Information Design in Insurance Markets: Selling Peaches in a Market for Lemons*

Daniel Garcia[†], Roee Teper[‡], and Matan Tsur[§]

September, 26th 2017

Abstract

This paper characterizes the optimal information structure in insurance markets in the presence of adverse selection. The optimal information structure minimizes ex-post risk subject to a participation constraint for insurees and a break-even constraint from insurers. In the unique optimal information structure, trade occurs with probability one and different risk-types are pooled together in the same signal. Surprisingly, these signals are not monotone so that low types are pooled with high types, while intermediate types are bundled together. We provide a simple algorithm that delivers the optimal information structure and derive comparative statics. We explore some applications and generalizations.

Keywords: Risk-Sharing, Adverse Selection, Information Design.

1 Introduction

Adverse selection is a key cause of market failure in competitive insurance markets. In an attempt to increase efficiency various policy interventions have been

*We are grateful to Eddie Dekel and Alessandro Lizzeri for useful comments and seminar participants at the Universite de Montreal for their suggestions.

[†]University of Vienna. Email to daniel.garcia@univie.ac.at

[‡]University of Pittsburgh. Email to rteper@pitt.edu

[§]University of Vienna. Email to matan.tsur@univie.ac.at

introduced. These interventions either affect the market mechanism, for example individual mandates or healthcare subsidies, or the information that is available to market participants. An important example of the latter are health insurance regulations which allow insurers to discriminate and charge premiums based on information about individual risk characteristics, such as preexisting conditions or genetic mutations. Price differentiation may increase participation of healthier individuals and alleviate adverse selection, but it also eliminates opportunities for risk cross-subsidization.

This observation points to a basic trade-off in insurance markets. The economic literature has analyzed adverse selection (e.g., Akerlof (1970)) and risk cross-subsidization (e.g., Hirshleifer (1971)) in isolation, but the social value of information in many insurance markets depends on both. This paper provides an equilibrium analysis of the optimal information structure in competitive insurance markets with adverse selection. Our results have important policy implication for the optimal limits of discrimination in health insurance regulations.

We consider a large policy space and assume the regulator can design a public rating system, which stochastically assigns a rating to each agent depending on her risk characteristics. The rating system determines the information based on which insurers can discriminate, i.e., the information structure in the market. *We show that in any such market there exists a unique optimal rating system, under which all individuals are insured and ratings pool risk characteristics in a negatively assortative fashion.* Importantly, the uniqueness of the optimal rating system and the negative assortative structure require no assumptions on the underlying distribution of risk or on agent's preferences aside from risk aversion. While our motivation and primary focus is on health insurance markets, the analysis and results apply to any competitive insurance market, including financial markets and labor markets.

Optimal Rating Systems

We consider an insurance market with risk averse agents and risk neutral insurers. Each agent has a privately known risk-type which is a distribution of medical costs. The insurance market is competitive and trade takes place as in Akerlof

(1970): insurers offer full coverage contracts at prices that equal the expected cost of the (average) patient choosing the contract. We assume that regulator cannot directly influence the market mechanism (for example with a mandate or price caps), but can design a rating system, which is a noisy public signal of each agent's risk-type (as in Blackwell (1951, 1953)). Insurers offer contracts contingent on public ratings.

Let us first consider a simple example to demonstrate the model. Suppose that each individual either has a genetic mutation or not, and either has a preexisting condition or not. There are four risk-types in the population: agents with no mutation and no preexisting condition, agents with a mutation and a preexisting condition, and so on. A rating system can specify that agents with no preexisting condition receive rating A and agents with a preexisting condition receive rating B. In turn, insurers offer each agent a contract at a price equal to the expected cost of the (average) agent choosing the contract within their rating group. The healthiest types will participate only if the average medical cost of the group of agents with no preexisting condition is sufficiently low. The regulator can also implement a more diversified risk pool by designing a rating system such that only a fraction of the population with a preexisting condition, receive rating B, while the rest of the population, including the remaining population with a preexisting condition, receive rating A.[1]

Given a set of risk types and a prior distribution, the problem of the regulator is to design the rating system that maximizes the ex-ante expected social welfare subject to the individual participation constraints. For exposition clarity, assume that types can be ordered from healthier to sicker, in the sense that types with higher expected medical costs are those who are willing to pay more for full coverage. We say that a sub-population of agents is an "interval of sickest agents" if the expected medical costs of each agent within the sub-population is (weakly) greater than the expected cost of any agent outside this sub-population.

We prove that the following algorithm yields the unique optimal rating system:

[1]Our approach also alleviates a technical difficulty. Since there are hundreds of individual risk characteristics that jointly predict the likelihood of developing sickness, the question of which information should insurers be allowed to use appears hopelessly cumbersome. By resorting to stochastic ratings, we present a general and tractable solution to the problem.

1. Absent any public rating system, if no participation constraint is violated, the process is complete and the optimal system reveals no information. Otherwise, move to step 2.

2. The population of the healthiest type (type 1 in this example) receive rating R_1 with probability 1. An interval of sickest agents in the population receives the same rating R_1, so that the posterior distribution of the average cost associated with the signal R_1 makes the healthiest type indifferent between buying insurance and not. Return to step 1 with the residual subpopulation of agents, if it is not empty. Otherwise, the process is complete.

In other words, the algorithm sequentially matches agents in a negatively assortative manner: in each iteration, the remaining healthiest type is matched with an interval of the sickest agents in the remaining population. The proof shows that three properties characterize the optimal rating system: 1) No exclusion; 2) No rents at the top; and 3) Negative assortative pooling.

To see this, suppose first that some individuals receive a signal and choose not to participate (because the average cost of their risk pool is too high). Then we can allocate only the agents who did not participate to an exclusive pool without influencing the prices in the other pools, which is a Pareto improvement. Second, observe that for any two risk pools, the participation constraint binds for the healthiest agents in the healthier pool (i.e. the pool with the lower average cost). Because if their participation constraint slacks, we can move some of the sickest individuals in the sicker pool (i.e. the pool with higher average cost) to the healthier pool without violating the participation constraints. The average price of the two pools always equals the expected costs of the agents participating in these pools, and thus the resulting outcome is a mean preserving contraction of the posterior distribution of prices, which is a welfare improvement.

It follows that the entire population of the healthiest types are allocated to a single pool, the average cost of this pool makes them indifferent, and this pool has a lower average cost than any other pool. The question is which are the types pooled with the healthiest types. The key point of our analysis is that to minimize price dispersion, the additional agents in this pool come from the bottom (the sickest types). We call this negative assortative pooling.

We will see that the optimal rating system is more precise whenever: 1) the adverse selection problem is more acute and 2) there is less uncertainty about the medical costs, either because private information is more precise or there is less idiosyncratic risk. These comparative statics suggest that restricting the use of genetic information is more likely to spur welfare in health insurance, where most individuals are covered, than in annuity markets, where only a small fraction of the population participates (see, e.g., Chiappori (2006)). Moreover, as accurate genetic information becomes more widespread, the regulations restricting their use should be more lax.

Related Literature

This work relates to four strands of the literature. First, following the seminal work of Hirshleifer (1971), a number of papers have shown that under quite general conditions the release of public information in insurance markets is socially harmful (see Schlee (2001) and references therein). These models do not consider agents with private information, thus leaving out an important positive effect of information disclosure: making private information public may reduces adverse selection. Our analysis focuses on both forces and fully characterizes the optimal information structure.

Second, there is a small but influential literature on the role of information in markets with adverse selection. Levin (2001) shows that in a lemons market, if both sellers and buyers are not fully informed, revealing more information may be inefficient (because it changes the distribution of seller's valuations). But if the seller is fully informed, as is the case in our setup, the release of public information always improves welfare. In a recent contribution, Bar-Isaac et al. (2017) provide a detailed analysis of multidimensional information structures in equilibrium outcomes in labor markets with risk-neutral workers and firms. In these papers, every agent is risk-neutral, and hence there is no intrinsic cost of providing more information.

Most related to our work is a recent study by Goldstein and Leitner (2015) on public information disclosure in financial markets in which the motive for trade is not insurance but rather to obtain outside liquidity to finance a profitable investment. They consider a risk neutral seller of a risky asset that receives an additional payoff bump only if her return is above some threshold. Importantly, the seller does not have a preference for mean preserving spreads of distribution of returns.[2] The optimal of the algorithm we construct hinges on risk aversion (in particular, a preference for mean preserving contractions), and therefore the optimal disclosure rule in their liquidity market is different. They first consider a case where the seller does not have private information regarding the asset, and show that the optimal disclosure rule is a cutoff rule. In our model, absent private information, it is never optimal to reveal any accurate information. In the private information case, Goldstein and Leitner do not fully characterize the optimal disclosure rule, which need not be unique. But if the expected payoff of the asset is sufficiently low, they show that the optimal rule has a non-monotonicity property in that some of the high types are matched with some low types, but a subset of low types is always excluded. In our insurance market model, all types are uniquely matched negatively assortatively. Despite these differences, our view is that these results are complementary and reinforce the point that non-monotonicty is an important feature of the optimal information disclosure regardless of the motive for trade.

Finally, our work is related to the literature on strategic persuasion, the underpinnings of which were introduced by Aumann and Maschler (Aumann and Maschler (1995)), and was revived by Kamenica and Gentzkow (2011). Conceptually, the models studied in that literature are different than ours. In particular, there is no conflict of interests between the regulator (the information designer) and market participants. This also implies that no commitment issues arise. Technically, the ex-post participation constraints make our analysis novel and we need not make use of concavification techniques. Also related is the work

[2] If the mean of the distribution of returns is below the threshold, the seller strictly prefers a mean preserving spread of this distribution (she is risk loving). If the mean of the distribution of returns is above the threshold, the seller is strictly worse-off with a mean preserving spread (she is risk averse).

of Bergemann and Morris (2013, 2016), who characterize the set of outcomes that can be supported in a Bayes-Nash equilibrium for some information structure as the Bayesian extension of the correlated equilibrium of the basic game. We use these ideas to characterize the set of outcomes that can be supported in an insurance markets with perfect competition and an arbitrary information structure.

Organization

The rest of the paper is organized as follows. In Section 2 we present the general framework and discuss the main assumptions. In Section 3 we provide a full characterization of the optimal test and provide some comparative statics results. In Section 4 we extend the model to allow the planner to directly cross-subsidize across pools and show that the optimal test remains non-monotonic. In Section 5 we characterize the set of allocations for the case of three types, which allows to set-identify the coefficient of risk-aversion using market-level information and provide a method to compute the set of equilibrium payoffs for arbitrary information structures, which allows to perform robust counterfactual. Section 6 contains some concluding remarks. All proofs appear in the Appendix.

2 General Set-up

We consider a heterogeneous population of risk averse agents, or patients, who are subject to some idiosyncratic health risk. We start by assuming that heterogeneity across agents is through their medical risks. Agents in the population are distributed over a finite set of types $\Theta = \{1, 2, ..., N\}$ according to the probability distribution μ. A type i agent is associated with a distribution of medical costs,[3] $f_i \in \Delta(X)$, where $X \subset \Re$. Every patient has a utility function $u : \Re \to \Re$ which is continuous, strictly increasing and strictly concave. We let $\theta_1, \ldots, \theta_N$ be the expected medical cost in type i, $\theta_i = E_{f_i}(x)$ and we label the types so that and $\theta_N > \theta_{N-1} > \ldots > \theta_1$. Let $U_i = E_{f_i}\big(u(w-x)\big)$ be type i's expected utility,

[3]A risk-type can be derived from a primitive cost function which depends on the profile of individual characteristics of the patient.

where w is the individual's wealth. We will later discuss the implications of type dependent wealth levels. Let ϕ_i be the willingness-to-pay for full insurance for type i. That is, $u(w - \phi_i) = U_i$. We make the following assumption:

Assumption 1. $\phi_l > \phi_i$ if and only if $\theta_l > \theta_i$,

In words, individuals with higher expected medical costs, are willing to pay more to obtain insurance. Assumption 1 and the assumption that all types have the same initial wealth, simplify the exposition but neither is necessary for our results to hold, as we show in Subsection 3.3.

In addition, there is a number of risk-neutral competitive insurers who compete in prices and are restricted to offer full insurance contracts.

Information: We assume that the patient knows her type (her individual risk characteristics) and the insurer knows only the prior distribution over realized states. Since every patient knows the state, in what follows we refer to the realized state as the patient's type. [4]

Definition: A *rating system*, σ, is a probability distribution over a set of signals (or, ratings) $S = (s_0, ..., s_M)$ conditional on the realization of the type $i \in \Theta$.

We assume that the set of available ratings is sufficiently rich so that $\mid S \mid = M > N$. The realization of the rating is public information, and creates *risk pools*. A pool associated with the a rating $s \in S$ is the posterior distribution of types among the individuals receiving the rating s. Let $\sigma_{ji} = \Pr_\sigma(s_j \mid i)$ so that $\sum_{j=1}^{M} \sigma_{ji} = 1$ for all $i = 1, 2, .., N$ and

$$\Pr_{\sigma}(i|s = s_j) = \frac{\Pr(s = s_j|i)\Pr(i)}{\Pr(s_j)} = \frac{\sigma_{ji}\mu_i}{\sum_{l=1}^{N} \sigma_{jl}\mu_l}.$$

We denote by $E_j(\theta)$ the expected medical cost conditional on receiving rating j.

We assume that a benevolent regulator designs the rating system at the ex-ante stage in order to maximize the utilitarian welfare of patients with Pareto weights given by the prior distribution.

[4] In Subsection 4.3 we discuss alternative assumptions regarding the information accessible to patients and insurers.

Trading Mechanisms: After observing the realization of the rating system, trade occurs. There are several possible mechanisms that determine the price. Following Akerlof (1970), we focus on simple contracts whereby the insurer provides full insurance covering the entire medical expenses in exchange for a fixed price, or premium. While in general there may be many such prices, we focus on the minimum price which achieves the most efficient allocation. Hence, the price associated with signal j satisfies,

$$t_j = \min\{t \ : \ t = E_j\big(\theta | i \in A(t)\big) \text{ and } \ A(t) = \{i : t \leq \phi_i\}\}.$$

In words, $A(t)$ is the set of types willing to accept price t and the expected cost of covering an average patient with rating s_j who is willing to accept is $E_j\big(\theta \mid i \in A(t)\big)$.[5] The price is well defined and, by Assumption 1, the set of types $A(t_j)$ is an interval.[6]

Prior to concluding this section with a few remarks, we summarize the timing of the model. The regulator designs a rating system, and patients privately learn their types. Then, public ratings are realized according to the designed system and patients' types. Lastly, trade occurs, the outcome of the lottery f_i is realized, and consumption takes place.

Remark 1: While we assume perfect competition, all our results extend to the case in which the price is computed using a constant load λ on the actuarially fair price so that $t = (1+\lambda)E_j\big(\theta \mid i \in A(t)\big)$ as long as it holds that $(1+\lambda)\theta_i \leq \phi_i$ for all types i.[7]

Remark 2: The model applies to any competitive market in which risk-neutral agents provide insurance to risk-averse agents who have private informa-

[5]Notice that as long as the willingness-to-pay for insurance of each type is monotone with her mean payoff, then $i \in A(t) \implies l \in A(t), \forall l < i$ (e.g., these sets are uniquely identified intervals of types).

[6]Note also that since the patient extracts all the rents, the optimal test does not incorporate the rent-extraction versus rent-creation, which is in the heart of some recent contributions in the information design in markets (e.g. Roesler and Szentes (2017)).

[7]Estimates of loadings range from 6% in some health insurance markets to 30% in long-term care insurance markets (Brown and Finkelstein (2011)), but they do not vary so much across groups (Mitchell et al. (1999)).

tion. In particular, we can relabel the notation to represent an asset market, whereby risk-averse sellers own assets whose return is distributed according to f_i and the regulator can disclose information about the quality of the asset. We explore this application further in Section 4.

3 Optimal Test Design

This section contains the main results of the paper. It is shown that there exists a unique optimal rating system and that such a system is the outcome of simple algorithm. A characterization of the optimal rating system is provided.

The optimal rating system solves the following problem,

$$\max_{\sigma \in \Delta^{M \times N}} \sum_{i=1}^{N} \mu_i \sum_{j=1}^{M} \sigma_{ji} \Big(u(w - t_j) 1_{t_j \leq \phi_i} + U_i 1_{t_j > \phi_i} \Big)$$

$$t_j = \min_t E_j\big(\theta \mid i \in A(t)\big)$$

$$\sum_{j=1}^{M} \sigma_{ji} = 1, \forall i$$

The first-best allocation requires full insurance of both risk sources (μ and f) to the patient. Therefore, the set of Pareto-efficient allocations satisfying ex-ante individual rationality (IR) is spanned by a scalar $\pi \geq 0$, which specifies the insurer's profit with the constraint that $u(w - E_\mu(\theta) - \pi) \geq E_\mu U$. On the opposite extreme, in the case of no trade, each type receives expected utility U_i and the ex-ante expected utility is $E_\mu U = \sum_{i=1}^{N} \mu_i U_i$.

We begin with the analysis of a number of useful benchmarks. First, in the absence of any public information, our model reduces to Akerlof's market for lemons with risk. From Assumption 1, it follows that the market price satisfies $t = \min_t E_\mu(\theta | i \in A(t))$. If $E_\mu(\theta) \leq \phi_1$, then all the agents will trade at price $E_\mu(\theta)$ which is a Pareto-efficient allocation. If $E_\mu(\theta \mid \theta \geq \theta_i) > \phi_i$ for all $i \neq N$, market breakdown occurs because only the sickest type obtains insurance. Second, a rating system that perfectly reveals the type of every patient induces trade with probability one but provides no cross-subsidization between risk-types, so that $t_i = \theta_i$ and expected utility is $\sum_{i=1}^{N} \mu_i u(w - \theta_i)$. This allocation is never Pareto-optimal but may be superior to the allocation without private information (for instance, if there is market breakdown).

The following observations simplify the problem of characterizing the optimal rating system. First, for every rating system σ we can define a distribution of prices $\Pr_\sigma(t_j = t \mid i)$, and two tests are equivalent if they generate the same distribution of prices. It is then without loss of generality to focus on rating systems such that every two signals $s_j, s_{j'}$ induce different posterior expected medical costs, for otherwise the regulator could just merge them into one signal and obtain the same expected payoff. Second, it is without the loss of generality to consider rating systems that implement *no exclusion,* that is all types are insured.

Lemma 0. An optimal rating system satisfies no exclusion.

To see this, suppose that a given type i does not buy insurance following some signal s_j and $\sigma_{ji} > 0$, then t_j does not depend on σ_{ji} and we can therefore construct an alternative test $\hat{\sigma}$ which equals σ except that $\hat{\sigma}_{ji} = 0$ and there exists some signal s_{M+1} with $\hat{\sigma}_{(M+1)i} = \sigma_{ji}$ and $\hat{\sigma}_{(M+1)k} = 0$ otherwise, which strictly improves on σ.

Therefore, following signal s_j, we can identify an associated equilibrium price $t_j = E_j(\theta) = \sum_{j=1}^{N} \Pr_j(\theta_i)\theta_i$, and write the planner's maximization problem as

$$\max_{\sigma \in \Delta^{M \times N}} \sum_{i=1}^{N} \mu_i \sum_{j=1}^{M} \sigma_{ji} u(w - t_j)$$

$$t_j = E_j(\theta) = \sum_{i=1}^{N} \frac{\sigma_{ji}\mu_i}{\sum_{l=1}^{N} \sigma_{jl}\mu_l} \theta_i$$

$$t_j \leq \phi_i, \forall i : \sigma_{ji} > 0$$

$$\sum_{j=1}^{M} \sigma_{ji} = 1, \forall i$$

Effectively, the regulator chooses a rating system to maximize the ex-ante expected utility of a randomly chosen patient after trade, subject to her ex-post participation and a break-even constraint from the insurer. This break-even constraint is in the spirit of the Bayes-neutrality condition under Bayesian Persuasion but has a classical meaning in insurance markets: the premium must be actuarially fair, given the information available in the market. The participation constraint of the insurer is more novel and it arises naturally in markets. A solution to this problem exists because the set of rating systems is compact and the feasible set is non-empty (e.g. full information is always feasible).

Our main result shows that a simple algorithm yields the uniquely optimal rating system.

Theorem 1. The following algorithm yields the unique optimal test. Let $l \in \mathbb{N}$ be a counter variable, set $l = 1$, and denote $\mu^l = \mu$. Then:

Step l_1: If $E_{\mu^l}(\theta) \leq \phi_l$, then set $\sigma_{0i} = 1 - \sum_{j=1}^{l} \sigma_{ji} \forall i$ and stop. Otherwise, create signal s_l with $\sigma_{ll} = 1$, $\sigma_{li} > 0$ only if $\forall k > i$, $\sum_{j=1}^{l} \sigma_{jk} = 1$, and $t_l = E_{\mu^l}(\theta|s_l) = \phi_l$. Proceed to Step l_2.

Step l_2: Stop if there are no individuals remaining in the population. Otherwise, define the prior on the remaining types by

$$\mu_i^{l+1} = \frac{\mu_i^l (1 - \sigma_{li})}{\sum_{k=l}^{N} \mu_k^l (1 - \sigma_{lk})},$$

increase l by one (that is, $l = l + 1$), and go to step $l_{1.}$

Informally, the algorithm can be described as follows. In each "rating stage", if there is no adverse selection (that is, the healthiest individuals in the remaining population purchase insurance), then all types are pooled under the same rating. Otherwise, the healthiest individuals in the population that were not yet rated are pooled with a an interval of the sickest individuals,[8] so that the participation constraint of the healthiest individuals is binding. Then, the process is being repeated with the individuals in population not yet rated.

[8]As discussed in the introduction, we say that a sub-population of agents is an "interval of sickest agents" if the expected medical costs of each agent within the sub-population is (weakly) greater than the expected cost of any agent outside this sub-population.

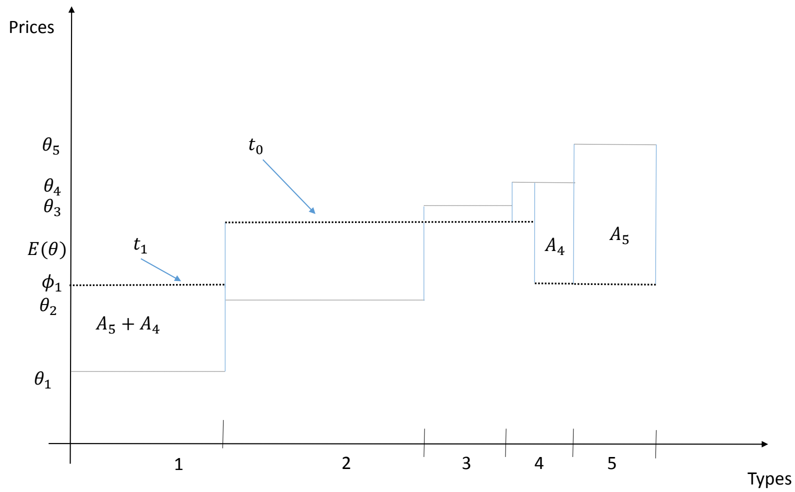

Figure 1: The optimal rating system

This is depicted in Figure 1 for the case of five types. In this case, $E(\theta) > \phi_1$ so the first best is not possible. The optimal rating system pools together types $1, 5$ and a fraction of type 4 in such a way that $\mu_1(\phi_1 - \theta_1) = \mu_5(\theta_5 - \phi_1) + \mu_4\sigma_{14}(\theta_4 - \phi_1)$. Since, the residual distribution of types induces an expected cost that lies below ϕ_2 the algorithm assigns the remaining types to a single signal, s_0, with price t_0 satisfying $\sum \mu_i^2(t_0 - \theta_i) = 0$.

The proof will identify three properties which are necessary and sufficient to characterize the optimal rating system. The first is full insurance, which we have discussed above. The second states that the healthiest type for any rating provided by the system receives no rents (that is, her participation constraint is binding), except for the rating associated with highest average cost. Formally, for a given rating systemσ, we say there are *no rents at the top* if whenever $i = \min\{k : \sigma_{jk} > 0\}$ and $t_j < \max\{t_{j'} : j'\}$, then $t_j = \phi_i$.

Lastly, and perhaps the most noticeable feature of the optimal rating system, is a property we refer to as *negative assortative pooling.* It states that for any two ratings, the second healthiest agent receiving the rating associated with the lower average cost is sicker than any agent in the pool with the higher cost. Namely, if

$t_j < t_{j'}$ and $i > \min\{k : \sigma_{jk} > 0\}$, then $i \geq i'$ for every i' such that $\sigma_{j'i'} > 0$.

Theorem 2. A rating system is optimal if and only if it satisfies no exclusion, no rents at the top, and negative assortative pooling.

The intuition for why the optimal test must satisfy no rents at the top is quite simple. We here provide the intuition for the necessity of negative assortative pooling, resorting to an example with three types (High, Middle, and Low). The details of the general case appear in the appendix. From no rents at the top, we know that the entire population of healthiest agents receive the same rating, henceforth rating A, and rating A has the lowest average cost among all the ratings. Suppose, towards a contradiction, that there some middle types receive rating A and some low types receive rating B. Then, the following re-allocation, depicted in Figure ?, is a Pareto improvement. Namely, create a new rating, henceforth rating C, and move a proportion of middle types from rating A to rating C. The average cost of rating A can either increase or decrease (depending on whether the expected medical cost of the middle type is below or above the average of rating A), and hence we move some low types so that the average cost in rating A does not change. In case the average cost in rating A increases, we can move a fraction of the low types from rating B to rating A. In the case that the average cost in rating A decreases, we can move a portion of the low types from rating A to rating C. Importantly, in either case, the participation constraint of the middle types in rating C slacks, and we can therefore move some of the low types from rating B to rating C so that the average cost among those who receive rating C is between the average costs of ratings A and B without violating any participation constraints. Since the average cost across all ratings does not change, the resulting allocation reduces price dispersion and achieves the same average price which is a welfare improvement.

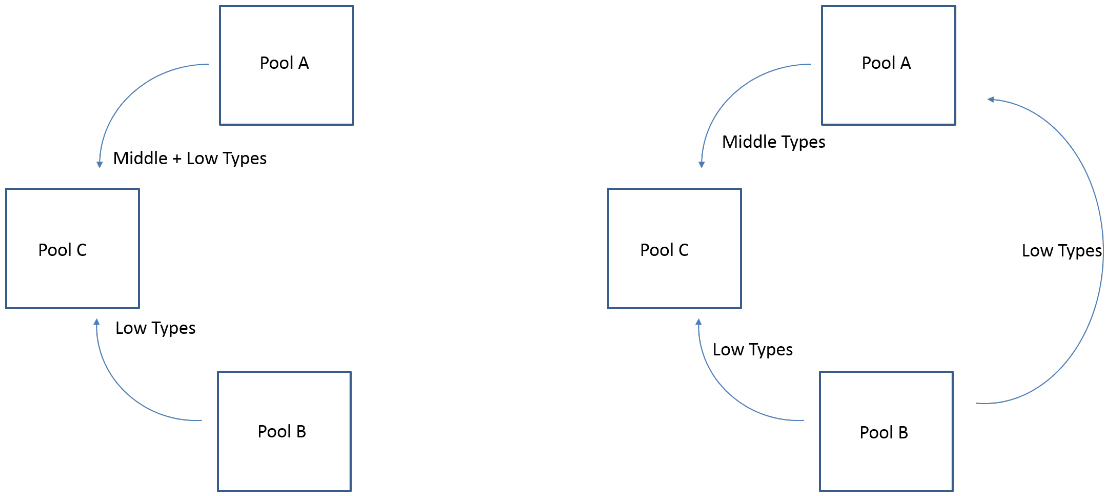

Figure 2: Improving Allocation

3.1 Welfare and Comparative Statics

An important perspective is offered by comparing this algorithm with two commonly used benchmarks: pure community rating (no information) and health-based pricing (full information). Pure community rating is optimal if and only if the participation constraints are not binding under perfect insurance. If (some of) these constraints bind, however, pure community rating does not ensure full trade and is Pareto-dominated by the optimal rating system. Health-based pricing is never optimal since at least some types are willing to pay for insurance more than the actuarially fair price, which allows the regulator to pool resources more efficiently.

In order to better understand the welfare gains that the optimal rating system may yield, as compared with other commonly used pricing systems we use a simple example with three types and CARA utility.

Example 1. The health expenditure of a given individual can be written as $x = \alpha\epsilon_i + (1-\alpha)\epsilon_A$, where ϵ_i is known by the agent and ϵ_A is not. Individual's preferences are represented by a CARA utility function with coefficient of risk-aversionγ. We assume that $\epsilon_i \in \{\epsilon_1, \epsilon_2, \epsilon_3\}$. It follows that $\phi_i = E(x \mid \epsilon_b) + \gamma(1-\alpha)^2 Var(\epsilon_a) = \theta_i + \Delta$.

- If $\Delta > \Delta_1$, no discrimination is needed: $\sigma_{0i} = 1$ for all i with $t_0 = E(\theta)$.

- If $\Delta \in [\Delta_1, \Delta_2)$, the healthiest and sickest types are given a higher rating (together with some middle types):$\sigma_{11} = 1$, $\sigma_{13} = 1$; $\sigma_{02} \in (0,1)$, $\sigma_{12} = 1 - \sigma_{02}$. As a result, $t_1 = \theta_1 + \Delta$, $t_0 = \theta_2$.

- If $\Delta \in [\Delta_2, \Delta_3)$, the healthiest and some of the sickest types are given a higher rating and the remaining types are pooled together:$\sigma_{11} = 1$, $\sigma_{13} > 0$; $\sigma_{02} = 1$ and $\sigma_{03} = 1 - \sigma_{13}$. It follows that $t_1 = \theta_1 + \Delta$, $\theta_2 \leq t_0 = E(\theta \mid s_0) \leq \theta_2 + \Delta$.

- If $\Delta > \Delta_3$, there are three different ratings. $\sigma_{11} = 1$, $\sigma_{22} = 1$ and $\sigma_{13} > 0$, $\sigma_{23} > 0$, $\sigma_{03} = 1 - \sigma_{13} - \sigma_{23} > 0$. It follows that $t_1 = \theta_1 + \Delta$,$t_2 = \theta_2 + \Delta$ and $t_0 = \theta_3$.

Notice that α affects both the level of risk (measured by $\Delta = \gamma(1-\alpha)^2 Var(\epsilon_a)$) and the dispersion of types ($\theta_i = \alpha\epsilon_i + (1-\alpha)\epsilon_a$). In order to calibrate this parameters, we follow Handel et al. (2015)and we assume that ϵ_a is lognormal with mean 6 (in thousands of US \$) and $Var(\epsilon_a) = 60$. We approximate the lognormal distribution of types so that $\epsilon_i \in \{2, 6.85, 23\}$ with probabilities $(0.5, 0.4, 0.1)$ and we choose $\gamma = 0.05$.

The ex-ante welfare for this market as a function of α is depicted in Figure 3. The blue line depicts the certainty equivalent of a randomly chosen individual under the optimal rating system. The green line depicts the certainty equivalent under full community rating (no information). These two lines coincide if $\alpha < 0.3366$ since the allocation then is efficient. The green line has jumps whenever one type leaves the market. The orange line depicts the welfare under full information (health-based pricing). Full information is optimal only if there is no ex-ante information ($\alpha = 0$) or there is no ex-post risk ($\alpha = 1$).

This example also suggests that there is a clear relation between the level of idiosyncratic risk and the efficiency of the market under an optimal rating system. More risky environments increase the wedge between the expected cost and the willingness-to-pay and allow the regulator to pool types more efficiently. Indeed, as we show formally in the Appendix, if we compare two different markets $M_1$and

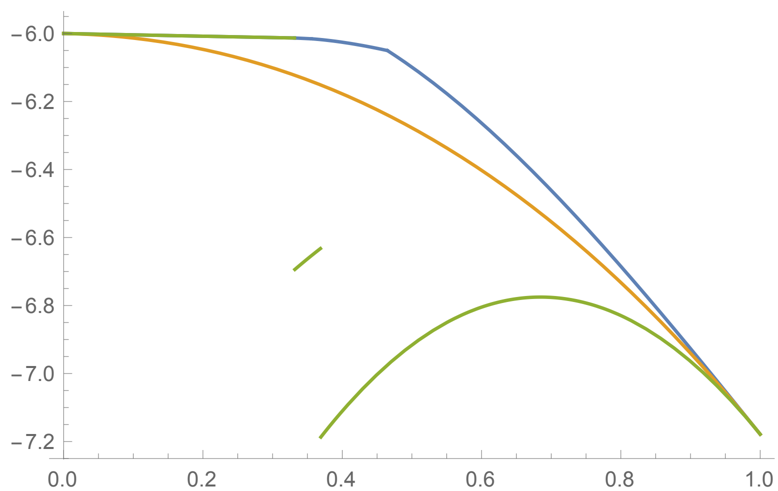

Figure 3: Welfare Comparisons across Regimes

M_2 such that for all types i $f_i^1 \succeq f_i^2$ in the SOSD sense, then the ex-ante expected utility is higher in M_1, although some agents are better-off in M_2.

3.2 (Non)-Monotonic Preferences

Assumption 1 posits that agents with higher expected medical costs are willing to pay more for full coverage. This assumption is typically made for tractability of the Akerlof model (without it, the acceptance sets $A(t)$ are no longer ordered intervals), and also because adverse selection is less severe if preference are not monotonic (because individuals with lower expected medical costs are willing to pay more for insurance). But in some cases of interest preference need not be monotonic. For example, in life insurance markets, individuals who are expected to live longer, are cheaper to insure and are also likely to be more risk-averse (Finkelstein and McGarry (2006)). In asset markets, portfolios with higher returns may have higher variance and if sellers are sufficiently risk-averse their preference relation over them may be non-monotonic.

Fortunately, however, a simple modification of the algorithm presented in Theorem 1 works in this more general environment. Notice that for relatively healthy types (those without rents) what matters is the order of their willingness-to-pay, while for sicker types (those who get rents) what matters is their expected cost. Therefore, we rearrange the algorithm to assign higher consumption to those agents with lower willingness-to-pay for insurance (lower ϕ), regardless of their actual type.

Corollary 1. The following algorithm provides the unique optimal test. Let $l \in \mathbb{N}$ be a counter variable, set $l = 1$, and denote $\mu^l = \mu$. Then:

Step l_1**:** If $E_{\mu^l}(\theta) \leq \min \phi_i : \mu^l > 0 := \phi_{j(l)}$, then set $\sigma_{0i} = 1, \forall i$ and stop. Otherwise, create signal s_l such $\sigma_{lj(l)} = 1$ and $\sigma_{li} > 0$ only if $\forall r > i$, $\sum_{k \leq l} \sigma_{kr} = 1$ and $t_l = E(\theta | s_l) = \phi_{j(l)}$ and move to Step l_2.

Step 2: Stop if there are no individuals remaining in the population. Otherwise, define the prior on the remaining types by

$$\mu_i^{-l} = \frac{\mu_i^{-(l-1)}(1 - \sigma_{li})}{\sum_{k=1}^{N} \mu_k^{-(l-1)}(1 - \sigma_{lk})}$$

increase l by one (that is, $l = l + 1$), and go to step $l_{1.}$.

3.3 Partially Informed Patients

In many situations, it is natural to assume that agents have only some partial information about their own medical risks. For instance, agents may be aware of their pre-existing conditions and perhaps mutations but may not be fully aware of the likelihood of developing certain illnesses and the associated costs. In these cases, it may well be that the regulator, who has access to the history of diagnoses of each patient, can more accurately predict their future expenses and can leverage in this (superior) information to design better rating systems. Two natural questions arise: does the regulator want to make use of this additional information? and, if so, how do market outcomes change as agents become more informed? These questions are particularly relevant in light of regulations like GINA and the recent improvements in genetic testing.

We assume that agents do not know their risk-type precisely but have instead a coarser partition. Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ be a partition on the set Θ and we shall assume monotonicity so that if $\theta_i > \theta_j > \theta_l$ and $i, l \in P_k$ then $j \in P_k$. Since the regulator observes the actual type and, conditional on the type, there is no additional information in the signals, for a given partition $P_1, ..., P_k$ we define a rating system $\sigma \in \Delta^{M \times N}$ as a distribution over public signals $s_1, \ldots, s_M$ that depends on each agent's risk type only so that σ_{ji} is the probability that an agent of type i receives public signal s_j.

Observe that the participation constraints depend on both the private and public signals, since the agent knows that the regulator has access to more information about her type. We let ϕ^k denote the maximal price that an agent is willing to pay for full coverage if she receives the private signal P_k at the interim stage (before she receives the public signal). That is, $u(w - \phi^k) = \sum_{i \in P_k} \tilde{\mu}_i^k U_i$, where $\tilde{\mu}_i^k = \Pr(i|P_k)$. Likewise, ϕ_j^k denotes the maximal price that an agent is will to pay for full coverage after she receives private signal P_k and public signal s_j. That is, $u(w - \phi_j^k) = \sum_{i \in P_k} \Pr(i|s_j, P_k) U_i = \sum_{i \in P_k} \frac{\tilde{\mu}_i^k \sigma_{ji}}{\sum_l \tilde{\mu}_l^k \sigma_{jl}} U_i$. We assume that at the interim stage, preferences are monotonic so that if $k < k'$, $\phi^k < \phi^{k'}$.

The novel feature is that the rating system may reveal new information to the agents. The key question is whether it is optimal to do so. Observe that a public signal that reveals new information, implies $\Pr(i|s_j, \eta_k) \neq \Pr(i|P_k)$ and $\phi^k \neq \phi_j^k$, which increases the dispersion of the outside options, and hence exacerbates adverse selection. But the optimal rating system will pool healthier agents (i.e. those willing to pay less for insurance) with sicker agents (those with high expected medical costs), and it will be more efficient to allocate these sicker agents based on more precise information of the risk type. We will see that the optimal rating system is characterized by three properties: 1) negative assortative pooling; 2) information garbling at the top: an interval of the healthiest agents do not recieve new information about their risk type; and 3) finer information at the bottom: an interval of the sickest agents are allocated to risk pools according to their risk-types.

Proposition 2. The following algorithm yields and optimal test.

Let $l \in \mathbb{N}$ be a counter variable, set the counter $l = 1$, and the prior distribution $\mu^1 = \mu$.

Step l_1: If $E_{\mu^l}(\theta) \leq \phi^l$ then set $\sigma_{0i} = 1 - \sum_{j=1}^{l-1} \sigma_{ji}$, $\forall i$ and stop. Otherwise, create signal s_l with $\sigma_{li} = 1$, $\forall i \in P_l$, and $\sigma_{li} > 0$ for $i \notin P_l$ only if $\forall r > i$, $\sum_{j=1}^{l} \sigma_{jr} = 1$, and such that $t_l = E_{\mu^l}(\theta|s_l) = \phi^l$. Proceed to Step l_2.

Step l_2: Stop if there are no individuals remaining in the population. Otherwise, define the prior on the remaining types by

$$\mu_i^{l+1} = \frac{\mu_i^l(1-\sigma_{li})}{\sum_{r=l}^{N} \mu_r^l(1-\sigma_{lr})},$$

increase l by one (that is, $l = l+1$), and go to step $l_{1.}$.

In words, step l_1 first checks whether there is adverse selection at the interim stage. If there is not, the rating system does not reveal information. If there is averse selection at the interim stage, then the algorithm creates a new risk pool consisting of the entire population of healthiest types which are selected based only by the private signal. To this pool, we add an interval of the sickest types so that the healthiest agents are indifferent. These agents from the bottom are select by their true risk type. The optimal rating system is characterized by negative assortative pooling whereby the agents at top are selected using the most coarse information (their private signal) and agents from the bottom are selected by the most precise information (their true type).

As a result, private information that is more precise induce the planner to reveal more information at the top because (i) the participation constraints are tighter and (ii) the test is measurable with respect to a more informative information structure at the top In the context of design of insurance contracts, it suggests that as more and more insurees obtain genetic information for medical reasons, the amount of information that insurance companies should have access to should also increase.

A more difficult question is what would happen if the regulator does not have access to all the information held by agents. In this case, the regulator may face a trade-off between reducing dispersion in prices (reclassification risk) with more coverage (adverse selection). This trade-off is at the core of earlier work Handel et al. (2015) and is beyond the scope of the present work. We note here, however, that if the regulator is committed to ensure full insurance (for fairness reasons or otherwise), the optimal rating system can be computed using the algorithm in Theorem 1 and substituting the IR constraint of each type for the one corresponding to the most optimistic posterior distribution in the support. Absent this commitment, full insurance may not be optimal as Example 2 illustrates.

Example 2. Patients and the regulator differ in their assessment of individual risks. The regulator has access to the genetic code of the agent and can distinguish$\theta_1 = 1$, $\theta_2 = 3$ both equally likely. Patients do not know their genes but know their behavior and can distinguish two types ($\tilde{\theta}_1 = 1$ and $\tilde{\theta}_2 = 3$) also equally likely. Suppose that both pieces of news are uncorrelated and that $\phi_i = E(x \mid \tilde{\theta}_i, s) + \Delta$.

- Full insurance ex-post does not release additional information. Hence, it is feasible iff $E(x) \leq 1 + \Delta$ or $\Delta \geq 1$
- Suppose there exists a signal, t_2 such that $\sigma_{12} = 0$ and $\sigma_{22} > 0$. Under full-trade it must be true that $t_2 = \theta_2 \leq E(x \mid 1, s_2) + \Delta$ or $E(x \mid 1, s_2) \geq 3 - \Delta$.
  - In such a case, those agents with $\tilde{\theta}_1 = 1$ will trade at $t_1 = E(x \mid \tilde{\theta}_1, s_1) + \Delta$, which is feasible only if $\Delta \geq \frac{1-E(x|\tilde{\theta_1},\theta_1)}{1+\sigma_{12}} + \frac{\sigma_{12}(3-E(x|\tilde{\theta}_1,\theta_2))}{1+\sigma_{12}}$. This pins down σ_{12} under full trade.
  - An alternative is to exclude from trade those types with $\tilde{\theta}_1 = 1$, which is feasible if $\Delta \geq \frac{1-E(x|\tilde{\theta}_2,\theta_1)}{1+\sigma_{12}} + \frac{\sigma_{12}(3-E(x|\tilde{\theta_2},\theta_1))}{1+\sigma_{12}}$. This cross-subsidizes more across genetic pools but reduces welfare through participation.
- If, $E(x \mid 1, s_2) > 3 - \Delta$ full trade is not even feasible.

3.4 Partially Informed Insurers

In many circumstances, the planner (or intermediary) cannot fully control the amount of information available in the market. However, as argued by Bergemann and Morris (2016), it is likely that the regulator has access to that information herself and can, therefore, condition her information release on the realization of these additional sources. To give one simple example, suppose that, as it is the case in the US, insurance companies can offer different contracts in different regions. Since it is common knowledge among market participants that the distribution of types across regions differs, the price offered (for a given signal) also changes. It follows quite naturally, that the regulator should optimally choose a different rating system in each region. The test in each region is then computed using the algorithm given in Theorem 1. Moreover, since the region-by-region algorithm was feasible even if insurers cannot condition in the origin of the patient, but it was generically not optimal, it follows that patients are worse-off when insurers can price-discriminate across regions.

4 Extensions

4.1 Cross-Subsidization

In many markets, regulators can directly intervene in insurance markets through taxes and subsidies. For instance, the Affordable Care Act specifies both a restriction on the information used to price individual policies and a redistributive scheme across policies (the so-called risk-corridor) in order to subsidize disadvantaged groups.[9] This raises a number of questions: should the regulator use both direct intervention and information design? and if so, then does the design of the optimal rating system change?

To answer these questions we introduce the possibility of taxing and subsidizing by the regulator. Fix a rating system σ and an associated set of prices t. The consumption of an individual who receives signal s_j is $c_j = w - t_j(1 + \tau_j) + s_j$, where τ_j is the tax rate on pool j and s_j the per-capita subsidy. In the benchmark model, we had $\tau_j = s_j = 0$ for all j. We shall refer to this case as the No-Taxation (NT) allocation. We assume that redistribution is not perfect so that a certain fraction of total revenue is lost. The budget-balance condition is

$$\alpha \sum_j \sum_i \mu_i \sigma_{ji} t_j \tau_j = \sum_j \sum_i \mu_i \sigma_{ji} s_j$$

[9]The ACA also introduces direct subsidies to policy-holders depending on their income. Since poorer individuals tend to have worse health status, these subsidies can also be interpreted as redistribution across pools.

for some $\alpha \in [0,1]$ which measures the efficiency of the system. [10]Since taxation is distortionary, it follows that if $\tau_j > 0$, $s_j = 0$. Notice also that both full insurance and no rents at the top remain optimal in this setup since the regulator has now more tools. Because for all signals except the worst we have $c_j = w - \phi_j < w - \theta_j$, it is never optimal to choose (σ, τ, s) such that $t_j > \phi_j$ and $s_j > 0$ since it involves excessive taxation and more mass at higher prices. As a result, it is without loss of generality to focus on policy configurations such that $s_j = 0$ for all $j \neq 0$.

A particularly relevant policy configuration uses taxation as the only tool for redistribution. We shall refer to it as the Ramsey (R) allocation. The R allocation has $\sigma_{jj} = 1$ and $\sigma_{ji} = 0$ otherwise for all j such that $w - \phi_j > c_0$ with $(1 + \tau_j)\theta_j = \phi_j$ and

$$\sum \mu_i \sigma_{0i} c_0 = \alpha \sum_{j:\phi_j > c_0} \mu_j (\phi_j - \theta_j) + \sum_i \mu_i \sigma_{0\mathrm{i}} (w - \theta_i).$$

Notice that if $\alpha = 1$, this allocation has the same expected consumption as the NT allocation but since redistribution is costless, it has a less dispersed distribution of consumption. It follows that for $\alpha = 1$, R is optimal. More generally, there exists a range $[\alpha_1, 1]$ such that if $\alpha \in [\alpha_1, 1]$ this allocation remains optimal. To find α_1 it suffices to consider an alternative policy $(\hat{\sigma}, \hat{\tau})$ such that $\hat{\sigma}_{0N} = (1 - \beta)$, $\hat{\sigma}_{kN} = \beta$ for some k such that $w - \phi_k > c_0$ and $\hat{\sigma}_{ji} = \sigma_{ji}$ otherwise. In order to ensure participation of type k, we need to adjust the tax rate so that

$$\frac{\mu_k \theta_k + \beta \mu_N \theta_N}{\mu_k + \beta \mu_{\mathbf{N}}} (1 + \tau_k) = \phi_k.$$

Using this expression we can obtain $T_k(\beta) = \tau_k \sum_i \mu_i \hat{\sigma}_{ki} \theta_i$, the tax revenue coming from pool k under $\hat{\sigma}$. Differentiating around $\beta = 0$, yields the following expression for the change in tax revenue brought about by an increase in the

[10]There is a large literature showcasing inefficiencies associated with the implementation of risk-corridors in Medicare. For a recent review see Geruso ...

fraction of low types who obtain signal k.

$$T_k'(0) \;=\; \frac{d\tau}{d\beta}\mu_k\theta_k + \tau_k\mu_N\theta_N = (1+\tau_k)\mu_k\theta_k - \mu_N\theta_N = \mu_k(\phi_k - \theta_N)$$

We can now compute the consumption of those individuals who obtain the worst rating under $\hat{\sigma}$ as,

$$c_0 \;=\; w - \frac{\sum_i \mu_i\sigma_{0i}\theta_i - \beta\mu_N\theta_N}{\sum_i \mu_i\sigma_i - \beta\mu_N} + \frac{\alpha\sum_{j\neq k}\mu_j(\phi_j-\theta_j) + \alpha T_k(\beta)}{\sum_i \mu_i\sigma_i - \beta\mu_N}$$

It follows that, for β small enough,

$$\begin{aligned}\frac{dc_0}{d\beta} &= \frac{\mu_N}{\sum_i \mu_i\sigma_{0i}}(\theta_N - t_0 + \alpha(\phi_k - \theta_N) - (c_0 - w - t_0)) \\ &= \frac{\mu_N}{\sum_i \mu_i\sigma_{0i}}(\theta_N - (w - c_0) - \alpha(\theta_N - (w - c_k)).\end{aligned}$$

Hence, total welfare improves if and only if

$$u(c_k) - u(c_0) > u'(c_0)\{\alpha(\theta_N - (w - c_k)) - (\theta_N - (w - c_0)))\}.$$

The left-hand side represents the increment in utility for those (few) who jump to a better signal. The right-hand side represents the cost for those who remain evaluated at their marginal utility in the Ramsey allocation. The first term in brackets measures the consumption premium of the sickest type when receiving the worst possible signal and measures the marginal benefit for the remaining types when she gets excluded. The second term represents the drop in the per-capita subsidy needed to ensure that type k still participates. As mentioned above, the R allocation is optimal if $\alpha = 1$, since in that case, the change in consumption is exactly $c_k - c_0$ and, therefore, concavity ensures that it is always true. If $\alpha = 0$, the right-hand side becomes negative since the lowest type is better than the mean type in the worst pool, so it is always true. More generally, this condition holds if and only if it holds for the sickest typek^*such that $w - \phi_{k^*} > c_0$, so if $\alpha < \alpha_1$ implicitly defined as the (smallest) root of the following equation

$$u(c_{k^*}) - u(c_0) = u'(c_0)\{\alpha(\theta_N - (w - c_{k^*})) - (\theta_N - (w - c_0))\}$$

where c_0 depends on α_1 in the Ramsey allocation.[11]

On the other hand, taxation is clearly suboptimal if $\alpha = 0$. Indeed, ifα is low enough, taxation is inefficient and it follows from Theorem 1 that the NT allocation is optimal. In this case, it may not be sufficient to consider simple deviations so a full characterization of the range of parameters for which the NT allocation is optimal is beyond the scope of the present paper. A necessary condition, however, is that a perturbation of the rating system in which a (small) fraction of type l individuals move from pool k to the worst pool must be suboptimal. Formally,

$$u(c_k) - u(c_0) > u'(c_0)\{\alpha(\theta_l - (w - c_k)) - (\theta_l - (w - c_0))\}$$

for all k such that $\phi_k > c_0$ and all $l \neq k$ such that $\sigma_{kl} > 0$. It can be easily seen that this condition holds for some (k, l) if and only if it holds for $k = 1$ and $l = N$, since these two types have the highest dispersion. While formally similar to the earlier expression, these two allocations are different, and therefore, the marginal utility of an agent assigned to the worst signal is higher in the NT allocation. It follows that

$$\alpha_2 \leq \frac{u(c_1) - u(c_0)}{u'(c_0)(\theta_N - \phi_1)} + \frac{\theta_N - t_0}{\theta_N - \phi_1} < \alpha_1.$$

For $\alpha \in [\alpha_1, \alpha_2]$ the regulator would choose to use both instruments simultaneously. In such a case, a simple extension of the arguments we used in Theorem 2 shows that the optimal rating system must have the negative assortative property since the incentives to reduce dispersion in the distribution of prices are the same but there is an additional advantage of reducing the mass at each of the signals inducing low prices. This establishes that the qualitative features of the

[11]There is a unique solution for this equation if the coefficient of absolute risk-aversion is not too high.

optimal rating system remain true even if some direct redistribution is feasible. In particular:

Proposition 3. Suppose the regulator can directly intervene the market with taxes and subsides (τ, s) and a rating system σ and let αbe the efficiency of taxation. Then, there exists $1 > \alpha_1 > \alpha_2 > 0$ such that

1. If $\alpha > \alpha_1$ the Ramsey allocation is optimal and $\sigma_{ji} = 1$ if $j = 1$ and $\phi_j > c_0$while $\sigma_{0i} = 1$ otherwise.

2. If $\alpha < \alpha_2$ the No-Taxation allocation is optimal and $\sigma$is characterized by Theorem 1.

3. If $\alpha \in [\alpha_1, \alpha_2]$, then there exists at least some j such that $\tau_j > 0$, $s_0 > 0$ and σ satisfies the conditions in Lemma 1 and 2.

To better understand these conditions, consider the following Example:

Example 3: Consider the CARA case with a uniform prior over $\Theta = \{0, 2, 4\}$, $\phi_i = \theta_i + 1$

- In the NT allocation, we have that $\sigma_{11} = 1$, $\sigma_{13} = \frac{1}{3}$ and $\sigma_{0i} = 1 - \sigma_{1i}$. As a result, $c_0 = \frac{6}{5}$ and $c_1 = 3$. This is optimal for all $\alpha < \alpha_2 = \frac{1}{15}(11 - 5e^{-\frac{9}{5}}) \approx 0.67$.

- In the Ramsey allocation, we have that $\sigma_{11} = 1$, $\tau_1 = \frac{1}{3}$ and $\sigma_{02} = \sigma_{01} = 1$. As a result, $c_0 = 1 + \frac{\alpha}{2}$ and $c_1 = 3$. This is optimal for all $\alpha > \alpha_1 \approx 0.72$.

- If $\alpha \in [\alpha_2, \alpha_1]$. Then $\sigma_{13} > 0$ and $\tau_1 > 0$.

4.2 Monotonic Tests

A natural concern for policy-makers is that the optimal test is non-monotone, in the sense that healthier types may pay higher premia. If we impose that the allocation induced by the test is monotone, in the sense that the expected utility of healthier types is higher, two properties of the optimal test would fail to hold. First, while it may no longer hold that for those types with $\phi_i < t_0$, $\sum_j \sigma_{ji} u(w - t_j) = u(w - \phi_i)$, it will still hold that for those agents such that $u(w - \phi_i) > \min \sum_j \sigma_{ji} u(w - t_j)$, $\sum_j \sigma_{ji} u(w - t_j) = u(w - \phi_i)$. Second, and perhaps most obvious, since the test cannot treat a sicker types better than healthier types, Lemma B cannot hold. The resulting allocation has the following properties.

Proposition 4. There is an optimal monotonic test with the following features:

1. All types $i = N, N-1, .., l$, with $\phi_l \geq E(\theta \mid \theta \geq \theta_l)$ are treated equally ($\sigma_{ji} = \sigma_{jN}$) have positive probability of receiving any equilibrium price, including $t_0 = E(\theta \mid \theta \geq \theta_l)$.

2. Types $i = l-1, l-2, ..., k$, with $k \geq l+1$, have positive probability of receiving price $t_j = \phi_i$ but also receive some prices $t_k < \phi_i$ with positive probability. Furthermore, $\sum_j \sigma_{ji} u(w - t_j) = \sum_j \sigma_{jN} u(w - t_j)$.

3. Types $i = k, ..., 1$ receive price $t_j = \phi_i$ with probability one and $u(w - \phi_i) > \sum_j \sigma_{j(i+1)} u(w - t_j)$.

A simple example may suffice to illustrate the bite of this restriction.

Example 4: Consider the CARA case with a uniform prior over $\Theta = \{0, 6, 7, 8\}$ with $\phi_1 = 5$, $\phi_2 = 6$ and $\phi_3 = 8$.

- Regardless of ϕ_2, in the allocation implemented by the unconstrained optimal rating, $\sigma_{11} = \sigma_{13} = \sigma_{14} = 1$ and $\sigma_{22} = 1$.

- In the allocation implemented by the constrained optimal rating, $\hat{\sigma}_{11} = 1$, $\hat{\sigma}_{13} = \hat{\sigma}_{14} = 0.997$, $\hat{\sigma}_{12} = 0.011$, $\hat{\sigma}_{22} = 1 - \sigma_{12}$ and $\hat{\sigma}_{33} = \hat{\sigma}_{34} = 1 - \hat{\sigma}_{13}$. This distribution is compound lottery that gives σ with probability 0.989 and with probability 0.011 induces a lottery between $t_1 = 5$ (with probability 0.86) and $t_3 = 7.5$ (with probability 0.14).

5 Conclusion

In this paper we have characterized the optimal information structure in an insurance market under adverse selection and competition among insurers. The combination of adverse selection and risk-aversion implies that the optimal information structure minimizes ex-post risk subject to ensuring full trade. In order to do so, it pools together high and low types as a way to cross-subsidize.

Several extensions seem promising. First, we have chosen the simplest possible market structure. In particular, we have assumed that insurers compete in prices and are bound to offer full coverage. Relaxing competition would introduce an additional margin to the problem of the planner, since different information structures induce different splits of the pie for buyers and sellers. Allowing insurers to offer quantity-price pairs, as in Rothschild and Stiglitz (1976), has the obvious drawback that equilibrium existence is not guaranteed and so the maximization problem of the planner is not always well-defined. A possible way forward is to follow Handel et al. (2015) and focus on a market configuration with two active policies and use Riley equilibrium as solution concept.

Finally, we have assumed throughout that the planner has access to at least as much information as buyers. While this may be true in some setups, it is obviously not true in others. Relaxing this assumption is challenging since the planner may now decide to exclude some types from trade as in Goldstein and Leitner (2015).

References

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Aumann, Robert J and Michael Maschler, *Repeated Games with Incomplete Information*, MIT press, 1995.

Bar-Isaac, Heski, Ian Jewitt, and Clare Leaver, "Multidimensional Asymmetric Information, Adverse Selection, and Efficiency.," 2017.

Bergemann, Dirk and Stephen Morris, "Robust Predictions in Games With Incomplete Information," *Econometrica*, 2013, *81* (4), 1251–1308.

_ **and** _ , "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *American Economic Review*, 2016, *106* (5), 586–91.

Blackwell, David, "Comparison of experiments," in "Proceedings of the second Berkeley symposium on mathematical statistics and probability" The Regents of the University of California 1951.

_ , "Equivalent Comparisons of Experiments," *Ann. Math. Statist.*, 1953, *24* (2), 265–272.

Brown, Jeffrey R and Amy Finkelstein, "Insuring long-term care in the United States," *The Journal of Economic Perspectives*, 2011, *25* (4), 119–141.

Chiappori, Pierre-André, "The welfare effects of predictive medicine," *Competitive Failures in Insurance Markets: Theory and policy implications*, 2006.

Finkelstein, Amy and Kathleen McGarry, "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *The American economic review*, 2006, *96* (4), 938.

Goldstein, Itay and Yaron Leitner, "Stress tests and information disclosure," 2015.

Handel, Ben, Igal Hendel, and Michael D Whinston, "Equilibria in health exchanges: Adverse selection versus reclassification risk," *Econometrica*, 2015, *83* (4), 1261–1313.

Hirshleifer, Jack, "The Private and Social Value of Information and the Reward to Inventive Activity," *The American Economic Review*, 1971, *61* (4), 561–574.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *The American Economic Review*, 2011, *101* (6), 2590–2615.

Levin, Jonathan, "Information and the Market for Lemons," *The RAND Journal of Economics*, 2001, *32* (4), 657–666.

Mitchell, Olivia S, James M Poterba, Mark J Warshawsky, and Jeffrey R Brown, "New evidence on the money's worth of individual annuities," in "American Economic Review" Citeseer 1999.

Roesler, Anne-Katrin and Balazs Szentes, "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, July 2017, *107* (7), 2072–80.

Rothschild, Michael and Joseph Stiglitz, "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information," *The quarterly journal of economics*, 1976, pp. 629–649.

Schlee, Edward E., "The Value of Information in Efficient Risk-Sharing Arrangements," *The American Economic Review*, 2001, *91* (3), 509–524.

A Omitted Proofs

We begin with two useful lemmas that help us derive the characterization result in Theorem 2. We then move to Theorem 1, show that the algorithm formalized within the theorem yields a system with the properties presented in Theorem 2, implying it is optimal, and then show that this is indeed the unique optimal rating system.

Lemma 1. Let σ be an optimal rating system and suppose that $t_j < t_0$ and $\sigma_{ji} > 0$. The following is true:

1. (*No rents at the top*) If $i = \min\{k : \sigma_{jk} > 0\}$, then $t_j = \phi_i$.

2. If $t_j < \phi_i$, then $\theta_i \geq t_0$ (with strict inequality unless $\sigma_{0k} = 0$ for all $k \neq i$).

Proof. We begin with claim (1). Suppose, for a contradiction, that there exists an optimal rating system σ with some signal j and for all types i with $\sigma_{ji} > 0$, $t_j < \phi_i$. Let l be such that $\sigma_{0l} > 0$, and $\theta_l \geq t_0$. We create a new rating system $\hat{\sigma}$ as follows. First, $\hat{\sigma}_{jl} = \sigma_{jl} + (\sigma_{0l} - \hat{\sigma}_{0l}) > \sigma_{jl}$, and $\hat{\sigma}_{0l} = \lambda\sigma_{0l}$ for $0 < \lambda < 1$, such that $\hat{t}_j = E_{j,\hat{\sigma}}(\theta) = \min\{\beta t_0 + (1-\beta)t_j, \phi_i\} > t_j$. That is,

$$\Pr_{\sigma}(s_j)t_j + \mu_l \lambda \sigma_{0l}\theta_l = \big(\Pr_{\sigma}(s_j) + \mu_l\lambda\sigma_{0l}\big)\min\{\beta t_0 + (1-\beta)t_j, \phi_i\},$$

which defines λ as a function of β. For every $1 > \beta > 0$ small enough, this is IR and feasible if σ was optimal. Since $\hat{\sigma}_{0l} < \sigma_{0l}$ and $\theta_l \geq t_0$, $t_j < \hat{t}_j < \hat{t}_0 \leq t_0$, for β small enough. On the other hand, because there is full trade, $E_{\hat{\sigma}}(t) = E_{\sigma}(t)$, so that σ induces a mean-preserving spread of the distribution of prices under $\hat{\sigma}$.

We now proceed to prove claim (2). We proceed by contradiction and assume that $\sigma_{ji} > 0$, $\theta_i \leq t_0$ but $t_j < \phi_i$ and i is not the unique type that has positive probability of receiving the worst signal.

The idea is to create a new signal s_{M+1} with types i (and perhaps some other type) from the support of s_j and some types from the support of s_0, such that $t_{M+1} = \min\{\frac{t_0+t_j}{2}, \phi_i\}$, and t_0 and t_j do not change. There are two cases to consider.

1. If $\theta_i \geq t_j$. Then, moving type i out of the support of s_j leads to an drop in t_j. Hence, we move a representative sample of those types that are pooled in s_0 to substitute the types in the support of s_j we move out (so that t_0 and t_j remain constant). Similarly, we take some additional proportion of this sample types to a new signal s_{M+1} with i which now has $t_{M+1} = \min\{\frac{t_0+t_j}{2}, \phi_i\}$. The result is that t_0 and t_j did not change, while types in s_0 and type j pool their resources more evenly, leading to an ex-ante welfare improvement.

Formally, suppose that $\theta_i \geq t_j$ and consider a test $\hat{\sigma}$ with the following features:

i) $\hat{\sigma}_{0l} = (1-\tau)\sigma_{0l}$ for all $l = 1, \ldots, N$ and $1 > \tau > 0$; ii) $\hat{\sigma}_{jl} = \sigma_{jl} + \alpha\tau\sigma_{0l}$ for all $l \neq i$; iii) $\hat{\sigma}_{ji} = (1-\lambda)\sigma_{ji} + \alpha\tau\sigma_{0i}$ and $\alpha \geq 0$; iv)$\hat{\sigma}_{(M+1)i} = \lambda\sigma_{ji} + (1-\alpha)\tau\sigma_{0i}$; and v) $\hat{\sigma}_{(M+1)l} = (1-\alpha)\tau\sigma_{0l}$ for $l \neq i$ with $\sigma_{kl} = \hat{\sigma}_{kl}$ otherwise. By construction $t_0 = \hat{t}_0$ and,

$$R_{M+1} = \sum \mu_l \hat{\sigma}_{(M+1)l}\theta_l = \mu_i((1-\alpha)\tau\sigma_{0i} + \lambda\sigma_{ji})\theta_i + \sum_{l\neq i} \mu_l(1-\alpha)\tau\sigma_{0l}\theta_l$$

$$= (1-\alpha)\tau \Pr_{\sigma}(s_0)t_0 + \mu_i\lambda\sigma_{ji}\theta_i$$

We choose α, given λ and τ to satisfy

$$R_{M+1} = \big(\Pr_{\sigma}(s_0)(1-\alpha)\tau + \mu_i\lambda\sigma_{ji}\big)\min\{\frac{t_0+\theta_i}{2}, \phi_i\}$$

which guarantees that $\hat{t}_{M+1} = \min\{\frac{t_0+\theta_i}{2}, \phi_i\}$. In particular,

$$1-\alpha = \frac{\mu_i\lambda\sigma_{ji}\max\{\frac{\theta_i-t_0}{2}, \theta_i-\phi_i\}}{\tau\Pr_\sigma(s_0)\min\{\frac{\theta_i-t_0}{2}, \phi_i-t_0\}}.$$

Since $\theta_i \geq t_0$ and $\phi_i \geq t_0$, $1-\alpha \geq 0$. In order to guarantee that $\alpha \geq 0$, it must be the case that

$$\tau > \lambda\frac{\mu_i\lambda\sigma_{ji}\max\{\frac{\theta_i-t_0}{2}, \theta_i-\phi_i\}}{\Pr_\sigma(s_0)\min\{\frac{\theta_i-t_0}{2}, \phi_i-t_0\}}. = \bar{\tau}$$

Similarly, we have

$$R_j = \sum_{l=1}^{N} \mu_l\hat{\sigma}_{jl}\theta_l = \alpha\tau\Pr_{\sigma}(s_0)t_0 + \sum_{l\neq i}\mu_l\sigma_{jl}\theta_l + (1-\lambda)\sigma_{ji}\mu_i\theta_i,$$

where we choose τ such that,

$$R_j = \left(\alpha\tau\Pr_{\sigma}(s_0) + \sum_{l\neq i}\mu_l\sigma_{jl} + (1-\lambda)\sigma_{ji}\mu_i\right)t_j$$

since $\theta_i \geq t_j$, $\sum_{l\neq i}\mu_l\sigma_{jl}(\theta_l - t_j) + \mu_i(1-\lambda)(\theta_i - t_j) \geq 0$ and therefore this determines τ and fixes $\hat{t}_j = t_j$. In particular, define

$$\tau(\lambda) = \frac{\sum_{l\neq i}\mu_l\sigma_{jl}(\theta_l - t_j) + (1-\lambda)\mu_i\sigma_{ji}(\theta_i - t_j)}{\Pr(s_0)(t_j - t_0)} + \lambda\mu_i\sigma_{ji}\bar{\tau},$$

which is non-negative, linear in λ and equals zero when $\lambda = 0$. It follows that both equations (R_j and R_{M+1}) are compatible for all $\lambda \in (0, \bar{\lambda})$ where $\tau(\bar{\lambda}) = \min 0, \bar{\tau}$.

Notice finally that if $t_0 = \theta_i$ but $\sigma_{0l} > 0$ for some $l \neq i$, then $\sigma_{ji} = 0$ for any $j \neq 0$ and we can set $\alpha = 0$.

2. If $\theta_i < t_j$. In this case, moving type i from the support of signal s_j leads to an increase in the price of this account, so we cannot simply replace him by types with positive support on s_0. However, in such a case there must be an additional type l such that $\sigma_{jl} > 0$ with $\theta_l > t_j$ (for otherwise the average cost of a patient in s_j cannot be above θ_i). Let β be such that $\beta\theta_i + (1-\beta)\theta_l = t_j$. Now consider $\hat{\sigma}_{0k} = (1-\tau)\sigma_{0k}$ for $k = 1, \ldots, N$ and some $1 > \tau > 0$, $\hat{\sigma}_{ji} = (1-\lambda)\sigma_{ji}$, $\hat{\sigma}_{jl} = \frac{1-\beta}{\beta}\frac{\mu_i}{\mu_l}\hat{\sigma}_{ji}$, which comes from ensuring that the relative proportions of types l and i that remain in account j are equal to $\beta/(1-\beta)$, so that $t_j = \hat{t}_j$. Similarly,$\sigma_{(M+1)i} = \lambda\sigma_{ji} + \tau\sigma_{0i}$, $\hat{\sigma}_{(M+1)l} = \sigma_{jl} - \frac{1-\beta}{\beta}\frac{\mu_i}{\mu_l}\hat{\sigma}_{ji} + \tau\sigma_{0l}$ which is positive for λ large enough.

As a result, we can take $\hat{t}_{M+1} = \min\{\frac{t_0+t_j}{2}, \phi_i\}$. By construction this is feasible and IR and we have that $\hat{t_0} = t_0$ and $\hat{t_j} = t_j$ for an appropriately pair (λ, τ). The improvement follows since $t_0 = \hat{t_0} > \hat{t_j} > t_j$, both distributions generate the same mean price and $\Pr_\sigma(t_0) > \Pr_{\hat{\sigma}}(t_0)$, $\Pr_\sigma(t_j) > \Pr_{\hat{\sigma}}(t_j)$.

Finally notice that if $\sigma_{0i} > 0$ but $\sigma_{0k} = 0$ for any type $k \neq i$, then the perturbation in Lemma 1 does not produce an improvement because $\hat{t}_{M+1} = t_0$.

QED.

As a result of Lemma 1, for every signal s_j (except s_0), there exists a unique healthiest type for which i) $t_j = \phi_i$, ii) $\sigma_{ji} = 1$ and moreover for all other types k, if $\sigma_{jk} > 0$ for $k \neq i$, then $t_j < \theta_k$. Therefore, without loss of generality, we may relabel the signals (except s_0) such that $t_j = \phi_i$ if and only if $i = j$. For instance, s_1 satisfies, $t_1 = \phi_1$ and $\sigma_{11} = 1$. As a result, the negative-assortative property can be written as follows: if for every type l and signal $j < l$ such that $\sum_{k \leq j} \sigma_{kl} < 1$, then $\sigma_{jl'} = 0$ for all types $j < l' < l$.

Lemma 2. The optimal rating system has the negative-assortative property.

Proof. Assume that the statement is not true, so that there exists a type l and signal $j > l$ such that $\sum_{k \geq j} \sigma_{kl} < 1$, but for some type $l > l' > j$, $\sigma_{jl'} > 0$. Again, it must hold that there exists some other signal $j < j' < l$ (or s_0) such that $\sigma_{j'l} > 0$. It must be true that for any optimal σ,

$$R_j = \sum_{i=1}^{j} \mu_i \sigma_{ji} \theta_i = t_j \sum_{i=1}^{j} \mu_i \sigma_{ji}$$

$$\implies \mu_l \sigma_{jl}(\theta_l - t_j) + \mu_{l'} \sigma_{jl'}(\theta_l - t_j) = \sum_{i \neq l,l'}^{j} \mu_i \sigma_{ji}(t_j - \theta_j) \equiv \Delta$$

Observe that $\Delta > 0$ since $\theta_l > \theta_{l'} > t_j$. Consider then $\hat{\sigma}$such that for all $i \neq l, l'$, and for all k,$\hat{\sigma}_{kl} = \sigma_{kl}$, for all $k \neq j, j'$, $\hat{\sigma}_{kl} = \sigma_{kl}$ and $\hat{\sigma}_{kl'} = \sigma_{kl'}$ and

$$\mu_l \hat{\sigma}_{jl}(\theta_l - t_j) + \mu_{l'} \hat{\sigma}_{jl'}(\theta_{l'} - t_j) = \Delta \iff \mu_{l'} \hat{\sigma}_{jl'} = \frac{\Delta}{\theta_{l'} - t_j} - \frac{\mu_l \hat{\sigma}_{jl}(\theta_l - t_j)}{\theta_{l'} - t_j}$$

Therefore, the joint probability of signal j conditional on types l or l' is higher in σthan in$\hat{\sigma}$ iff

$$\mu_1 \sigma_{j1} + \frac{\Delta}{\theta_{l'} - t_j} - \frac{\mu_l \sigma_{jl}(\theta_l - t_j)}{\theta_{l'} - t_j} < \mu_l \hat{\sigma}_{jl} + \frac{\Delta}{\theta_{l'} - t_j} - \frac{\mu_l \hat{\sigma}_{jl}(\theta_l - t_j)}{\theta_{l'} - t_j}$$

$$\iff \mu_1 \sigma_{jl} \frac{\theta_{l'} - \theta_l}{\theta_{l'} - t_j} < \mu_1 \hat{\sigma}_{N1} \frac{\theta_{l'} - \theta_l}{\theta_{l'} - t_j} \iff \sigma_{jl} < \hat{\sigma}_{jl}$$

since $\theta_{l'} < \theta_l$ but $\theta_{l'} > t_j$. Such a change is feasible by assumption since $\sigma_{jl} > 0$ and $\sigma_{j'l} > 0$. Hence, we have

$$\Pr_\sigma(t_j) = \mu_l \sigma_{jl} + \mu_{l'} \sigma_{jl'} + \sum_{i \neq l,l'}^{j} \mu_i \sigma_{ji} > \mu_l \hat{\sigma}_{jl} + \mu_{l'} \hat{\sigma}_{jl'} + \sum_{i \neq l,l'}^{j} \mu_i \sigma_{jl'} = \Pr_{\hat{\sigma}}(t_j)$$

Thus, following the argument above, the probability on the highest of the two prices (t_j rather than $t_{j'}$) is lower under $\hat{\sigma}$, so that $\hat{t}_{j'} < t_j$, and since $E_\sigma(t) = E_{\hat{\sigma}}(t)$, the distribution of prices underσ is a mean preserving spread of the distribution of prices under $\hat{\sigma}$.

QED

Proof of Theorem 2. By Lemma 1, an optimal test can be characterized by a set of associated prices $(t_1, t_2, ..., t_k, t_0) = (\phi_1, \phi_2, ..., \phi_k, E(\theta|s_0))$ for some $k < N$. By Lemma 2, σ_{ji} is uniquely determined for all signals $j \neq 0$. This establishes that, up to signal s_0, the optimal test is uniquely determined. Now suppose that two different tests σ and $\hat{\sigma}$ are consistent with Lemmas A and B, then it must be that we can write the associated lists of prices for each test as $(t_1, t_2, ..., t_k, t_0)$ and $(t_1, t_2, ..., t_k, \hat{t}_{k+1}, ..., \hat{t}_l, \hat{t}_0)$.[12] Since, by Lemma 2, σ_{ji} is uniquely determined for all signals $j \neq 0$, $\sigma_{ji} = \hat{\sigma}_{ji}$ for all $j < k$ and for all i. This implies that,

$$\Pr_{\hat{\sigma}}(s_0)\hat{t}_0 + \Pr_{\hat{\sigma}}(s_l)\hat{t}_l + ... + \Pr_{\hat{\sigma}}(s_{k-1})\hat{t}_{k+1} = \Pr_{\sigma}(s_0)t_0$$

On the other hand we have

$$\Pr_{\hat{\sigma}}(s_0)\hat{t}_0 + \Pr_{\hat{\sigma}}(s_l)\hat{t}_l + ... + \Pr_{\hat{\sigma}}(s_{k-1})\hat{t}_{k+1} = \Pr_{\hat{\sigma}}(s_0)\hat{t}_0 + \Pr_{\hat{\sigma}}(s_l)\phi_l + ... + \Pr_{\hat{\sigma}}(s_{k-1})\phi_{k+1}$$
$$< \phi_{k+1} \sum_{j \geq k+1} \Pr_{\hat{\sigma}}(s_j) = \phi_{k+1} \Pr_{\sigma}(s_0) \leq \Pr_{\sigma}(s_0)t_0,$$

where the first equality follows by Lemma 1, the inequality follows by the ordering of the willingness-to-pay for insurance and the fact that $\hat{t}_0$ is by definition the highest price and all those signals strictly positive probability, the next follows by the fact that $\sum_{j \geq k} \Pr_{\hat{\sigma}}(s_j) = \sum_{j \geq k} \Pr_{\sigma}(s_j)$, and the last inequality follows from the fact that σ is feasible and IR. Hence, we have a contradiction.

QED

The following lemma will be useful in proving Theorem 1.

Lemma 3 (recursion). If σ is an optimal test for prior $\mu_1, \ldots, \mu_N$, then fix a signal s_j, and consider the test $\sigma^j_{ki} = \frac{\sigma_{ki}}{1-\sigma_{ji}}$, for all $k \neq j$ and $\sigma^j_{ji} = 0$. It must be that σ^j is the optimal test given the prior $(\mu^j_1, \ldots, \mu^j_N)$ where $\mu^j_i = \frac{\mu_j(1-\sigma_{ji})}{\sum_{k=1}^N \mu_k(1-\sigma_{jk})}$.

Proof. Follows directly from the maximization problem. Suppose that a test

[12] It is easy to see that if $l = k$ both tests must be identical because $t_0 = E_{\sigma}(\theta|s_0) = E_{\hat{\sigma}}(\theta|s_0) = \hat{t}_0$. Therefore, wlog, we can take $\hat{\sigma}$to have more signals than σ.

σ is optimal. Given Lemmas A and B either σ implements the first best (in which case it contains a single signal) or there exists some signal s_M with associated probability σ_{Mi}. Since σ is optimal and IR it solves the maximization problem above. Consider the following maximization problem,

$$\max_{\sigma\in\Delta^{(M-1)\times N}} \sum_{i=1}^{N} \mu_i^M \sum_{k\neq j}^{M-1} \sigma_{ki} u(w - t_k)$$

$$t_k = \sum_{i=1}^{N} \frac{\sigma_{ki}\mu_i^M}{\sum_{l=1}^{N} \sigma_{kl}\mu_l^M}\theta_i$$

$$t_k \leq \phi_i, \forall i : \sigma_{ki} > 0$$

$$\sum_{k=1}^{M-1} \sigma_{ki} = 1, \forall i$$

and notice that it is completely independent of s_M, given the new prior μ_i^M. Let σ^M its solution. It follows that $\sigma = (\sigma_M, \sigma^M)$.

QED

Proof of Theorem 1. From Lemma 3, we construct the optimal test by recursively eliminating the signals corresponding for the highest types that remain. We now show that the algorithm is compatible with Lemmas 1 and 2, since there is a unique such test, this will prove optimality. First notice that the algorithm must stop in at most N steps since $\mu_i^l > 0$ only if $i > l$. Furthermore, it holds that that for every price $t_i > t_0$, $t_i = \phi_i$ and $\sigma_{ii} = 1$. This satisfies condition 1 in Lemma 1. On the other hand, t_0 is the price corresponding to the signal created in the last Step (say Step l) because types are ordered by their willingness-to-pay for insurance. Similarly, there must exist at most one type who is both in the support of s_0 and in the support of some other signal s_j, since by construction if $i < l$, $\sigma_{0i} = 0$ and if $i > l$ but $\sigma_{0i} = 0$ then $\sum_{k\leq l} \sigma_{ki'} = 1$ for all $i' > i$. Either this type is then the lowest such type in the support of s_0 or it is the only such type. In any case, this satisfies condition 2 in Lemma 1. Finally, Lemma 2 follows directly from the construction using the lowest available types.

QED

Proof of Proposition 2. We first show that σ must satisfy Lemmas A and B with respect to the interim beliefs it induces. We then show that if this is the case, the optimal allocation has two new properties. First, healthy types' beliefs are fully determined by their initial information (no information at the top). Second, sick types are fully revealed and pooled according to negative assortative matching. We finally show that this algorithm delivers this allocation as its output.

For the first part suppose that σ is optimal and let μ_{il}^{j} be the belief that individuals of type i who observe signal j attach to the event that they belong to type l. Notice that the partitional structure of the prior information implies that i) $\mu_{il}^{j}=0$ if $i \in P_k$ and $l \in P_{k'}$with $k \neq k'$; ii) $\mu_{il}^{j}=\mu_{i'l}^{j}=\mu_{kl}^{j}=\frac{\mu_l \sigma_{jl}}{\sum \mu_l \sigma_{jl}}$ otherwise. Let ϕ_k^j be the willingness-to-pay of an individual who has belief μ_{kl}^{j} that she belongs to type l. Since the principal knows the information that the agent has, it follows naturally that full trade must still hold in any optimal test. Moreover, we can consider only rating systems in which the highest type in the support of two different signals belongs to two different partitions. To see this notice that if there are two different signals j, j' such that $t_j \geq \phi_k^j \geq \phi_{k'}^j$ and $t_{j'} \geq \phi_k^{j'} \geq \phi_{k''}^j$ for some$k$ and for all k' such that $\sigma_{ji}>0$ for $i \in P_{k'}$ and $\sigma_{ji'}>0$ for $i \in P_{k''}$. It follows from Jensen's inequality that an alternative $\hat{\sigma}$is feasible whereby $\hat{\sigma}_{ji}=\hat{\sigma}_{ji}+\hat{\sigma}_{j'i}$ and yields an improvement.

Suppose then that there is an optimal rating system in which some type $l \in P_k$ derives positive rents. That is, there exists some signal j with $t_j<\phi_k^j \leq \phi_{k'}^j$, $\sigma_{jl}^k>0$ and for all k' such that $\sigma_{j'i}^k>0$ with $i \in P_{k'}$. If there is another signal j' such that $\sigma_{j'l}=0$ and $t_{j'}<t_j$ then any mixture $\beta\sigma_{j'i}+\sigma_{ji}$ yields an improvement and is feasible for β small enough since ϕ is independent of β. If there does not exist such a signal, since two signals cannot have their highest type from the same element of the partition, it must then be because $t_j=t_0$. Hence, there are no rents at the top.

We now prove that for every signal j there exists a single partition k such that $\sigma_{ji}>0$ for some $i \in P_k$ and $\theta_j^k>t_0$. Suppose not and let k be such that $\sigma_{ji}>0$ for some $i \in P_k$ and $t_j=\phi_j^{k'}$ but $\theta_j^k>t_0$. If $\sigma_{0i'}>0$ for some $i' \in P_k$, then directly adding $\hat{\sigma}_{0i}=\sigma_{0i}+\sigma_{ji}$ for all $i \in P_k$ is feasible and profitable.

Therefore, we assume that $\sigma_{0i'} = 0$ for all $i \in P_k$. Now we construct $\hat{\sigma}$so as to keep t_j and $t_0$unchanged and we introduce a new signal s_{N+1} such that $t_{N+1} = \phi_j^k$ with some additional types in s_0. But because types in P_k were absent from s_0 $\phi_{N+1}^k = \phi_j^k \geq \phi_{N+1}^{k'}$ for all k' such that $\sigma_{ji'} > 0$ for some $i' \in P_{k'}$ by monotonicity. Hence, this is feasible and induces a mean-preserving contraction.

To see that the negative-assortative property holds for the bottom, simply assume that $\sigma_{ji} > 0$ and $\sum_{l \leq j} \sigma_{li'} < 1$ with $\theta_i < \theta_{i'} < t_0$. If $i \in P_k$ and $i' \in P_{k'}$ this is not optimal by Lemma B. Hence, assume that $i, i' \in P_k$. If in σ there exists some signal j' such that $t_{j'} = \phi_{j'}^k$ then, $\hat{\sigma}$such that $\hat{\sigma}_{j'i} = 1$ for all $i \in P_k$ is an improvement. Otherwise, $t_{j'} > \phi_{j'}^k$ for all $j' \neq j$. But if this is so, the argument used in Lemma B still holds because the change in beliefs does not affect the allocation (the IR constraints are not binding).

Finally, consider the algorithm described. It clearly satisfies Lemmas A and B on the induced interim beliefs since it is an extension of the algorithm in Theorem 2. Second, if it creates a signal j such that $t_j = \phi(\mu_k^j)$, $\sigma_{ji} = 1$ for all $i \in b_k$ and hence $\mu_k^j = \mu_k$. Third, it obviously satisfies full revelation at the bottom. But since the argument proving uniqueness in the allocation in Theorem 1 is still valid with respect to the interim beliefs and interim beliefs are uniquely determined, the allocation is unique. Hence, the algorithm implements the optimal test.

QED

Proof of Proposition 3. Most of the analysis follows from the text. Three things remain to be shown. First, it is without loss of generality to consider schedules whereby $s_j = 0$ for all $j \neq 0$. Second, the Ramsey allocation is optimal if and only if $\alpha < \alpha_1$ so that the deviation considered in the text is the most profitable one. Finally, we need to show that Lemmas A and B continue to hold in this more general environment.

We begin with the first of these statements. Consider a policy (σ, τ, s) with $s_k > 0$,$t_k > \phi_k$ and $\phi_k > c_0$. We know that $\tau_k = 0$ and, therefore,

$$\sum_i \mu_i \sigma_{ki}(\theta_i - s_k) = \sum_i \mu_i \sigma_{ki} \phi_k.$$

Consider now the following alternative scheme $(\hat{\sigma}, \tau, \hat{s})$, with $\hat{\sigma}_{ki} = (1 - \beta)\sigma_{ki}$

for all $i \neq j$ and $\hat{\sigma}_{0i} = \sigma_{0i} + \beta\sigma_{ki}$ for all $i \neq j$ and $\hat{\sigma}_{ji} = \sigma_{ji}$ otherwise. Notice that neither tax rates nor the total taxable base change, so tax revenue and the associated waste is the same in both policies. As a result, any policy that reduces dispersion is beneficial. As before, we show that one can increase $\hat{c}_0$ while keeping $c_k = w - \phi_k$ by redistributing some individuals from k to 0. Subsidies adjust so that

$$\mu_k(\theta_k - \hat{s}_k - \phi_k) + (1-\beta)\sum_{i \neq k} \mu_i\sigma_{ki}(\theta_i - \hat{s}_k - \phi_k) = 0.$$

Notice that $\sum_i \mu_i(\sigma_{ki}s_k - \hat{\sigma}_{ki}\hat{s_k}) = \beta\sum_{i \neq k} \mu_i\sigma_{ki}(\theta_i - \phi_k)$ is the change in subsidies needed to ensure participation in k under the alternative policy. The consumption of the individuals receiving the worst signal

$$\sum_i \mu_i(\sigma_{0i} + \beta\sigma_{ki})(\theta_i - \hat{s}_0 - (w - \hat{c}_0)) = \sum_i \mu_i\sigma_{0i}(\theta_i - s_0 - (w - c_0)) = 0.$$

Simple algebra yields,

$$\sum_i \mu_i\sigma_{0i}(c_0 - \hat{c}_0) + \beta\sum_i \mu_i\sigma_{ki}(\phi_k - (w - c_0)) = 0.$$

Since $c_k > c_0$by definition, the second term is positive. Hence, $c_0 < \hat{c}_0$ and there is a profitable redistribution.

We now show that the Ramsey allocation is optimal if and only if $\alpha \geq \alpha_1$. The only if part is immediate so we only need to prove that if $\alpha < \alpha_1$there is no alternative (σ, τ, s) yielding higher expected surplus. Clearly, replacing θ_N with any other type with $\sigma_{0i} > 0$ is worse because the cost of the redistribution increases without affecting the benefit. Similarly, any deviation involving some $k > k^*$ is inefficient if deviation at $k^*$is inefficient because the left-hand side increases in c_k less than the right-hand side. Two additional deviations need to be checked. First, the second derivative of the value function with respect to β is simply $-u''(c_0)\frac{dc_0}{d\beta} < 0$ so that the problem of choosing the optimal β is concave. Finally, deviations involving more than one type are equivalent (for the R allocation) to deviations of the average type but since it is suboptimal to choose any other type than the worst, the average deviation cannot improve welfare if the deviation with the worst type does not. Hence, the R allocation is optimal iff $\alpha \geq \alpha_1$ as desired.

To see that Lemmas A and B must still hold consider an allocation in which types l,l' with $\theta_l < \theta_{l'}$and signals j and j' such that $\phi_j > \phi_{j'}$ and $\sigma_{jl} > 0$ but $\sigma_{j'l'} > 0$. Consider $\hat{\sigma}$ such that $\hat{\sigma}_{jl} + \hat{\sigma}_{j'l} = \sigma_{jl} + \sigma_{j'l}$ and $\hat{\sigma}_{jl'} + \hat{\sigma}_{j'l'} = \sigma_{jl'} + \sigma_{j'l'}$and $\hat{\sigma}_{kl} = \sigma_{kl}$ otherwise. We pick $\hat{\sigma}$ so that the expected pre-tax following signal j is constant but now $\sum \mu_i \hat{\sigma}_{ji} < \sum_i \mu_i \sigma_{ji}$. If the wasted revenue was constant, this would induce an improvement by Lemmas A and B, but the wasted revenue is actually lower so this must be an strict improvement.

QED

B Online Appendix

Proposition B.1. The following comparative statics hold:

1. Suppose that for every type i, $f_i^1 \prec f_i^2$ in the Second Order Stochastic Dominance sense, and denote by $\sigma(f)$ be the optimal test corresponding with each distribution. Then, $E_{\sigma(f^2)}U \geq E_{\sigma(f^1)}U$ with strict inequality if the allocation is not first best.

2. Let ρ be the Arrow-Pratt coefficient of risk-aversion and let u_1 and u_2 be two risk-averse utility functions. If $\rho_1 > \rho_2$, then for every v, $E_{\sigma^1}V \geq E_{\sigma^2}V$.

Proof of Proposition 2. The first result is straightforward since for every i, $\phi_i^1 < \phi_i^2$ but $\theta_i^1 = \theta_i^2$. In such a case, i) the optimal test under f^1 is feasible under f^2 and if there exists some j such that $t_j < t_0$under f^1, then consider $\hat{\sigma}$ such that $\hat{\sigma}_{ki} = \sigma_{ki}(f^1)$ for all $k \neq j$ and $\hat{\sigma}_{ji} = \sigma_{ji}(f^1) + (1-\tau)\sigma_{oi}(f^1)$ such that $\frac{t_j+\tau t_0}{1+\tau} = \frac{\phi_j^1+\tau t_0}{1+\tau} = \phi_j^2$. Clearly, $\hat{\sigma}$ is feasible and IR under f^2 and improves upon σ. For the second part, simply notice that for every i, $\phi_i^1 \geq \phi_i^2$ (see, e.g. Ross (1981) or Pratt (1970)), so that σ^2 is feasible under u_1. Furthermore, by Theorem 1, the prices under σ^1 are $(\phi_1^1, \phi_2^1, \ldots \phi_j^1, t_0^1)$ while that under σ^2 is $(\phi_1^2, \phi_2^2, ..., \phi_l^2, t_0^2)$, with $j \leq l$ and $t_0^1 \leq t_0^2$. If $j = l$, then it follows naturally that the second distribution is a mean-preserving spread of the first since all the prices are lower except of the highest, and the probability of the highest of the prices is lower under the first, while both have the same mean. Suppose then that $j < l$ and consider $\hat{\sigma}^2$ such that for $\hat{\sigma}_{ki}^2 = \sigma_{ki}^2$ for all $k \leq j$ and $\hat{\sigma}_{0i}^2 = \sum_{k>j} \sigma_{ki}^2$. While $\hat{\sigma}^2$ is not IR according to u^2 it yields clearly a higher expected utility since it bundles together different signals. Now it follows that σ^1 is a mean-preserving contraction of $\hat{\sigma}^2$ since all prices are higher under σ^1 except the highest and the probability attached to the highest of the prices is higher under $\hat{\sigma}^2$ while both have the same mean.

QED

Proof of Proposition 4. Part 3 follows directly by Theorem 1 since for types $i = k+1, ..., N$ monotonicity does not impose any additional restrictions. All these types pay price $t_i = \phi_i$ and get pooled with some lower types. Now fix an allocation and let $\bar{u}_0 = \max_{j \leq k} \sum \sigma_{ji} u(w - t_j)$. Monotonicty requires that

$\bar{u}_0 = u_k$, but optimality requires that $u_k = \min_{j \leq k} \sum \sigma_{ji} u(w - t_j)$. Hence, $u_i = \bar{u}_0$ for all $i \leq k$. Since IR must hold ex-post for every signal, $\sigma_{ji} = 0$ for all $j = i$. This establishes part 2. The equal treatment property for types $i \leq l$ where l is the highest type such that $\sum_{i \leq l} \mu_i \theta_i \leq \sum_{i \leq l} \mu_i \phi_l$ follows from Theorem 1 with the additional monotonicity constraint since the optimal profile has a decreasing expected utility.

QED