*Powell*' modules of Scipy optimization.

The resulting set $\hat{\theta}_{GMM}$ of parameters can be seen in Table 2. The comparison between the moments in the data $m(x)$ and the moments from our GMM fitted model $m(x|\hat{\theta}_{GMM})$ is displayed in two bar plots in Figure 8.



(a) X_j (b) X_j^{30}

Figure 8: GMM results
Moments by quality bin, Data vs. Model fit (estimation)

The first four representative physicians take a corner, $\kappa_j = 0$ solution. The rest are at the inner solution, satisfying the FOC equation (4), choosing an increasingly large threshold in relation to their higher V_j. Patients satisfying $\kappa_i > \bar{\kappa}_{\max}$, which place high value on physician quality, are concentrated at the highest bins, especially number 10, and they would be even more if were the search parameter λ_s were higher.

Even considering the fact that our set-up of physician quality bins amounts to a sort of "pre-fitting" of the data to our model, the fitting itself is reasonably succesful, every single of the twenty moments in the model approaching its data analogue to a high degree, in some cases nearing equality.

6.3 Identification

Though fulfilling the minimal condition that we have more equations (20) than moments (15), we need to further ensure the correct identification of each individual parameter. This is straightforward to argue for in the case of the $\vec{V}$ vector, each of the ten V_j parameters directly raises demand for representative physician j, increasing the value of both moments for the respective quality bin. Identification of the other five moments comes from the interplay between the ten X moments and the other ten X^{30} moments.

To get a visual understanding, the Figure 13 in the Appendix A.6 plots the change in the distribution of X_j and X_j^{30} across quality bins for variations in the five remaining parameters. We briefly explain for each parameter the variation in the data which allows for its identification:

λ_s This parameter determines the "quality" of search in the Logit model. As it increases towards ∞, demand concentrates exclusively at V_{10}. As it tends to 0, demand is uniformly distributed across the ten bins. This sensitivity is more pronounced for the X^{30} moments, sick leaves for patients above $\kappa_{\max}$, for the reasons mentioned in subsection 6.1: this moment is almost independent from γ_i, which conversely makes it more sensitive to V_j, which is in turn scaled by λ_s. This parameter serves to explain, for a given $\vec{V}$, the relative steepness of the distribution of X^{30} with respect to X.

$\bar{\kappa}_{\max}$ This parameter directly controls the scale of overall X^{30} with respect to X. The lower it is, the more aggregate X^{30} approaches aggregate X. It should be such that the X^{30}/X ratio resembles the ratio of sick leaves of 30+ days over total sick leaves issued in the data, and indeed it does (21.75% vs. 20.05%).

p The *punishment* function parameter p determines the level of X_j at which physicians who choose an inner solution $\bar{\kappa_j}$ stay, and the first physician bin which plays an inner solution. Raising or increasing p takes the X_j levels of the six representative physicians of highest quality away from the horizontal they currently occupy, regardless of the magnitude of V_j they hold.

μ, σ For these two parameters we run into complications. Jointly, they *can* be identified, as they account for variations in the location of the first inner solution bin beyond what the p parameter can achieve, but they can't be readily separated. The effect of a decrease in μ is equivalent to that of an increase in σ, furthermore, the distribution is insensitive to an increase in μ as it is to decreasing σ. In this regard we can say that μ and σ jointly describe a feasible plane where the distribution we achieved is possible for values in both parameters, but they fail at reducing this 2-dimensional plane to a singleton, which would be the exact identification of both.

This last result is unfortunate, it implies we can only *partially* identify the distribution $G(\gamma)$, placing a question mark over our whole enterprise. We nonetheless carry on, but if this issue is to be alleviated, more data moments are to be made available for fitting. It's possible that the mean and variance of $G(\gamma)$ would present heterogenous effects for Q_j as they do on X_j, allowing distinction between the two. In that case, data on physicians' clientele would be crucial. Perhaps physician variation across r_j bins would be another solution, rather than our choice of a single r for all. We will touch on this at the end of this paper.

For now, we go on with our estimated parameters and run simulations to showcase our framework.

7 Counterfactual and Discussion

With our fitted model at hand, we are in a position to run different scenarios. In particular, what might be the ideal choice of p for the punishment function by a policymaker?

Ideally, a policy maker would be able to stamp out fraud directly, which in our model would imply that they would be able to have direct influence over the physicians' choice of $\bar{\kappa_j}$, namely to place a specified lower bound on it, some level $\underline{\kappa}$ under which patients are under no circumstance to receive sick leave. That is, in a way, the status quo, except for the fact that enforcement is unfeasible. The "overdetermination" of roles which Markussen and Røed (2017) speak about not only has the consequence that physician serve opposite financial interests (their own and the taxpayers'), but also that, in this occasion, they're better able to cover their tracks. The policymaker, as principal, would like to establish surveillance over their agent, the physician, on whether they've granted unwarranted sick leaves. Only that their only source of information is the agent herself. The agent in charge of diagnosis is the same as then issues sick leave, as such they would have every incentive to fit one to the other, whether legitimately or illegitimately. We mentioned higher scrutiny *does* take place on high duration sick leaves, automatically sent up to the overseer for approval, requiring medical. This isn't the case for the routine conditions which make up the majority of observations in our sample, nor would it be practical that it were so.

This isn't necessarily a question of deliberate fraud either. Carlsen et al. (2020) establish that under all but the most extreme circumstances (read: set of parameters), physicians have no incentive to be distrustful of their patients' self-reported symptoms, whether in fact suspicious of foul play or not. In that regard, our policy of choice would seek to realign the incentives of the physician to that of the policymaker, rather than merely being a question of catching and punishing willing offenders, such that physicians self-police into subjecting diagnosis to some higher level of scrutiny on their part.

Consider for a moment the policymaker *were* able to place restrictions on $\bar{\kappa_j}$. Under our current parameters, if physician strategies, i.e. strictness, were imposed at a fourth of the value of the upper bound, that is, at $\underline{\kappa} = 1/4\,\bar{\kappa}_{\max}$, total issuance would go down by 32%. In particular, the lowest 32% of patients previously receiving sick leave, ordered by their 'medical necessity' κ_i, would be cut off.

How might the policymaker effect such a change through feasible policy means? The careful reader might argue that, in fact, we have already defined the punishment function $P(\cdot)$ as being the purview of the institutional overseer. As misdiagnosis itself isn't observed, punishment takes the form of a probability and magnitude of fine as a function of sick leaves issued over a given period (X_j), where we proposed $P(X_j) = p/2 \cdot (X_j)^2$ as a polynomial approximation. Under this definition, the policymaker would have control over the parameter p.

Figure 9 illustrates the *aggregate* amount of sick leaves issued in the physician-patient market as a function of different relative levels of p. The dashed horizontal line is the level of aggregate sick leaves that would be achieved by direct control over physicians' $\bar{\kappa_j}$, placing them at $\underline{\kappa} = 1/4\,\bar{\kappa}_{\max}$.

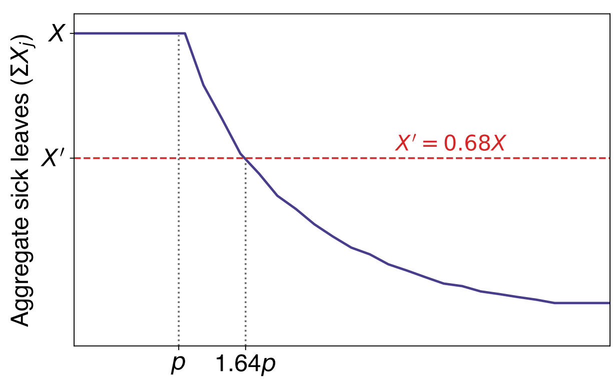


Figure 9: Aggregate sick leaves in our fitted model for different values of p

We don't directly address the question of welfare in this paper. The optimal choice of p, itself informed by the optimal choice of $\underline{\kappa}$, should derive from some model which can numerically formalize its different repercussions. Within our framework itself is present the downside: raising p presses physicians into higher strictness, reducing overall welfare, as physicians must cut back their services, facing higher expected penalties for a reduced revenue stream — and patients are in many cases induced to visit lower quality physicians, which decreases their utility with respect to the previous status quo, and in some cases patients are driven out of the market entirely.

The flipside, the reason why a policimaker would *want* to reduce the issuance of sick leaves, is not present in any shape or form in our framework. These negative consequences are one of the main subjects of the economic literature on sick leaves. Two motives stand out: sick leave fraud puts financial strain on either the taxpayer or employer, whoever covers the worker's wages; the work absence facilitated by easy access to sick leave has an impact on aggregate productivity. Such effects would need to be included in an estimation of the overall welfare impact of cracking down on sick leave issuance, but they fall well outside of the scope of our analysis.

Final comments

We have presented a framework which seeks to include and explain several behaviors in the sick leave market in a parsimonious way: assortative matching of high issuance physicians with high "usage" patients, leniency as a response to increasing competition among physicians, heterogenous leniency, imperfect matching. All of these are constructed to be endogenous outcomes from rational choices by both types of agents, physicians and patients, as well as certain modelling decisions on our part: heterogenous physicians, fuzzy search, dual source patient utility. In this respect, theoretically, our framework and both its computed specifications give reasonable insight into this market.

As for its use as a guide to policy beyond providing conceptual understanding, that is, correctly predicting the physicians' response to a raise in the probability and/or magnitude of the fine they face for excessive sick leaves granted, its reliability would depend on the plausibility of our chosen functional form for the different distributions and functions and the quality of the data used and performed estimation to calibrate their parameters. It's our view that we achieve the former, but perhaps not the latter. The chosen functional forms are spare in requirements and follow precedents set in the literature. The data we have available, however, is rather thin in comparison to relevant papers on this subject, which affects the quality of our estimation.

As we discussed above, our GMM exercise would benefit greatly from additional sources of information, both in terms of additional variation across physicians, as well as having more parameters be retrieved elsewhere and be set in place by the time we perform estimation, such that we require fewer degrees of freedom. Chief among the possible revisions and extensions to this program: information on the client base of physicians, i.e. a source of Q_j, and an alternative, separate method to gauge physician quality, perhaps relating observable qualities like education, tenure, place of work and so on to disagregated data on salaries and visit revenue. Such disaggregated data on its own could prove useful to find more horizontal variation across physicians, setting up bins of r_j and/or τ_j for their own sake. Ensuring proper identification is foremost among the next steps, but would also require a considerable amount of leg work. *You can only do so much in one semester.*

References

Autor, D., Duggan, M. and Gruber, J. (2014). Moral hazard and claims deterrence in private disability insurance, *American Economic Journal: Applied Economics* **6**(4): 110–141.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision, *Journal of Economic Behavior & Organization* **131**: 17–23. Experimental and Behavioral Economics of Healthcare.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: Experimental evidence, *Health Economics* **26**(2): 243–262.

Carlsen, B., Lind, J. T. and Nyborg, K. (2020). Why physicians are lousy gatekeepers: Sicklisting decisions when patients have private information on symptoms, *Health Economics* **29**(7): 778–789.

Choné, P. and Ma, C. (2011). Optimal health care contract under physician agency, *Annals of Economics and Statistics* **1**(101/102): 229–256.

Chunyu, L., Volpin, P. F. and Zhu, X. (2024). Do paid sick leave mandates increase productivity?, *Available at SSRN 4096707* .

Clemens, J. and Gottlieb, J. D. (2014). Do physicians' financial incentives affect medical treatment and patient health?, *American Economic Review* **104**(4): 1320–1349.

Crea, G. (2019). Physician altruism and moral hazard: (no) evidence from finnish national prescriptions data, *Journal of Health Economics* **65**: 153–169.

Currie, J., Li, A. and Schnell, M. (2023). The effects of competition on physician prescribing, *Working Paper 30889*, National Bureau of Economic Research.

FONASA (2024). Arancel de prestaciones de salud año 2024. modalidad libre elección, Available at `https://www.fonasa.cl/sites/fonasa/adjuntos/1LibroArancelMLE2024`.

Gaynor, M., Mehta, N. and Richards-Shubik, S. (2023). Optimal contracting with altruistic agents: Medicare payments for dialysis drugs, *American Economic Review* **113**(6): 1530–1571.

Godager, G. and Wiesen, D. (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism, *Journal of Health Economics* **32**(6): 1105–1116.

Hennig-Schmidt, H., Selten, R. and Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation, *Journal of Health Economics* **30**(4): 637–646.

Hennig-Schmidt, H. and Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students, *Social Science & Medicine* **108**: 156–165.

Henrekson, M. and Persson, M. (2004). The effects on sick leave of changes in the sickness insurance system, *Journal of Labor Economics* **22**(1): 87–113.

Instituto de Salud Pública, U. A. B. (2024). *Encuesta Nacional de Salud 2024*, Ipsos.

Iversen, T. (2004). The effects of a patient shortage on general practitioners' future income and list of patients, *Journal of Health Economics* **23**(4): 673–694. Contains contributions from the Grossman Symposium.

Iversen, T. and Lurås, H. (2000). Economic motives and professional norms: the case of general medical practice, *Journal of Economic Behavior & Organization* **43**(4): 447–470.

Johansson, P. and Palme, M. (2005). Moral hazard and sickness insurance, *Journal of Public Economics* **89**(9): 1879–1890.

Kesternich, I., Schumacher, H. and Winter, J. (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus?, *Journal of Public Economics* **131**: 1–11.

Ljungqvist, L. and Sargent, T. J. (2012). *Search, Matching, and Unemployment*, The MIT Press, pp. 159–224.

Markussen, S., Mykletun, A. and Røed, K. (2012). The case for presenteeism — evidence from norway's sickness insurance program, *Journal of Public Economics* **96**(11): 959–972. Fiscal Federalism.

Markussen, S. and Røed, K. (2017). The market for paid sick leave, *Journal of Health Economics* **55**: 244–261.

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior, *Technical report*, UC Berkeley: Institute of Urban and Regional Development.
URL: *https://escholarship.org/uc/item/61s3q2xr*

Oteíza, B. (2023). *Playing whac-a-mole: Effects of intervening the supply side of fraudulent sick leaves market in chile*, Master's thesis, Pontificia Universidad Católica de Chile.

Paola, M. D., Scoppa, V. and Pupo, V. (2014). Absenteeism in the italian public sector: The effects of changes in sick leave policy, *Journal of Labor Economics* **32**(2): 337–360.

Robertson, C., Rose, S. and Kesselheim, A. S. (2012). Effect of financial relationships on the behaviors of health care professionals: A review of the evidence, *Journal of Law, Medicine & Ethics* **40**(3): 452–466.

Schnell, M. (2022). Physician behavior in the presence of a secondary market: The case of prescription opioids, *Princeton University Department of Economics Working Paper* **5**: 383–410.

Stearns, J. and White, C. (2018). Can paid sick leave mandates reduce leave-taking?, *Labour Economics* **51**: 227–246.

Appendix

A.1 The Cause of the "Strategic Effect"

The main difference between our equilibrium and that of Schnell (2022) is the presence of a "strategic effect", wherein physician j takes into account the behavior of other physicians in the selection of her own $\bar{\kappa}_j$. Of the different modifications we made to Schnell's framework, we argue that it is the absence of *additive separability* across patients in physician utility $U_j(\cdot)$ which accounts for this.

In our context of unbounded maximization by physician j, the absence of additive separability implies she can't consider each patient individually when it comes to whether she's willing to allow (or induce) their visit, such a decision no longer being independent from other patients' visits in its impact to the physician's utility: the marginal utility of an additional patient is dependent on the aggregate of clients up to that point, both in terms of the visit itself as well as in the number of sick leaves granted up to that point.

Let's illustrate this point. Consider for a moment a finite number of patients $1, ..., k$, where each patient is inputed directly as an argument in our physician j's utility function $U_j(\cdot)$, like so: $U_j(1, ..., k)$. If $U_j(\cdot)$ has the property of additive separability, this means it may be reformulated as the addition of individual functions for each patient:

$$U_j(1, ..., k) = v_{j1}(1) + ... + v_{jk}(k) = \sum_{i=1}^{k} v_{ji}(i)$$

Unconstrained optimization in this context implies she's willing to see any patient whose $v_{ji}(i)$ is non-negative, such that her optimal level of utility is:

$$U_j^*(1, ..., k) = \sum_{i:\, v_{ji}(i) \geq 0} v_{ji}(i)$$

For our physician-patient model, where physician utility is increasing in κ_i, such patient selection is achieved by the physician through the choice of κ_j^*, excluding all patients i whose $\kappa_i < \kappa_j^*$. Supposing our patients are well-ordered in κ_i, the selection of such a κ_j^* would be one where the marginal consumer i affords a non-negative $v_{ji}(i)$, and the inframarginal consumer $i+1$ satisfies $v_{j,i+1}(i+1) < 0$. Ignoring for a moment that patients themselves have a *choice* of visiting—depending upon a second dimension γ —, we would then have:

$$U_j^*(1, ..., k) = \sum_{i:\, \kappa_i \geq \kappa_j^*} v_{ji}(i)$$

If instead of a discrete set we consider a mass of consumers $\mathcal{I}$ characterized by their level of κ_i, and we make the simplifying assumption that v_{ji} takes on the same form v_j for every $i \in \mathcal{I}$, our function $U_j(\cdot)$ could be expressed as:

$$U_j^*(\mathcal{I}) = \int_{\kappa_j^*}^{\infty} v_j(\kappa)\, dF(\kappa)$$

$U_j^*(\mathcal{I})$ represents the optimal value of $U_j(\mathcal{I})$, where physician j only sees patients who provide her with non-negative marginal utility, i.e. such that $\kappa_i \geq \kappa_j^*$. We can find this optimal value κ_j^* by looking at *threshold* equilibria, where physician utility is also dependent on the threshold $\bar{\kappa_j}$ over which she's willing to see patients:

$$U_j(\mathcal{I}, \bar{\kappa_j}) = \int_{\bar{\kappa_j}}^{\infty} v_j(\kappa)\, dF(\kappa)$$

The optimal value of the threshold $\bar{\kappa_j}$ corresponds to that of κ_j^*. Assuming $U_j(\cdot)$ is twicely differentiable and concave in $\bar{\kappa_j}$, this solution may be arrived at through the FOC:

$$\frac{\partial U_j(\mathcal{I})}{\partial \bar{\kappa_j}} \equiv -v_j(\bar{\kappa_j}) f(\bar{\kappa_j}) = 0$$

Solving for this would yield the optimal $\bar{\kappa_j} = \kappa_j^*$.

Schnell (2022) is an example of just such a treatment, which specifies the physician's additively separable utility by patient in the following manner:

$$v_{ji}(\kappa_i) \equiv R_j + \beta_j h(\kappa_i)$$

where R_j is a parameter which stands for revenue by visit, and $\beta_j h(\kappa_i)$ represents the physician's 'altruistic' utility over the health impact of a prescription drug to a patient with 'pain level' κ_i.

The optimal threshold κ_j^* is then obtained out of the maximization:[12]

$$\max_{\bar{\kappa_j}} \int_{\bar{\kappa_j}}^{\infty} R_j + \beta_j h(\kappa)\, dF(\kappa)$$

Resulting in the following FOC:

$$R_j = -\beta_j h'(\bar{\kappa_j})$$

which, as is immediately apparent, doesn't depend upon the behavior of other physicians, more specifically, on *their* choice of $\bar{\kappa_j}$. This equation merely establishes that, at the threshold, marginal benefit by patient—revenue R_j—must equal marginal "cost", the altruistic "cost" $\beta_j h(\bar{\kappa_j})$. A way to interpret this is that physicians will be willing to see any patient which render them positive marginal utility, unconcerned with their total market share, and thus, unconcerned with what other physicians may do to take away clientele.

A treatment like Schnell's is rendered inviable by our choice of utility function. Her parameter of revenue would in our model imply the linearity of our revenue *function* $R_j(\cdot)$ and our $P(\cdot)$ function over *aggregate* licenses granted. Strict convexity of $P(\cdot)$ would forestall its formulation as Schnell-like $\beta_j h'(\bar{\kappa_j})$ terms for each patient,

[12]Once again, ignoring patient choice and γ,. We're discussing a *simplified* version of the patient search framework for illustrative purposes.

because the impact on doctor j's utility in granting patient i a license would no longer independent from the granting of licenses of other patients. Likewise, strict concavity of $R_j(\cdot)$ belies a simple "r" parameter such that $R'_j(Q_j) = rQ_j$.

More formally, when $U_j(1, ..., k)$ *isn't* additively separable across patients, the value κ_j^* such that if $\kappa_i \geq \kappa_j^*$ patient i provides positive marginal utility, isn't independent of current clientele, because what before was a properly defined object, marginal utility by patient i, $v_j(\kappa_i)$, can no longer be so identified. The marginal utility i provides to j as the kth client (assuming some order over clients) is not necessarily the same he'd provide as the $k+1$th client, and so, as the kth client he could provide 0 utility, impliying $\kappa_i = \kappa_j^*$, whereas as the $k+1$th he could be inframarginal, such that $\kappa_i < \kappa_j^*$.

κ_j^* is not longer independent of clientele mass $\mathcal{I}$ as before, but a function of it, $\kappa_j^*(\mathcal{I})$. More specifically, in the models we consider it will depend on the *cardinality* of current clientele, $|\mathcal{I}|$, such that our physician's choice of marginal consumer will depend on her aggregate level of patient demand, where before it didn't. This effect is introduced through the *strict non-linearity* of our physician utility function $U_j(\cdot)$, either through the strict concavity of $R_j(\cdot)$ over expected patient demand Q_j, or the strict convexity of $P_j(\cdot)$ over total expected sick leaves granted.

When either of those is the case, $aggregate$ levels enter into the equation and form part of the optimality condition. Our general FOC reflects this:

$$R'_j(Q_j)\frac{\partial Q_j}{\partial \bar{\kappa_j}} = P'_j(X_j)\frac{\partial X_j}{\partial \bar{\kappa_j}} \tag{12}$$

where either or both of $R'_j(\cdot)$ and $P'_j(\cdot)$ are a non-constant function over aggregates.

When such aggregates, either Q_j or X_j, come into play, physicians come to consider their *market share*, which doesn't depend exclusively on their own choice of $\bar{\kappa_j}$. Each patient's strategy S_i is constructed taking into account the whole of $\{(V_j, \bar{\kappa_j})\}_{i=1}^J$.

Imagine equation (12) holds for some doctor j, and some doctor $l \neq j$ decides to lower her $\bar{\kappa_l}$, enough that it changes the value of s_{ij} for some mass of clients. The value of Q_j would then change, and therefore that of $R'_j(Q_j)$. If $P'_j(X_j)$ didn't vary by the same amount, (12) would no longer be an equality, leading j to modify her choice of $\bar{\kappa_j}$ to make equality hold.

This intelligible line of reasoning links the presence of a "strategic effect" to the non-additive separability of $U_j(\cdot)$: doctor j takes into account other physicians' strategy in her own choice of $\bar{\kappa_j}$ because of the present of *aggregate* amounts of clientele in her optimality conditions, which is so because utility isn't additively separable across clients.

A.2 Schnell (2022) with Logit Choice

To prove our point that it is additive separability which accounts for a possible strategic effect, we reformulate Schnell (2022) in the manner of a McFadden Logit — with some quirks.

The way in which Schnell devises her physician's utility formula is implicitly Bernoulli-like:

$$\int_{\kappa}\int_{\gamma} u(\kappa,\gamma)\cdot p(\kappa,\gamma)\, dG(\gamma)dF(\kappa)$$

where $u(\kappa,\gamma)$ is the utility function of the physician from patients characterized by a given (κ,γ) tuple, and $p(\kappa,\gamma)$ stands for the density of clients atomized in this tuple.

We define $u(\kappa,\gamma)$ as Schnell did: $\beta_j h(\kappa)+R_j$. The accompanying $p(\kappa,\gamma)$ fits the role that our own s_{ij} we have been playing thus far, which implicitly depends on the (κ_i,γ_i) tuple which characterizes patient i, and gives a measure of the change that they'll visit physician j. Physician utility can thus be rendered as:

$$\int_{\kappa}\int_{\gamma} [\beta_j h(\kappa)+R_j]\cdot s_{ij}(\kappa,\gamma)\, dG(\gamma)dF(\kappa)$$

In our modeling section, our "implicit" search model was a left-censored Logit choice model, which gave null probability of assignment to physicians from which patient i expected non-positive utility. It was defined as:

$$s_{ij}=\frac{\alpha_{ij}}{\sum_{k=1}^{J}\alpha_{ik}},\quad \text{where}\quad \alpha_{ij}=\begin{cases} e^{\lambda u_{ij}}, & \text{if}\quad u_{ij}>0\\ 0, & \text{if}\quad u_{ij}=0\end{cases}$$

Schnell's opioid-focused physician model provides an additional simplification: patients which aren't given a drug prescription, i.e. those such that $\kappa_i<\bar{\kappa_j}$, don't garner positive utility from a visit. As such, in the original model as in this reformulation, patients "below" $\bar{\kappa_j}$ have a null probability of visit, making the lower bound of integration for κ to be $\bar{\kappa_j}$.

As for γ, the lower bound of the inner integral remains the same as in Schnell in the absence of a secondary market, a value of γ_i such that $u_{ij}=0$, i.e. $h(\kappa_i)+\gamma_i-\tau^d-\tau^o=0$.[13]

The double-integral of physician utility to be maximized is then as follows:

$$\max_{\bar{\kappa_j}}\int_{\bar{\kappa_j}}^{\infty}\int_{\tau^d-\tau^o-h(\kappa)}^{\infty}[\beta_j h(\kappa)+R_j]\cdot\frac{e^{\lambda u_{ij}}}{\sum\limits_{k:\,u_{ik}>0} e^{\lambda u_{ik}}}\, dG(\gamma)dF(\kappa)$$

The FOC of this equation is:

$$\underbrace{[\beta_j h(\bar{\kappa_j})+R_j]}_{A}\underbrace{\int_{\tau^d-\tau^o-h(\bar{\kappa_j})}^{\infty}\frac{e^{\lambda u_{ij}}\mid\bar{\kappa_j}}{\sum\limits_{k:\,u_{ik}>0} e^{\lambda u_{ik}}\mid\bar{\kappa_j}}\, dG(\gamma)}_{B}\underbrace{f(\bar{\kappa_j})}_{C}=0 \tag{13}$$

where we use "$\mid\bar{\kappa_j}$" to clumsily indicate that we integrate γ over patients where $\kappa_i=\bar{\kappa_j}$. As an integral over a strictly positive value and the atom of a density function,

[13] Schnell splits the costs a patient will face into costs of visit (τ^d), cost of purchase (τ^o) and search cost (τ^s), whereas we consider only the first.

respectively, the factors B and C in the equation are non-negative. For this reason, the only way for equation (13) to be satisfied is by obeying the following equality:

$$R_j = -\beta_j h(\bar{\kappa_j}) \tag{14}$$

This is the same result as in the baseline model presented in Schnell (2022), without a secondary market. This equality requires that marginal utility in $\bar{\kappa_j}$ must be 0, i.e. that revenue (R_j) must equal "altruistic" loss ($\beta_j h(\bar{\kappa_j})$), which presupposes that the marginal patient granted prescription opioids suffers a net loss in utility (the negative externalities outweigh its medical benefit as a palliative).

Equation (14) also implies that the value of κ_j is chosen with no considerarion of the strategies employed by other physicians, they quite literally don't enter into the equation. As we have argued, additive separability over clients means the *aggregate* demand for physician j's services doesn't influence marginal utility by a single patient i, and so it has no sway in the optimality condition. Physician j follows simply the rule that she won't grant a prescription (or sick leave, as in our case) below $\bar{\kappa_j}$, however that might affect the size of her "clientele". In order for her to care about *aggregate* values, like her total expected demand, making her take her market share and thus the strategies of other doctors into account, *strict non-linearity* must be introduced into her utility function, either through $R_j(\cdot)$ or, as in our case, a function $P(\cdot)$.

A.3 Strategic Complementarity of $\bar{\kappa_j}$'s (Proof of Proposition 3.3)

We defined the components s_{ij} of S_i as:

$$s_{ij} \equiv g_i(u_{ij}, u_{i,-j})$$

where $g_i(\cdot)$ is a continuous function weakly increasing in the first argument u_{ij}, weakly decreasing in $u_{i,-j}$. Our defining s_{ij} this way has two interlinked corollaries:

Corollary A.1.

$$s_{ij} \mid \kappa_i < \bar{\kappa_j} \quad \leq \quad s_{ij} \mid \kappa_i \geq \bar{\kappa_j} \quad \textit{(with strict inequality if } \gamma_i > 0\textit{)}.$$

Proof. For a fixed value of V_j and κ_i, for all i, the value of $U_i(V_j, \bar{\kappa_j})$ is $V_j\kappa_i - \tau_j$ if $\kappa_i < \bar{\kappa_j}$, and $\gamma_i + V_j\kappa_i - \tau_j$ if $\kappa_i \geq \bar{\kappa_j}$, where $\gamma_i \geq 0$. Given that s_{ij} is weakly increasing in u_{ij}, and we see that u_{ij} is weakly higher if $\kappa_i \geq \bar{\kappa_j}$, the corollary follows. □

Corollary A.2.

$$\frac{\Delta s_{ij}}{\Delta \bar{\kappa_j}} \leq 0 \qquad\qquad \frac{\Delta s_{ij}}{\Delta \bar{\kappa_l}} \geq 0, \forall l \neq j$$

Proof. From the argumentation in corollary 1 follows that u_{ij} is weakly decreasing in $\bar{\kappa_j}$, $\forall j$. Take some $l \neq j$, then s_{ij} is defined in turn as weakly decreasing in u_{il}, which implies it is increasing in $\bar{\kappa_l}$. □

Both corollaries hinge upon our definition of $U_i(\cdot)$ as a step function over κ_i, such that it is discontinuous at $\kappa_i = \bar{\kappa_j}$, where there's a discrete jump of magnitude $\gamma_i \geq 0$. We use them in the next proposition.

Proposition A.3.

$$\frac{\partial Q_j}{\partial \bar{\kappa_j}}, \frac{\partial X_j}{\partial \bar{\kappa_j}} \leq 0 \qquad\qquad \frac{\partial Q_j}{\partial \bar{\kappa_l}}, \frac{\partial X_j}{\partial \bar{\kappa_l}} \geq 0, \forall l \neq j$$

Proof. We recall our function of patient demand for physician j is:

$$Q_j(\bar{\kappa_j}, \bar{\kappa}_{-j}) \;=\; \int_0^\infty \int_0^\infty s_{ij}(\kappa, \gamma)\, dG(\gamma)\, dF(\kappa)$$

Were physician j to increase her threshold from $\bar{\kappa_j}$ to $\bar{\kappa_j}' = \bar{\kappa_j} + \epsilon, \epsilon > 0$, for the mass of patients whose $\kappa_i \in [\bar{\kappa_j}, \bar{\kappa_j} + \epsilon)$, utility would weakly decrease and thus so would their s_{ij}, falling to $s_{ij}' \leq s_{ij}$ (see first corollary, think of s_{ij}' as $s_{ij} \mid \kappa_i < \bar{\kappa_j}$ as opposed to $s_{ij} \mid \kappa_i \geq \bar{\kappa_j}$).

The difference that would make for patient demand *caeteris paribus* given the $\bar{\kappa}_{-j}$ strategies of the other physicians (ommitted in our notation) is:[14]

$$\Delta Q_j = Q_j(\bar{\kappa_j} + \epsilon) - Q_j(\bar{\kappa_j}) = \int_{\bar{\kappa_j}}^{\bar{\kappa_j}+\epsilon} \int_0^\infty \{s_{ij}' - s_{ij}\}\, dG(\gamma)\, dF(\kappa)$$

By the definition of a partial derivative:[15]

$$\begin{aligned}\frac{\partial Q_j}{\partial \bar{\kappa_j}} &= \lim_{\epsilon \to 0} \frac{\int_{\bar{\kappa_j}}^{\bar{\kappa_j}+\epsilon} \int_0^\infty \{s_{ij}' - s_{ij}\}\, dG(\gamma)\, dF(\kappa)}{\epsilon} \\ &= \int_0^\infty \{s_{ij}(\bar{\kappa_j}, \gamma)' - s_{ij}(\bar{\kappa_j}, \gamma)\}\, dG(\gamma)\, f(\bar{\kappa_j})\end{aligned}$$

where the derivative is negative if s_{ij}' is strictly lower than s_{ij}, meaning a shift in leniency would affect the utility and optimal strategy of some positive mass of patients around $\bar{\kappa_j}$; or null, if $s_{ij}' = s_{ij}$.

In a similar vein, consider some physician $l \neq j$ shifting her threshold from $\bar{\kappa_l}$ to $\bar{\kappa_l}' = \bar{\kappa_l} + \epsilon, \epsilon > 0$. The fall in patient strategies in $[\bar{\kappa_j}, \bar{\kappa_j} + \epsilon)$ from s_{il} to $s_{il}^\star \leq s_{il}$ could provide a windfall for j (and all other physicians), raising s_{ij} to $s_{ij}^\star \geq s_{ij}$ (see previous corollary). In that case:

$$\Delta Q_j = \int_{\bar{\kappa_l}}^{\bar{\kappa_l}+\epsilon} \int_0^\infty \{s_{ij}^\star - s_{ij}\}\, dG(\gamma)\, dF(\kappa)$$

[14] Besides simplifying $Q_j(\bar{\kappa_j}, \bar{\kappa}_{-j})$ into $Q_j(\bar{\kappa_j})$, we also forego the arguments of $s_{ij}(k, \gamma)$, as we've done many times already, to avoid notation clutter.

[15] The notation for $s_{ij}(\bar{\kappa_j}, \gamma)'$ is to be taken to mean the strategy $s_{ij}' \leq s_{ij}$ of the mass of patients at different levels of γ atomized at $\kappa_i = \bar{\kappa_j}$.

thus rendering:

$$\begin{aligned}\frac{\partial Q_j}{\partial \bar{\kappa_l}} &= \lim_{\epsilon \to 0} \frac{\int_{\bar{\kappa_l}}^{\bar{\kappa_l}+\epsilon} \int_0^\infty \{s_{ij}{}' - s_{ij}\}\, dG(\gamma)\, dF(k)}{\epsilon} \\ &= \int_0^\infty \{s_{ij}(\bar{\kappa_l}, \gamma)^\star - s_{ij}(\bar{\kappa_l}, \gamma)\}\, dG(\gamma)\, f(\bar{\kappa_l})\end{aligned}$$

As for X_j, the steps taken are almost the same, only that after raising the threshold by ϵ the amount of sick leaves granted goes around $\bar{\kappa_j}$ goes to 0, thus giving out:

$$\frac{\partial X_j}{\partial \bar{\kappa_j}} = -\int_0^\infty s_{ij}(\bar{\kappa_j}, \gamma)\, dG(\gamma)\, f(\bar{\kappa_j})$$

such that it is simply required that j has positive demand for the mass of patients whose $\kappa_i = \bar{\kappa_j}$ for this derivative to be strictly negative, else it is null.

Likewise:

$$\frac{\partial X_j}{\partial \bar{\kappa_l}} = -\int_0^\infty s_{ij}(\bar{\kappa_l}, \gamma)\, dG(\gamma)\, f(\bar{\kappa_l})$$

□

Before dealing with Proposition 3.3 we need one last intermediary proposition.

Proposition A.4.

$$\frac{\partial^2 Q_j}{\partial \bar{\kappa_j} \partial \bar{\kappa_l}} = 0 \quad \textit{and} \quad \frac{\partial^2 X_j}{\partial \bar{\kappa_j} \partial \bar{\kappa_l}} = 0\ , \quad \forall j \textit{ and } l,\ \bar{\kappa_j} \neq \bar{\kappa_l}$$

Proof. Our proof will be heuristic. Consider

$$\frac{\partial Q_j}{\partial \bar{\kappa_j}} = \int_0^\infty \{s_{ij}(\bar{\kappa_j}, \gamma)' - s_{ij}(\bar{\kappa_j}, \gamma)\}\, dG(\gamma)\, f(\bar{\kappa_j})$$

Patient strategies s_{ij} are taken at the atom $\kappa_i = \bar{\kappa_j}$. For any other physician l such that $\bar{\kappa_l} \neq \bar{\kappa_j}$, an infinitesimally small change to $\bar{\kappa_l}$ doesn't affect the utility of patients at $\bar{\kappa_j}$, who will remain above or below l's threshold same as before. Thus their strategy s_{ij} also remains unaffected. If $\bar{\kappa_l}$ *does* equal $\bar{\kappa_j}$, then the following limit is not properly defined:

$$\lim_{\epsilon \to 0} \frac{\frac{\partial Q_j(\kappa_l + \epsilon)}{\partial \bar{\kappa_j}} - \frac{\partial Q_j(\kappa_l)}{\partial \bar{\kappa_j}}}{\epsilon}$$

because as ϵ approaches 0 from the right the limit is weakly positive (≥ 0), whereas approaching from the left it's strictly 0. Meaning the derivative $\frac{\partial^2 Q_j}{\partial \bar{\kappa_j} \partial \bar{\kappa_l}}$ doesn't exist for j and $l \neq j$ when $\bar{\kappa_j} = \bar{\kappa_l}$.

□

Proof of Proposition 3.3.

Consider some two physicians j, l, such that $\bar{\kappa_j} \neq \bar{\kappa_l}$. If we breakdown the mixed partial derivative of U_j over $\bar{\kappa_j}$ and $\bar{\kappa_l}$, we get:

$$\begin{aligned}
\frac{\partial^2 U_j}{\partial \bar{\kappa_j} \partial \bar{\kappa_l}} &= \underbrace{R_j''(Q_j(\bar{\kappa_j}, \bar{\kappa}_{-j}))}_{\leq 0} \cdot \underbrace{\frac{\partial Q_j}{\partial \bar{\kappa_l}}}_{\geq 0} \cdot \underbrace{\frac{\partial Q_j}{\partial \bar{\kappa_j}}}_{\leq 0} + \underbrace{R_j'(Q_j(\bar{\kappa_j}, \bar{\kappa}_{-j}))}_{\geq 0} \underbrace{\frac{\partial^2 Q_j}{\partial \bar{\kappa_j} \partial \bar{\kappa_l}}}_{=0} \\
&\quad - \underbrace{P''(X_j(\bar{\kappa_j}, \bar{\kappa}_{-j}))}_{\geq 0} \cdot \underbrace{\frac{\partial X_j}{\partial \bar{\kappa_l}}}_{\geq 0} \cdot \underbrace{\frac{\partial X_j}{\partial \bar{\kappa_j}}}_{\leq 0} - \underbrace{P'(X_j(\bar{\kappa_j}, \bar{\kappa}_{-j}))}_{\geq 0} \underbrace{\frac{\partial^2 X_j}{\partial \bar{\kappa_j} \partial \bar{\kappa_l}}}_{=0} \\
&= \underbrace{R_j''(Q_j(\bar{\kappa_j}, \bar{\kappa}_{-j})) \cdot \frac{\partial Q_j}{\partial \bar{\kappa_l}} \cdot \frac{\partial Q_j}{\partial \bar{\kappa_j}}}_{\geq 0} - \underbrace{P''(X_j(\bar{\kappa_j}, \bar{\kappa}_{-j})) \cdot \frac{\partial X_j}{\partial \bar{\kappa_l}} \cdot \frac{\partial X_j}{\partial \bar{\kappa_j}}}_{\leq 0} \quad \geq 0
\end{aligned}$$

and positive mixed derivative of physicians' utility over their strategy is the definition of strategic complementarity.

$\square$

A.4 Optimal Reserve Utility (Proof of Equation 8)

We follow the argument of Ljungqvist and Sargent (2012). Let for the moment $\tilde{F}_i(U) = \Pr\{u_{ij} \leq U\}$ be the distribution of utility patient i recieves from doctor visits, with $\tilde{F}_i(0) = 0$, $\tilde{F}_i(B) = 1$, for $B < \infty$. Each period, the patient draws an 'offer' U from $\tilde{F}_i$, which is to say, his search lead him to a physician j such that $u_{ij} = U$. He may either accept and have physician j assigned to him, recieving utility U per period henceforth, or look for a new physician next period.

Let u_t be the patient's utility for period t, where it's 0 if he doesn't visit a physician and U if he has accepted an assigned physician which renders him utility $u_{ij} = U$. This patient will seek to maximize $\mathbb{E}[\sum_{t=0}^{\infty} \beta^t u_t]$, where $0 < \beta < 1$ is his discount factor.

Let $V(U)$ be the expected value of $\sum_{t=0}^{\infty} \beta^t u_t$ of a patient with an offer U in hand. Assuming no recall, under optimal behavior the value function $V(U)$ satisfies the Bellman equation:

$$V(U) = \max\left\{\frac{U}{1-\beta},\ \beta \int V(U')\, d\tilde{F}_i(U')\right\} \tag{15}$$

The patient chooses an optimal value $\bar{U}_i$, such that offers $U \geq \bar{U}_i$ are accepted (the visit takes place). As such, the solution takes on the following form:

$$V(U) = \begin{cases} \frac{\bar{U}_i}{1-\beta} + \beta \int V(U')\, d\tilde{F}_i(U') & \text{if } \ U \leq \bar{U}_i, \\ \frac{U}{1-\beta} & \text{if } \ U \geq \bar{U}_i. \end{cases} \tag{16}$$

Using equation (16), we can convert the functional equation (15) into an ordinary equation in the reservation utility $\bar{U_i}$. Evaluating $V(\bar{U_i})$ and using equation (16), we have:

$$\frac{\bar{U_i}}{1-\beta} = \beta \int_0^{\bar{U_i}} \frac{\bar{U_i}}{1-\beta}\, d\tilde{F}_i(U') + \beta \int_{\bar{U_i}}^{B} \frac{U'}{1-\beta}\, d\tilde{F}_i(U')$$

or

$$\frac{\bar{U_i}}{1-\beta} \int_0^{\bar{U_i}} d\tilde{F}_i(U') + \frac{\bar{U_i}}{1-\beta} \int_{\bar{U_i}}^{B} d\tilde{F}_i(U') \\ = \beta \int_0^{\bar{U_i}} \frac{\bar{U_i}}{1-\beta}\, d\tilde{F}_i(U') + \beta \int_{\bar{U_i}}^{B} \frac{U'}{1-\beta}\, d\tilde{F}_i(U')$$

or

$$\bar{U_i} \int_0^{\bar{U_i}} d\tilde{F}_i(U') = \frac{1}{1-\beta} \int_{\bar{U_i}}^{B} \left(\beta U' - \bar{U_i}\right) d\tilde{F}_i(U').$$

Adding $\bar{U_i} \int_{\bar{U_i}}^{B} d\tilde{F}_i(U')$ to both sides gives

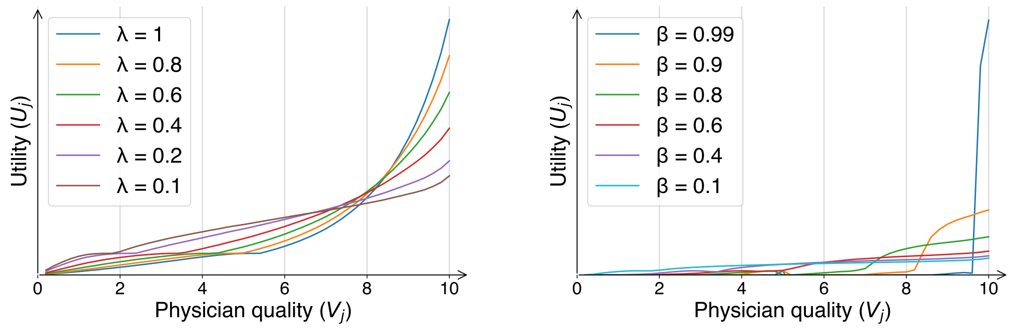
$$\bar{U_i} = \frac{\beta}{1-\beta} \int_{\bar{U_i}}^{B} \left(U' - \bar{U_i}\right) d\tilde{F}_i(U') \tag{17}$$

Finally, for our purposes, we discretize $\int_{\bar{U_i}}^{B} \left(U' - \bar{U_i}\right) d\tilde{F}_i(U')$ as a summation over the countable set of J doctors. The discrete equivalent of equation (17) is:

$$\bar{U_i} = \frac{\beta}{1-\beta} \sum_{j=1}^{J} \left\{ \frac{\mathbb{1}[u_{ij} \geq \bar{U_i}] \cdot (u_{ij} - \bar{U_i})}{\mathbb{1}[u_{ij} \geq \bar{U_i}]} \right\} \tag{8}$$

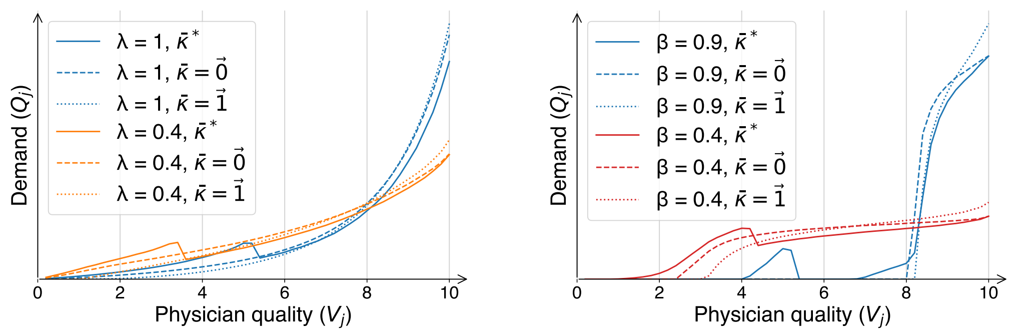
Which is the function we say characterizes patient thresholds in the explicit search model.

A.5 More Figures



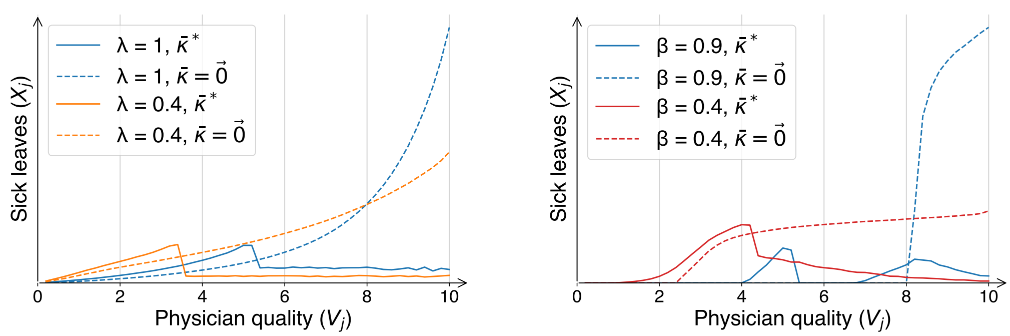
(a) Logit model (b) Sequential model

Figure 10: Equilibrium physician utility (U_j) by quality (V_j) for different model parameters, equilibrium $\bar{\kappa}^*$



(a) Logit model (b) Sequential model

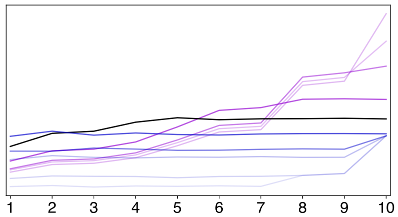
Figure 11: Comparison of patient demand (Q_j) by physician quality (V_j) for different model parameters and threshold vectors $\bar{\kappa}$



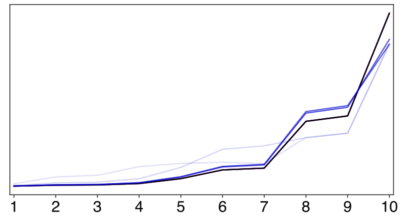
(a) Logit model (b) Sequential model

Figure 12: Comparison of sick leaves issued (X_j) by physician quality (V_j) for different model parameters and threshold vectors $\bar{\kappa}$

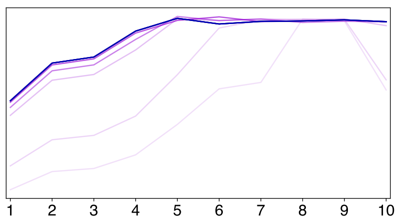
A.6 Identification Panels



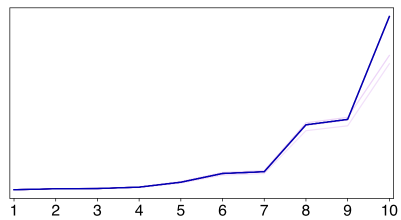
(a) X_j, variations in p



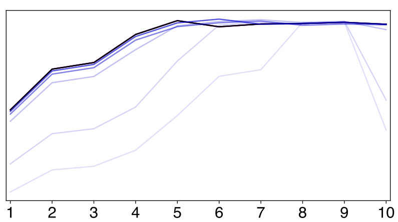
(b) X_j^{30}, variations in p



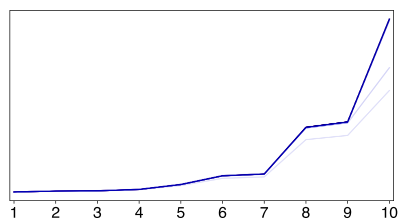
(c) X_j, variations in μ



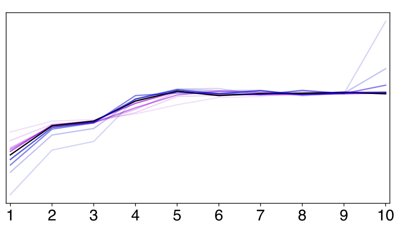
(d) X_j^{30}, variations in μ



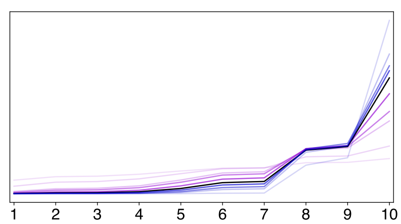
(e) X_j, variations in σ



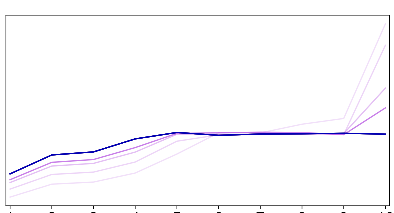
(f) X_j^{30}, variations in σ



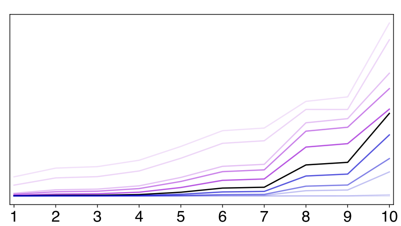
(g) X_j, variations in λ_s



(h) X_j^{30}, variations in λ_s



(i) X_j, variations in $\bar{\kappa}_{\max}$



(j) X_j^{30}, variations in $\bar{\kappa}_{\max}$

Figure 13: Sick leaves issued by quality bin as parameters vary.
Black: Estimated value. Violet: Lower. Blue: Higher.
Transparency increasing with distance from estimated value

A.7 Parameter Sources

Physician quality by bins (V_1 ... V_{10})

As mentioned in the main text, three criteria are considered for the *quality score*:

i Average duration of sick leaves issued.

ii Percentage of sick leaves of 30+ days over total in sampled period.

iii Amount of sick leaves of 30+ days issued.

For each of the criteria the *percentile rank* which each physician occupies is recorded, all in increasing order, i.e. from lowest to highest average duration, from lowest to highest percentage/amount of $\geq$ 30 days sick leaves. Call the percentile rank which physician j occupies in each category p_j^i, p_j^{ii} and p_j^{iii}. The *quality score* of physician j is the simple average of these three, that is:

$$q\_score_j = \frac{1}{3}p_j^i + \frac{1}{3}p_j^{ii} + \frac{1}{3}p_j^{iii}$$

The quality bins are then set up from the decile ranks of physicians in their *quality score*. Bin 1, for example, is simply the first decile rank of q_score, the lowest 10% of physicians according to their quality score. Each bin is therefore equally sized.

The parameters V_1 ... V_{10} associated with each bin are then to be estimated in our GMM exercise, subject to the constraint that they increase in alignment with quality. Simply put: $V_{10} > V_9 > ... > V_1$.

Revenue by visit (r_j) and cost of visit (τ_j)

We use a single r and τ for all physicians on this occasion. In the future a proper match to revenue bins would be a way to further square parameter estimation. Our procedure is arbitrary, but an alternative method would nonetheless stay within the same degree of magnitude as our final values for r and τ, so in this case arbitrariness is not all too relevant.

We select the ten most frequent $cie10$ categories in our sample, the international standard classification of medical conditions. These ten, as well as non-categorized conditions (seen as *empty* in the Table), account for 47.91% of all sick leaves in our database. We then match the conditions to a particular field of practice, though strictly speaking physicians aren't reduced to their own field when issuing sick leaves, so matching is fuzzy.

We then take a *full price* and the fee paid by FONASA affiliates in the Free Choice Modality (MLE) as a reference price, from a tariff book assembled by FONASA itself (2024). We take this to correspond to the physician's revenue by visit (r) and the patients' cost of visit (τ), respectively.

Our final value for r and τ is the weighted average of this prices, weighted by the relative frequency of each category and its matched full and benefit price. This amounts to taking the dot product of the '%' column with the 'Full Price' and 'Aff. Fee' columns in the Table below, respectively.

The resulting values are: $r = 21,182$ and $\tau = 8,944$. For simplicity, we select to use them in the thousands for estimation, i.e. 8.944 instead of 8,944. See the Table below for the referenced variable columns.

cie10 cat.	Instances	%	Informal Description	Matched Field	Full Price	Aff. Fee
M54	379,332	0.20	Back pain	Traumatology	$ 18,130	$ 7,250
F32	304,399	0.16	Depression	Psychiatry	$ 27,510	$ 11,000
F41	284,129	0.15	Anxiety	Psychiatry	$ 27,510	$ 11,000
F43	244,468	0.13	Stress	Psychiatry	$ 27,510	$ 11,000
J20	191,615	0.10	Bronchitis	General	$ 14,270	$ 7,440
A09	186,535	0.10	Gastroenteritis	General	$ 14,270	$ 7,440
empty	152,239					
M75	115,585	0.06	Shoulder	Traumatology	$ 18,130	$ 7,250
J00	92,660	0.05	Common cold	General	$ 14,270	$ 7,440
M51	54,800	0.03	Intervertrebal disc	Traumatology	$ 18,130	$ 7,250
K52	49,182	0.03	Gastroenteritis	General	$ 14,270	$ 7,440

Table 3: Most common health conditions reported by physicians in sick leaves issued, matched to corresponding fees of service