CESifo WORKING PAPERS

6853
2018

January 2018

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

Daniel Garcia, Roee Teper, Matan Tsur

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6853*
Category 11: Industrial Organisation

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

Abstract

This paper characterizes the optimal information structure in competitive insurance markets with adverse selection. A regulator assigns ratings to individuals according to their risk characteristics, insurers offer fixed insurance contracts to each rating group, and the market clears as in Akerlof (1970). The optimal rating system minimizes ex-ante risk subject to participation constraints. We prove that in any such market there exists a unique optimal system under which all individuals trade and the ratings match low risk types with high risk types negative assortatively. A simple algorithm yields the optimal system. We examine implications for government regulations of insurance markets.

JEL-Codes: D820.

Keywords: insurance markets, adverse selection, information design.

Daniel Garcia
University of Vienna
Vienna / Austria
daniel.garcia@univie.ac.at

Roee Teper
University of Pittsburgh
Pittsburgh / PA / USA
rteper@pitt.edu

Matan Tsur
University of Vienna
Vienna / Austria
matan.tsur@univie.ac.at

January 4, 2018

1 Introduction

As the combination of big data, artificial intelligence, and scientific innovations in predictive medicine improves the accuracy of risk estimates in insurance markets, a key policy question is how much information should insurers be allowed to use when offering contracts. In health insurance, the Genetic Information Non-discrimination Act of 2008 (GINA) and the Affordable Care Act of 2011 (ACA) have restricted the degree to which insurers can price discriminate based on information about individual risk characteristics, such as genetic mutations and preexisting conditions. The premise of such policies is that equalizing insurance premiums, rather than differentiating them, provides better insurance from an ex-ante perspective. However, providing more accurate information to insurers enables them to offer different contracts to individuals with different risk characteristics, alleviating adverse selection.

The tension between ex-ante insurance and ex-post participation appears to have substantial welfare consequences.[1] In health insurance, for example, individuals with preexisting conditions are turned down or priced out, while healthier and younger individuals opt out of expensive contracts. Since the seminal works of Akerlof (1970) and Hirshleifer (1971), the literature has studied the social value of information in settings with adverse selection or ex-ante insurance, but not with both. In this paper, we focus on this fundamental tradeoff and characterize the optimal information structure in competitive insurance markets with adverse selection.

We consider a market where a population of risk-averse agents buy insurance policies from risk-neutral sellers, as in Akerlof (1970). Each agent has a privately known risk type, which can be understood as a distribution of medical costs. There are no opportunities for signaling or screening in this market. Insurers offer fixed insurance contracts, which cover all medical expenses, and compete over prices. In the ex-ante stage, a regulator designs a rating system, which assigns a public rating to each agent depending on her risk type. Insurers can differentiate agents only by their rating, and thus the rating system determines

[1] See, e.g., Handel, Hendel, and Whinston (2015) for health insurance, Hendel and Lizzeri (2003) for life insurance, and Finkelstein, McGarry, and Sufi (2005) for long-term care insurance markets.

the information structure in the market.

The regulator designs a rating system in order to maximize the expected social welfare given the market structure. We show that in any such market there exists a unique optimal rating system. In the allocation induced by this optimal rating system, all individuals trade and the ratings match low risk types with high risk types negative assortatively. Importantly, these properties require no assumptions on the underlying distribution of risk or on agents' preferences aside from risk aversion. Although our primary focus is on health insurance markets, the analysis and results apply to any competitive market for insurance, including financial markets and labor markets. In contrast to many economic settings where the optimal policy pools similar types together,[2] the optimal information structure in these markets exhibits negative assortative pooling.

Optimal Rating System

In the insurance market described above, each rating induces a posterior distribution over the risk types, which we refer to as a *risk pool*. In equilibrium, each agent will be charged an insurance premium equal to the average medical cost of the participating agents in the same risk pool. Let us consider a simple example to illustrate this. Suppose that each agent 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 neither mutation nor preexisting condition, agents with both mutation and preexisting condition, and so on. If the rating system assigns a unique rating to each risk type, then each agent will pay the actuarially fair price equal to her own expected medical cost. The resulting allocation achieves full participation, but no risk sharing between different types. At the opposite extreme, if everyone receives the same rating, then the healthiest agents will not participate if the average cost in the entire population is too high.

The optimal rating system minimizes ex-ante risk subject to participation constraints. To achieve this, the regulator may decide to assign rating B to agents with a preexisting condition and assign rating A to the rest of the population. The healthiest agents, those who have neither a pre-existing condition nor a

[2]Examples are, inter alia, delegation, strategic persuasion, and costly verification models.

mutation, will participate only if the average medical cost of the agents in rating A is sufficiently low. The regulator can also implement a more diversified risk pool by assigning rating B only to a fraction of the population with a preexisting condition and assigning rating A to the rest of the population, which includes the remaining agents with a preexisting condition.

To simplify the exposition, assume that types can be ordered from healthier to sicker, such that agents of those types with higher expected medical costs are also willing to pay more for full coverage. We say that a subpopulation of agents is an *interval of sickest agents* if the expected medical cost of any agent within the subpopulation is (weakly) greater than the expected cost of any agent outside this subpopulation. Notice that an interval of sickest agents may include multiple risk types. Our main result is that the following algorithm yields the unique optimal rating system:

Step 1. If the average cost in the population is below the willingness-to-pay of the healthiest agent, the entire population receives the same rating and the process is complete. Otherwise, all agents of the healthiest type receive rating R_1. An interval of sickest agents receives the same rating R_1, so that the posterior distribution of the average cost associated with the rating R_1 makes the healthiest type indifferent between buying insurance and not.[3] Proceed to Step 2.

Step 2. The agents that have been rated in Step 1 are removed from the population. If the residual population is empty, the process if complete. Otherwise, the process returns to Step 1.

In other words, each iteration of the algorithm creates a new rating, which pools all agents of the healthiest type among those who have not yet been rated with an interval of sickest agents among those who have also not yet been rated.

The proof shows that three properties are necessary and sufficient to characterize the optimal rating system. The first property is that all agents purchase insurance. To see why this property is necessary, observe that if an agent were to not participate in the market, the regulator could create a new rating that perfectly reveals her type and induce her to trade. Since she was not partici-

[3] Such an interval of sickest agents always exists when the population is large or, equivalently, when allowing for stochastic ratings. The formal description appears in Section 3.

pating before, no other price is affected and the resulting allocation is a Pareto-improvement. The second property, which we refer to as *no rents at the top*, states that the participation constraint of the healthiest type in each risk pool is binding, except perhaps in the risk pool that has the highest average cost. The idea behind this property is that if the participation constraint were not binding for the healthiest type in a given risk pool, we could "move" some of the sickest agents in the worst risk pool to this pool without violating any constraints. Since the average price equals the average cost of participating agents, the resulting allocation achieves a mean preserving contraction of the posterior distribution of prices, which is a welfare improvement.

The third property, referred to as *negative assortative pooling*, is perhaps the most surprising. It states that if two risk types i and j are the healthiest in their respective risk pools with i healthier than j, then any type in j's risk pool is (weakly) healthier than any type in $i's$ risk pool. For example, if type i is the healthiest type in the population, then those agents who are pooled with type i form an interval of sickest agents. The idea behind negative assortative pooling is that it minimizes price dispersion. The proof then shows that the output of the algorithm described above is the unique rating system satisfying all three properties.

We find that the optimal rating system is more informative 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 because there is less idiosyncratic risk. These comparative statics suggest that restricting the use of genetic information is more likely to increase welfare in health insurance, where most agents are covered, than in annuity markets, where only a small fraction of the population participates (see, e.g., Chiappori (2006)). Moreover, our results suggest that as accurate genetic information becomes more widespread, the regulations restricting its use should become less strict.

Related Literature

This work relates to several strands of the literature. Following the seminal work of Hirshleifer (1971), a number of papers have shown that releasing 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 motive for information disclosure: making private information public may reduce adverse selection. On the other hand, there is a small but influential literature on the value of information in competitive markets with adverse selection (Levin (2001); Bar-Isaac, Jewitt, and Leaver (2017)). In these models agents are risk neutral, and hence there is no motive to reduce price dispersion. We analyze insurance markets where both forces are present and fully characterize the optimal information structure.

The papers most related to our work are Handel, Hendel, and Whinston (2015) and Goldstein and Leitner (2015). Handel, Hendel, and Whinston (2015) quantitatively study the effect of price discrimination by simulating health insurance exchanges (markets). They focus on two pricing schemes, no discrimination and perfect discrimination, which in our model correspond to a rating system that reveals no information and complete information, respectively. In their simulations, most of the markets in which insurers cannot price discriminate unravel. The average social welfare in these markets is nonetheless higher than in markets in which insurers can perfectly price discriminate. Our results characterize the constrained efficient discrimination policy. Section 3.1 compares the optimal rating system with these two policies from a welfare perspective.

Goldstein and Leitner (2015) study public information disclosure in financial markets where the motive for trade is to obtain outside liquidity to finance a profitable investment, rather than insurance. The optimal disclosure rule maximizes the volume of trade. The key difference with our setup is that in their model agents are risk neutral, and thus price dispersion does not matter. They first consider the symmetric information case, and show that the optimal disclosure rule is a cutoff rule. In our model, absent private information, it is never optimal to reveal any information. In the private information case, Goldstein and Leitner do not fully characterize the optimal disclosure rule, which need not be unique. Under some restrictions, they show that higher types are matched with lower types, but a subset of low types is always excluded. In our insurance market model, the lowest types are uniquely matched negative assortatively with the highest types. Despite these differences, our view is that these results are complementary and reinforce non-monotonicity as an important feature of optimal information disclosure regardless of the motive for trade.

Finally, with respect to the literature on strategic persuasion and information design (Aumann and Maschler (1995); Kamenica and Gentzkow (2011); Bergemann and Morris (2013, 2016)), our model considers a benevolent social planner who has no conflict of interest with market participants. Technically, the tension between ex-ante insurance and ex-post participation is what makes our analysis novel. In addition, our proofs are constructive and we need not make use of concavification techniques.

2 General Setup

We consider an insurance market consisting of identical risk-neutral insurers and a heterogeneous population of risk-averse agents who are subject to idiosyncratic health risks. Agents in the population are distributed over a finite set of risk types $\Theta = \{1, 2, ..., N\}$ according to the probability distribution μ. A type i agent has a distribution of medical costs $f_i \in \Delta(X)$ where[4] $X \subset \Re^+$. Every agent has a utility function $u : \Re \to \Re$ that is continuous, strictly increasing, and strictly concave. We let $\theta_1, \ldots, \theta_N$ be the expected medical cost of type i, $\theta_i = E_{f_i}(x)$. The types are labeled so that $\theta_N > \theta_{N-1} > \ldots > \theta_1$. Let $U_i = E_{f_i}\big(u(w - x)\big)$ be type i's expected utility, where w is the agent's wealth. The parameter ϕ_i denotes the willingness-to-pay for full insurance of an agent of type i. That is, $u(w - \phi_i) = U_i$.

We assume that the only source of heterogeneity across agents is their medical costs, and therefore agents have the same utility functions and wealth levels (for further discussion see Section 5). We also make the following assumption for expositional reasons:

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

In words, agents with higher expected medical costs are willing to pay more to obtain insurance. In Section 3.3 we relax Assumption 1 and show that all our

[4]An agent's risk type can be derived from a primitive cost function that depends on her profile of individual characteristics.

results remain true.

Information. Each agent knows her individual risk characteristics, and therefore her type. Insurers know only the prior distribution over the types. A regulator knows the risk type of every individual and can design a rating system which reveals public information about agents' expected medical costs.[5]

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

Let $\Delta^{M\times N}$ be the collection of all rating systems. For any rating system $\sigma \in \Delta^{M\times N}$, let $\sigma_{ji} = \Pr_\sigma(s_j \mid i)$ denote the probability of rating s_j conditional on the type being i. In particular, $\sum_{j=1}^{M} \sigma_{ji} = 1$ for all $i = 1, 2, .., N$. The set of available ratings is sufficiently rich so that $\mid S \mid = M > N$. It will be useful to define a *risk pool* associated with a rating $s \in S$ as the posterior distribution of types among the population receiving the rating s. The average expected medical cost of the risk pool associated with a rating s_j is denoted by $E_j(\theta)$. That is, $E_j(\theta) = \sum_{i=1}^{N} \Pr_\sigma(i|s_j)\theta_i$, where

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

Timing. The regulator designs a rating system, and agents privately learn their types. Then, public ratings are realized according to the designed system and agents' types. Lastly, trade occurs, the outcome of the lottery f_i is realized, and consumption takes place.

Trading Process. We adopt Akerlof's market for lemons and focus on fixed insurance contracts that provide full coverage in exchange for a fixed price, or premium. Insurers compete over prices and offer insurance contracts conditional on the information that they observe. While there may be multiple equilibrium prices, we focus on the minimum price that achieves the most efficient allocation.

[5] For example, to compute risk adjustments in Medicare, regulators use detailed information of individuals' medical histories (See, e.g., Geruso and Layton (2017) and the references therein). In Section 4 we extend the analysis to a case where the regulator and the agents do not have the same information.

Hence, the price associated with signal s_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. The equilibrium price of risk pool j equals the expected average cost of agents in the risk pool who are willing to trade at that price. Assumption 1 guarantees that the set of types $A(t_j)$ is an interval and, therefore, the price is well defined.

The regulator's problem. 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. The optimal rating system solves the following problem:

$$\max_{\sigma \in \Delta^{M\times N}} \sum_{i=1}^{N} \mu_i \sum_{j=1}^{M} \sigma_{ji} \Big(u(w - t_j) 1_{t_j \leq \phi_i} + U_i 1_{t_j > \phi_i} \Big)$$
$$s.t. \ t_j = \min_t E_j\big(\theta \mid i \in A(t)\big)$$

In the Appendix we show that it is without loss of generality to consider rating systems with at most $N+1$ signals (Remark A1).[6]

Remark 1. We have assumed that there is asymmetric information and the rating system reveals new information to insurers. An equivalent formulation of the model is that insurers can only offer contracts based on the information provided by the rating system. The key assumption is that the regulator has access to the same information as the market participants. In Section 4 we examine the consequences of relaxing this assumption.

Remark 2. The model applies to any competitive insurance market with adverse selection. In particular, we can relabel the notation to represent an asset market

[6]Although 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 $t = (1+\lambda)E_j\big(\theta \mid i \in A(t)\big)$ provided that $(1+\lambda)\theta_i \leq \phi_i$ for all types i. Note also that since firms obtain no rents, the optimal test does not incorporate the rent-extraction versus rent-creation trade-off, which is at the heart of some recent contributions to the literature on information design (e.g., Roesler and Szentes (2017)).

where risk-averse sellers know the distribution of the asset's return f_i, risk-neutral buyers are initially uninformed, and the regulator can disclose information about the asset's quality.

3 Optimal Rating System

This section characterizes the optimal rating system. The first-best allocation requires full insurance of both risk sources (μ and f). Therefore, the set of Pareto-efficient allocations satisfying ex-ante individual rationality is spanned by a scalar $\pi \geq 0$, which specifies the insurer's profit with the constraint that $u(w - E_\mu(\theta) - \pi) \geq E_\mu U$. In the absence of any public information, our model reduces to Akerlof's market for lemons with risk. The equilibrium price is the minimal price satisfying $t = E_\mu(\theta | i \in A(t))$. 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 agents of the sickest type obtain insurance. At the opposite extreme, a rating system that perfectly reveals the type of every agent induces each agent to trade at the price $t_i = \theta_i$. The expected utility is $\sum_{i=1}^{N} \mu_i u(w - \theta_i)$. This allocation is never Pareto-optimal because it provides no cross-subsidization between risk types, but it may be superior to the allocation without any public information.

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 rating systems are equivalent if they generate the same distribution of prices for all types. It is then without loss of generality to focus on rating systems such that every two signals $s_j, s_{j'}$ induce different risk pools since 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 an outcome of *no exclusion* where all types are insured.

Lemma 0. An optimal rating system satisfies no exclusion.

Proof. By Remark A1 in the Appendix, it is without loss of generality to assume that $\sigma_{Mi} = 0, \forall i$ since at most $N + 1 \leq M$ signals are necessary to implement an optimal allocation. Suppose that $\sigma_{ji} > 0$ and type $i \in \Theta$ does not buy insurance following some signal s_j; then $t_j = \min_t E_j\big(\theta \mid i \in A(t)\big)$ does not depend on σ_{ji}, and we can construct another rating system that strictly improves. The new rating system $\hat{\sigma}$ equals σ except that $\hat{\sigma}_{ji} = 0$ and $\hat{\sigma}_{Mi} = \sigma_{ji}$ and $\hat{\sigma}_{Mk} = 0, \forall k \neq i$. Thus, under $\hat{\sigma}$, type i's expected utility is strictly greater and types $k \neq i$ have the same expected utility. □

It follows that for each signal s_j we can identify an associated equilibrium price

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

and write the regulator's maximization problem as

$$\max_{\sigma \in \Delta^{M \times N}} \sum_{i=1}^{N} \mu_i \sum_{j=1}^{M} \sigma_{ji} u(w - t_j)$$
$$s.t.\ t_j = E_j(\theta)$$
$$t_j \leq \phi_i, \forall i : \sigma_{ji} > 0$$

The regulator chooses a rating system to maximize the ex-ante expected utility, subject to the ex-post participation and break-even constraints. The break-even constraint is in the spirit of the Bayes-neutrality condition of Bayesian Persuasion models 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 (with respect to the Bayesian Persuasion literature) and it arises naturally in insurance 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).[7] Our main result shows that there exists a unique optimal rating system that is the outcome of a simple algorithm.

[7]Notice that the participation constraints define a closed set since they can be rewritten as $\sigma_{ji}(t_j - \phi_i) \leq 0$.

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

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

Informally, each iteration of the algorithm assigns a new rating only to the agents who have not yet been rated (the remaining population). All agents of the healthiest type in the remaining population receive this rating. An interval of sickest agents[8] in the remaining population also receive this rating, up to the point where either the price equals the willingness-to-pay of the remaining healthiest type or the remaining population is exhausted.

Figure 1 depicts the outcome of this algorithm for a case with five types. The first best is not feasible because $E(\theta) > \phi_1$. The optimal rating system pools together types $\{1, 5\}$ and a fraction of type 4 in such a way that type 1 is indifferent (i.e., $t_1 = \phi_1$), and the insurer's expected profit from type 1 agents equals the expected loss from the interval of sickest agents:

$$\mu_1(\phi_1 - \theta_1) = \underbrace{\mu_5(\theta_5 - \phi_1)}_{A_5} + \underbrace{\mu_4\sigma_{14}(\theta_4 - \phi_1)}_{A_4}.$$

The distribution of the residual types is given by

$$\mu^2 = \frac{1}{\mu_2 + \mu_3 + (1 - \sigma_{14})\mu_4}(0, \mu_2, \mu_3, (1 - \sigma_{14})\mu_4, 0),$$

[8]Recall that a subpopulation of agents is an interval of sickest agents if the expected medical cost of any agent within the subpopulation is (weakly) greater than the expected cost of any agent outside this subpopulation.

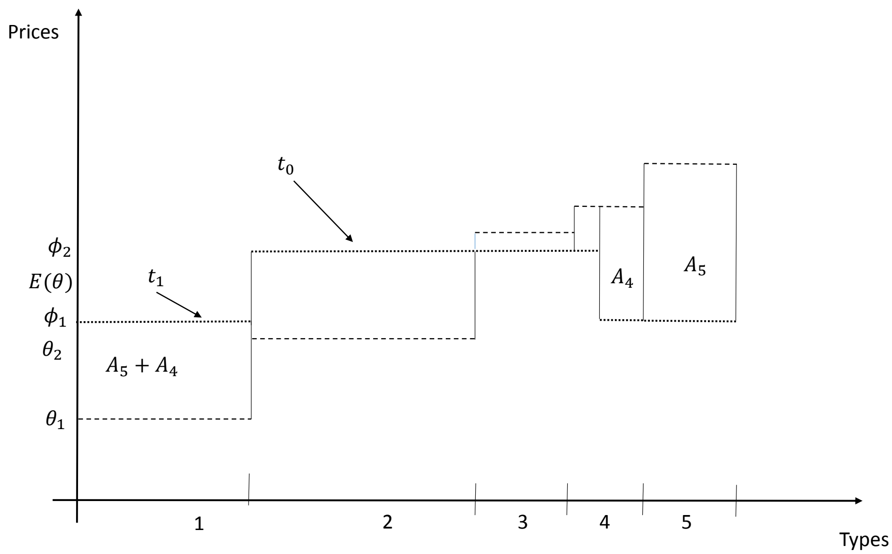

Figure 1: The optimal rating system

and it induces an expected cost of $\sum_{=1}^{5} \mu_i^2 \theta_i < \phi_2$. Therefore, the algorithm assigns rating s_0 to the remaining types with price $t_0 = \sum_{=1}^{5} \mu_i^2 \theta_i$.

The proof of Theorem 1 identifies three necessary properties, that are jointly sufficient to characterize the optimal rating system. The first property is *no exclusion,* which was discussed above. The second property states that the participation constraint of the healthiest type receiving a certain rating with positive probability is binding, except for the rating associated with the highest average cost. Formally, we say that a rating system σ satisfies *no rents at the top* if whenever $i = \min\{k : \sigma_{jk} > 0\}$ and $t_j < \max_{j'} t_{j'}$, then $t_j = \phi_i$.

The third property is *negative assortative pooling.* It states that if types $i, l \in \Theta$ are the healthiest types in their respective pools, and type i is healthier than type l, i.e., $i < l$, then any agent pooled with i is (weakly) sicker than any agent pooled with l. Formally, if there are two signals s_j and $s_{j'}$ such that $i = \min\{k : \sigma_{jk} > 0\} < \min\{k : \sigma_{j'k} > 0\} = l$, then $\min\{k \neq i : \sigma_{jk} > 0\} \geq \max\{k : \sigma_{j'k} > 0\}$.

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

The proof is given in the Appendix. We first explain the intuition for why these properties are necessary. We begin with no rents at the top. Suppose that there exists a rating with an associated risk pool whose average cost lies strictly below the willingness-to-pay of the healthiest agent in that pool. Since the participation constraints in this risk pool are not binding, we can "move" some of the the sickest agents from a risk pool that has a higher average cost to this risk pool without violating any constraint. The resulting allocation is welfare-improving because the price dispersion between these pools is reduced and the average price across these pools does not change, as the average price equals the average medical cost of the participating agents.

To see why negative assortative pooling is necessary, consider a case with three types $\{1, 2, 3\}$. The details of the general case are in the Appendix. It follows from no rents at the top that all agents of type 1 receive the same rating, henceforth rating A, and rating A has the lowest average cost of all the ratings. Suppose, towards a contradiction, that some agents of type 2 receive rating A and some agents of type 3 receive rating B. Then there exists a welfare improving rating system. Namely, create a new rating, henceforth rating C, and "move" some agents of type 2 from rating A to rating C. Since the average cost of rating A can either increase or decrease, we "move" some agents of type 3 to keep it constant. In the case where the average cost of rating A increases, we can subtract agents of type 3 from rating B and add them to rating A, as depicted in the right panel of Figure 2. In the case where the average cost of rating A decreases, we can subtract agents of type 3 from rating A and add them to rating C, as depicted in the left panel of Figure 2. Importantly, in either case, the average cost of rating C lies between that of ratings A and B, and the participation constraints of the agents that receive rating C are not binding. Therefore, we can "move" agents of type 3 from rating B to rating C without violating any participation constraints. The resulting allocation achieves the same average price and reduces price dispersion.

The final step shows that there exists a unique rating system that satisfies the three properties. To see this consider two potential candidates σ and $\hat{\sigma}$

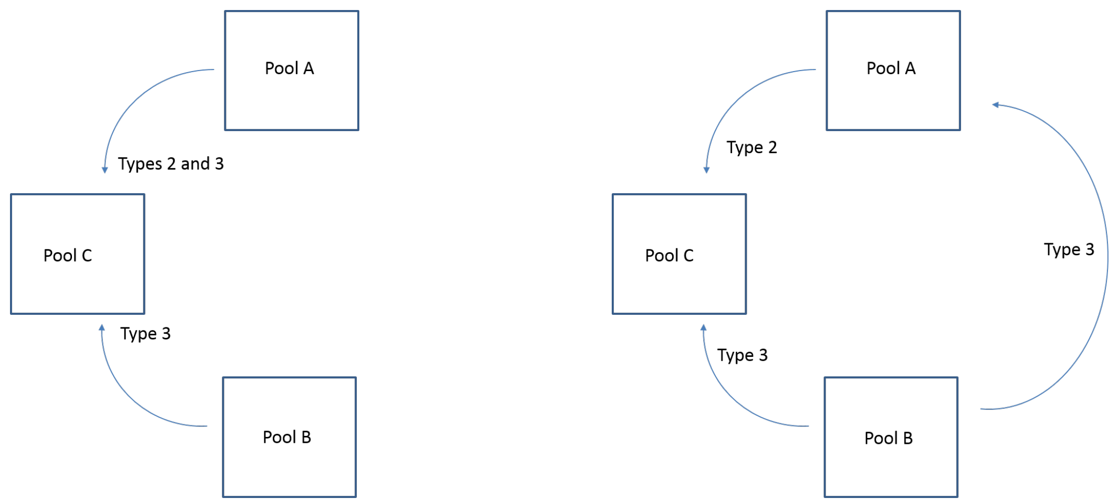

Figure 2: Improving Allocation

with associated number of signals M and $\hat{M} \geq M$. Since no agent is excluded, both rating systems achieve an allocation with the same average price. By no rents at the top, the prices associated with all signals except the worst equal the willingness-to-pay of a type. Hence, the lowest $M-1$ prices induced by each of the two rating systems coincide. By negative assortative pooling the probabilities of each of those $M-1$ signals must be the same in both rating systems. It follows that the price of the worst signal in σ equals the average price among the worst $\hat{M}-M+1$ signals in $\hat{\sigma}$, and that all types in the support of these signals are willing to trade at such a price. But then $\hat{\sigma}$ either violates no rents at the top or negative assortative pooling. It is straightforward to check that the output of our algorithm satisfies the three properties.

3.1 Welfare and Comparative Statics

In a recent paper, Handel, Hendel, and Whinston (2015) estimate the social welfare in health insurance exchanges under two pricing systems. Under a system of pure community rating, insurers are not allowed to price any individual risk characteristics, which constitutes the no information benchmark of our model.[9] Under a system of health-based pricing, insurers price the individual risk charac-

[9]In the exchanges set up by the Affordable Care Act, prices are determined by adjusted community ratings that depend only on age and smoking.

teristics, which is equivalent to the full-information benchmark in our model.

An important perspective emerges when we compare these systems to the optimal rating system. Pure community rating is optimal only if it achieves full trade and is otherwise Pareto dominated by the optimal rating system. Health-based pricing is never optimal because the regulator can pool individuals more efficiently by cross-subsidizing between risk types. In order to better understand the welfare gains of the optimal rating system, we use a simple example with three types and CARA utility.

Example 1. Following Handel, Hendel, and Whinston (2015), we assume that the health expenditure is given by $x = \alpha\epsilon_i + (1-\alpha)\epsilon_A$, where ϵ_i is known by the agent and ϵ_A is not. We assume that $\epsilon_i \in \{\epsilon_1, \epsilon_2, \epsilon_3\}$, and preferences are represented by a CARA utility function with a coefficient of risk-aversion γ. Therefore, $\theta_i = E(x|\epsilon_i)$ and $\phi_i = E(x \mid \epsilon_i) + \gamma(1-\alpha)^2 Var(\epsilon_a) \equiv \theta_i + \Delta$.

Thus, under health-based pricing, each type pays her actuarially fair price $t_i = \theta_i$. Under pure community rating, we either have full participation at price $E_\mu(\theta)$ (iff $\phi_1 \geq E_\mu(\theta)$); or partial unraveling, which occurs when only types 2 and 3 purchase insurance at price $E_\mu(\theta|i \neq 1)$; or complete unraveling, which occurs when only type 3 purchases insurance at price θ_3. Under the optimal rating system, all types receive the same rating if and only if $\phi_1 = \theta_1 + \Delta \geq E_\mu(\theta)$. Otherwise, there are 3 regions given by $\Delta_1 < \Delta_2 < \Delta_3$:

- If $\Delta \in [\Delta_1, \Delta_2)$, there are two ratings: the healthiest types, the sickest types, and some of the middle types receive rating s_1 and pay $t_1 = \theta_1 + \Delta$; the rest of the middle types receive rating s_0 and pay $t_0 = \theta_2$.

- If $\Delta \in [\Delta_2, \Delta_3)$, there are two ratings: the healthiest types and some of the sickest types receive rating s_1 and pay $t_1 = \theta_1 + \Delta$; and the remaining types receive signal s_0 and pay $t_0 = E(\theta \mid s_0)$ where $\theta_2 \leq t_0 < \theta_2 + \Delta$.

- If $\Delta \geq \Delta_3$, there are three different ratings: the healthiest types and some of the sickest types receive rating s_1 and pay $t_1 = \theta_1 + \Delta$; the middle types and some of the low types receive signal s_2 and pay $t_2 = \theta_2 + \Delta$; and the rest of the low types receive signal s_0 and pay $t_0 = \theta_3$.

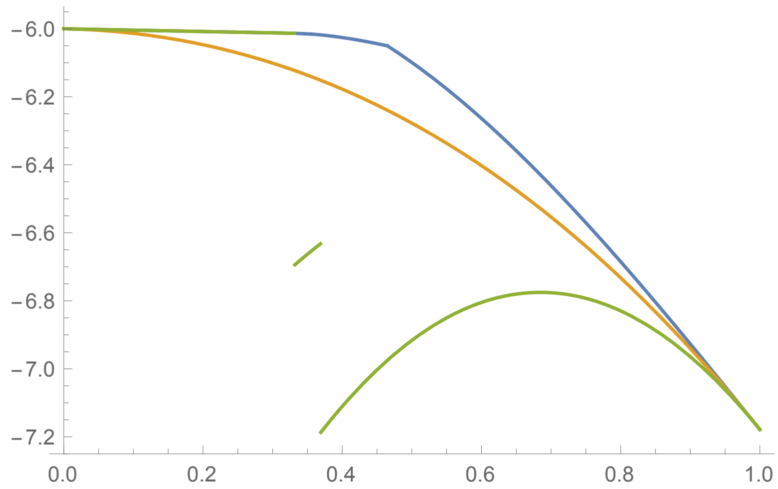

Figure 3: Welfare Comparisons across Regimes

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 these parameters, we again follow Handel, Hendel, and Whinston (2015) and assume that ϵ_a is log-normal with mean 6 (in thousands of US \$) and $Var(\epsilon_a) = 60$ and $\gamma = 0.05$. We approximate the log-normal distribution of types so that $\epsilon_i \in \{2, 6.85, 23\}$ with probabilities $(0.5, 0.4, 0.1)$.

The welfare for this market as a function of α is depicted in Figure 3. The blue line depicts the ex-ante certainty equivalent under the optimal rating system. The green line depicts the (ex-ante) certainty equivalent under pure community rating. The blue and green lines coincide if $\alpha < 0.33$ since both regimes implement the efficient allocation. As α increases, Δ decreases and the market under pure community rating unravels, as represented by the downward jumps in the green line. The orange line depicts the certainty equivalent under 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, if we compare two different markets M_1 and M_2 such that for all types $i \in \Theta$, f_i^2 is a mean preserving spread of f_i^1, then the ex-ante expected utility is higher in M_2.[10]

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, which is typically made for tractability, implies that the average cost of a given risk pool is always decreasing in participation (formally, $A(t)$ is always an interval). But in some cases of interest Assumption 1 may fail. For example, in life insurance markets, individuals who are expected to live longer are likely to be more risk-averse (Finkelstein and McGarry (2006)). In asset markets, portfolios with higher returns may have higher variance and if the sellers are sufficiently risk-averse, the preference relation may be non-monotonic.

Fortunately, a simple modification of the algorithm presented in Theorem 1 yields the optimal rating system in this more general environment. That is, the algorithm assigns higher consumption to those agents with lower willingness-to-pay for insurance (lower ϕ), regardless of their medical costs.

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

Step l_1. Let $j(l) \in \Theta$ be the type with the lowest willingness-to-pay in the support of μ^l, i.e., $\phi_{j(l)} = \min\{\phi_i : \mu_i^l > 0\}$. If $E_{\mu^l}(\theta) \leq \phi_{j(l)}$, then set $\sigma_{0i} = 1 - \sum_{j=1}^{l-1} \sigma_{ji}, \forall i$ and stop. Otherwise, create signal s_l such that $t_l = E(\theta|s_l) = \phi_{j(l)}$, where $\sigma_{lj(l)} = 1$ and $\sigma_{li} > 0$ only if $\forall r > i$, $\sum_{k \leq l} \sigma_{kr} = 1$. 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

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

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

[10]Observe that the optimal test under f^1 is feasible under f^2 because $\phi_i^1 > \phi_i^2$ and $\theta_i^1 = \theta_i^2$ for all i.

In each iteration, if the average medical cost of the population that has not yet been rated is below the willingness-to-pay of each agent that has not yet been rated, then the remaining types are pooled in the same rating. Otherwise, the agents that are willing to pay the least for insurance are pooled with the agents that have the highest expected medical costs (within the remaining population), so that the participation constraints of the former are binding. In the resulting allocation, an interval of agents with the lowest willingness-to-pay receive no rents, and an interval of the agents with the highest expected medical costs receive rents.

4 Informational (Dis)Advantage

We have characterized the optimal rating system under the assumptions that the regulator has access to the same information as market participants. This section explores the results of relaxing this assumption. We begin with the case where the regulator has more precise information than the agents. This information asymmetry may arise, for example, because the regulator can more accurately estimate the likelihood of developing certain illnesses or can better predict the costs that are associated with them. We will show that the optimal rating system is still characterized by the same properties, but the regulator can leverage his superior information to design better rating systems.

We extend our model by assuming that agents do not know their risk-type precisely, but instead have a coarser partition. Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ be a partition on the set Θ. The partition is monotonic so that if $\theta_i > \theta_j > \theta_l$ and $i, l \in P_k$, then $j \in P_k$. We denote by ϕ^k the maximal price that an agent is willing to pay for full coverage if she receives only the private signal P_k. That is, $u(w - \phi^k) = \sum_{i \in P_k} \tilde{\mu}_i^k U_i$, where $\tilde{\mu}_i^k = \Pr(i|P_k)$. The preferences are monotonic so that $k < k'$ if and only if $\phi^k < \phi^{k'}$.

The regulator can directly observe the type. Given a partition $\mathcal{P}$, we define a rating system $\sigma \in \Delta^{M \times N}$ as a distribution over public ratings $s_1, \ldots, s_M$ that depends only on each agent's risk type, so that σ_{ji} is the probability that an agent of type i receives public signal s_j. Let ϕ_j^k denote the maximal price that an agent is willing to pay for full coverage after she receives private signal P_k and public signal s_j. That is,

$$u(w-\phi_j^k) = \sum_{i\in P_k} \Pr(i|s_j, P_k)U_i = \sum_{i\in P_k} \frac{\tilde{\mu}_i^k \sigma_{ji}}{\sum_{l\in P_k} \tilde{\mu}_l^k \sigma_{jl}} U_i.$$

Proposition 1. The following algorithm yields an optimal rating system. Let $l \in \mathbb{N}$ be a counter variable and set the counter to $l = 1$, and $\mu^1 = \mu$.

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

In other words, if the average cost in the population that has not been rated is below the willingness-to-pay of the agents that have not been rated, then the rating system does not reveal any more information. Otherwise, the agents in the cell of the partition with the lowest willingness-to-pay are pooled with an interval of sickest agents, so that the healthiest type is indifferent. The proof is given in Appendix B. It applies the same arguments presented in the proof of Theorem 1.

Observe that this algorithm achieves negative assortative pooling where the agents at the top are selected by the coarser information (their element of the partition), while the agents at the bottom are selected by the finer information (their true type). The intuition is that a public rating that reveals new information increases the dispersion of the outside options. At the same time, negative assortative pooling is most efficient when the types at the bottom, whose participation constraints are not binding, are selected using the most precise information.

As a result, as agents possess finer information about their risk characteristics, the regulator designs a rating system that reveals more information. This is due to two reasons. First, Jensen's inequality implies that participation constraints are tighter, in the sense that the total rents that can be extracted from any interval of healthiest agents is lower. Second, mechanically, if there are more types, the regulator uses more signals. In the context of health insurance contracts, this result suggests that as more and more insurees obtain genetic information for medical reasons, insurance companies should be able to price discriminate based on richer information.

A more difficult problem is presented if the regulator has access to a coarser information structure than the agents do. The key issue here is that the regulator may have to promise informational rents in order to satisfy the agents' participation constraints, and thus it may be optimal to exclude some of the types (see Example B.1 in the Online Appendix for an illustration). Nevertheless, for the case where informational asymmetries are not too severe, in the sense that the regulator could implement full trade by revealing her information, it can be shown that the optimal rating system satisfies no rents at the top (with respect to the information held by the agents) and negative assortative pooling (with respect to the information held by the regulator).

To illustrate this, let us consider a situation where the regulator perfectly observes the expected medical costs of each agent, but not their willingness-to-pay for insurance. That is, an agent's type is a pair (i, ϕ) so that the expected medical cost is θ_i and the willingness-to-pay for full coverage ϕ is distributed according to G_i. The regulator knows θ_i and the prior distribution of ϕ conditional on i. Each agent knows θ_i and the realization of ϕ. A rating system is a probability distribution over a set of ratings $S = (s_0, ..., s_M)$ conditional only on the realization of the average cost θ_i. To place some structure, let us assume that agents can be ordered so that agents with higher expected medical costs are willing to pay more for insurance. That is, the willingness-to-pay of different cost types do not overlap, so that if $G_{i+1}(x) > 0$, then $G_i(x) = 1$.[11]

This information asymmetry may arise, for example, if agents have private information about the variance of the distribution of medical costs or about their preferences. As we have argued above, since there is private information, the optimal rating system need not satisfy the property of no exclusion. However,

[11]This is consistent with Assumption 1 since $\theta_{i+1} > \theta_i$.

as long as the regulator's objective function continues to exhibit a preference for mean preserving contractions of the price distribution (as would be the case if the heterogeneity concerns only the variance of the distribution), then the optimal rating system satisfies the other two properties of Theorem 2.

In this setting, we say that a rating system satisfies *no rents at the top** if whenever $\theta_i = \min\{\theta_l : \sigma_{jl} > 0\}$ and $t_j < \max\{t_{j'} : j'\}$, then $t_j \in supp(G_i)$.

Corollary 2. The optimal rating system satisfies no rents at the top* and negative assortative pooling.

The proof is given in Appendix B. It follows the same line of arguments used to prove Theorem 2.

Finally, in some markets regulators may not be able to fully control the information used to price contracts. For example, it may be that insurance companies can offer different contracts to individuals in different regions. In such a case, if the distribution of medical risks across regions differs, then individuals in healthier regions will be offered cheaper insurance contracts. Even if the regulator cannot eliminate regionally-based discrimination, he is still likely to have access to the information that insurers use to cherry pick their customers. It follows naturally that the regulator should optimally choose a different rating system in each region. The rating system in each region is computed using the algorithm given in Theorem 1. Importantly, since the region-by-region algorithm is feasible in the case where insurers cannot cherry pick, but is generically not optimal, it follows that agents are worse off under regionally-based discrimination.

5 Discussion

We have analyzed the problem of a regulator that can provide information to insurers and this information enables them to offer different contracts to individuals with different risk characteristics. We showed that the optimal rating system is uniquely characterized by three properties: no exclusion, no rents at the top, and negative assortative pooling. The key assumption is that the regulator has access to the same information as the agents. In view of the recent scientific innovations in predictive medicine and machine learning techniques, we think that this assumption is reasonable in most insurance markets. In various setting, we have argued that the optimal rating system still satisfies no rents at the top and negative assortative pooling, even though it need not induce full trade. We briefly discuss a number of possible extensions of the model.

Preferences and Welfare. We have assumed that (i) individuals are risk-averse expected-utility maximizers, and (ii) the regulator maximizes the utilitarian social welfare function with Pareto weights given by the prior distribution. The optimal rating system we have characterized remains optimal under more general models accommodating risk-aversion and a large class of social welfare functions.

To see this, notice first that the key feature of the model is that each agent's willingness to pay is greater than her average cost, $\phi_i > \theta_i$, which is true under any definition of risk aversion. Second, a rating system that does not achieve full trade is Pareto dominated, and thus the optimal rating system under any social welfare function satisfies no exclusion (Lemma 0). Third, the proof of Theorem 1 takes any rating system satisfying no exclusion and applies a sequence of perturbations yielding the unique rating system constructed by our algorithm. Each test perturbation is a mean preserving contraction of the distribution of consumption profiles. Therefore, the proof holds true for any social welfare function that respects SOSD. For example, maximin (Rawls (2009)) and leximin (Sen (1977)) social preferences, which put all their weight on the worse-off members of society, and the quadratic social welfare function (Epstein and Segal (1992)), which maximizes a mean-variance value function of the interim utilities, all respect SOSD.

Heterogeneity. We have assumed that the only source of heterogeneity is the distribution of medical costs. The assumption implies that a rating system that induces full trade at lower price dispersion is welfare-improving. Without it, the social planner need not have a preference for mean preserving contractions of the price distribution, and consequently a general characterization of the optimal rating system is not possible. For example, suppose that each agent is willing to trade at a price equal to the average medical cost in the population, $t = E_\mu(\theta)$. A rating system that creates a single risk pool induces full trade, but this need not be optimal if the marginal utility from consumption depends also on agents' wealth levels.[12] In practice, health insurance policy is often used as a safety net with the (implicit) aim of reducing inequality and poverty (see, e.g., Finkelstein, Hendren, and Shepard (2017)), but the additional trade-offs involved are beyond the scope of the present paper.

Market Structure. We have assumed that insurers offer fixed insurance contracts and compete over prices, while agents face an all-or-nothing decision. On the supply side, if insurers can offer quantity-price pairs, as in Rothschild and Stiglitz (1976), the only equilibrium outcome, when it exists, requires full separation of types, and so no further cross-subsidization is feasible. A possible way forward is to follow Handel, Hendel, and Whinston (2015) and focus on a market configuration with two active policies, using Riley equilibrium as the solution concept. Relaxing competition would introduce an additional dimension to the problem of the regulator, since different information structures induce different splits of the pie for buyers and sellers. On the demand side, a natural extension is to allow agents to buy partial insurance. Partial insurance adds a new trade-off for the regulator because increasing the mass of high-cost agents in a pool reduces the participation of low-cost agents on the extensive margin. It is easy to see that some pooling remains optimal but a full characterization of the optimal information structure is left for future work.

Taxes and Subsidies. We have analyzed the problem of a regulator that can influence the market outcomes only through information design. There are, of course, a range of more direct policy interventions. The Affordable Care Act, for example, specifies a broad redistributive scheme across contract pools (the so-called risk-corridor), compensating insurers with excessive costs.[13] A natural question to ask is how to optimally combine information design and fiscal policies.

More specifically, suppose that, as in our model, the regulator designs a rating system that determines the composition of the risk pools, and then the price of each risk pool is determined by market competition (it equals the average cost

[12] Observe that a rating system that creates diverse risk pools consisting of healthy rich agents and poor sick agents may better equalize the consumption profile.

[13] 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.

of the agents participating in the risk pool). In addition, the regulator sets a tax rate and a subsidy for each of the risk pools, with the constraint that the policy should be budget-balanced. For instance, the regulator may choose a rating system that perfectly reveals each type and set up taxes and subsidies so as to smooth consumption subject to the participation constraint. The optimal tax scheme in this case is given by the Ramsey problem: the planner taxes an interval of the healthiest agents so that their participation constraints are binding, and redistributes the proceeds to equalize the consumption of everyone else. If the tax system is fully efficient, in the sense that there is no waste associated with raising taxes, it follows that this "Ramsey allocation" is optimal. In other words, it is more efficient to redistribute directly through taxes and subsidies than through diversification of risk pools. The intuition is that negative assortative pooling promises a very high consumption level to a subset of the sickest types, whereas direct redistribution achieves a more even allocation.

More generally, if taxes and subsidies are not fully efficient and consequently a fraction of the tax revenue is lost, information design becomes an important redistributive policy tool. The Online Appendix presents a formal analysis of this case. Proposition B1 shows that there exists a threshold level such that if the tax system is more efficient than this threshold, the Ramsey allocation is optimal, while if it is less efficient, the optimal rating system is characterized by the properties presented in Theorem 2.

Implementation. The main obstacle to implementing the optimal rating system is that the regulator needs to have access to at least as much information as market participants and be able to accurately predict the medical costs. While these assumptions may seem strong, we contend that in various cases of interest regulators have access to detailed information about medical histories and genetic data, and the availability of big data together with machine learning techniques yield accurate predictions. In fact, regulators routinely estimate the average cost of different risk pools and cross-subsidize between them. For example, in Medicare and ACA exchanges, insurers are compensated whenever the expected cost in their pool exceeds the average cost in the population of insurees (Geruso and Layton (2017)). As far as we can tell, the amount of information required to perform these computations is similar to the information used in the optimal rating system.

Another implementation obstacle is that negative assortative pooling induces payoffs that are non-monotone in type, in the sense that the expected utility of healthier types may be lower. Non-monotonicity is a potential concern for policy-makers because such allocations may appear unfair, and also because of moral hazard (agents may want to mimic lower types). The Online Appendix presents an analysis of the optimal rating system under the additional restrictions of monotone expected payoffs.

References

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

Aumann, R. J., and M. Maschler (1995): *Repeated Games with Incomplete Information*. MIT Press.

Bar-Isaac, H., I. Jewitt, and C. Leaver (2017): "Multidimensional Asymmetric Information, Adverse Selection, and Efficiency," Discussion paper.

Bergemann, D., and S. Morris (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica: Journal of the Econometric Society*, 81(4), 1251–1308.

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

Chiappori, P.-A. (2006): "The Welfare Effects of Predictive Medicine," *Competitive Failures in Insurance Markets: Theory and Policy Implications*.

Epstein, L. G., and U. Segal (1992): "Quadratic Social Welfare Functions," *Journal of Political Economy*, 100(4), 691–712.

Finkelstein, A., N. Hendren, and M. Shepard (2017): "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts," Discussion paper, NBER Working Paper 23668.

Finkelstein, A., and K. McGarry (2006): "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *The American Economic Review*, 96(4), 938.

Finkelstein, A., K. McGarry, and A. Sufi (2005): "Dynamic Inefficiencies in Insurance Markets: Evidence from Long-Term Care Insurance," *The American Economic Review*, 95(2), 224–228.

Geruso, M., and T. J. Layton (2017): "Selection in Health Insurance Markets and Its Policy Remedies," *Journal of Economic Perspectives*, 31(4), 23–50.

Goldstein, I., and Y. Leitner (2015): "Stress Tests and Information Disclosure," Discussion paper.

Handel, B., I. Hendel, and M. D. Whinston (2015): "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk," *Econometrica*, 83(4), 1261–1313.

Hendel, I., and A. Lizzeri (2003): "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *The Quarterly Journal of Economics*, 118(1), 299–328.

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

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

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

Rawls, J. (2009): *A Theory of Justice*. Harvard University Press.

Roesler, A.-K., and B. Szentes (2017): "Buyer-Optimal Learning and Monopoly Pricing," *The American Economic Review*, 107(7), 2072–80.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, pp. 629–649.

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

Sen, A. (1977): "On Weights and Measures: Informational Constraints in Social Welfare Analysis," *Econometrica: Journal of the Econometric Society*, pp. 1539–1572.

A Appendix

We begin by establishing that the optimal rating system will use at most $N+1<\infty$ ratings. We then provide two useful lemmas that help us derive the characterization result in Theorem 2. We conclude the proof of our main results by showing that the output of the algorithm provided in Theorem 1 is the unique rating system that satisfies the properties of Theorem 2.

Given any rating system, we refer to the rating inducing the highest equilibrium price as s_0 and price $t_0 = E(\theta|s_0, i \in A(t_0))$.

Remark A1. The optimal rating system has at most $N+1$ ratings.

Proof. Fix σ with ratings $s_0, \ldots, s_{M-1}$ and suppose $M > N+1$. By the pigeonhole principle, there must exist some $i \in \Theta$ such that i is in the support of two ratings s_k and s_j, $k, j \neq 0$, and i is the healthiest type who participates in each of these ratings. Since ratings are not redundant, $t_k \neq t_j$, and assume that $t_k > t_j$. Type i is the maximal type who is willing to participate in both ratings, and therefore $\phi_l \geq \phi_i > t_k > t_j$ for all types $l \in \Theta$ who participate in rating j. We now construct a welfare-improving rating system, $\hat{\sigma}$: $\hat{\sigma}_{jl} = \sigma_{jl} + (1-\beta)\sigma_{0l}$, $\hat{\sigma}_{0l} = \beta\sigma_{0l}$, $\hat{\sigma}_{j'l} = \sigma_{j'l}$ for all $l \in \Theta$ and $j' \neq 0, j$. Thus, for β large enough, $\phi_i \geq \hat{t}_j$. To see that this is an improvement, notice that the only change in the allocation pertains to ratings s_0 and s_j. By construction, notice that $\hat{t}_0 = t_0$, $\sum_i \sigma_{0i} > \sum_i \hat{\sigma}_{0i}$, $\sum_i \sigma_{0i} t_0 + \sum_i \sigma_{ji} t_j = \sum_i \hat{\sigma}_{0i} \hat{t}_0 + \sum_i \hat{\sigma}_{ji} \hat{t}_j$, and $t_j > \hat{t}_j > t_0$. It follows that the distribution of prices under $\hat{\sigma}$ is a mean-preserving contraction of that under σ. $\square$

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. Claim (1) follows from the same argument used to prove Remark A1. Suppose, for a contradiction, that there exists an optimal rating system σ with some signal $s_j \neq s_0$ and for all types i with $\sigma_{ji} > 0$, $t_j < \phi_i$. We now construct a welfare-improving rating system, $\hat{\sigma}$: $\hat{\sigma}_{jl} = \sigma_{jl} + (1-\beta)\sigma_{0l}$, $\hat{\sigma}_{0l} = \beta\sigma_{0l}$, $\hat{\sigma}_{j'l} = \sigma_{j'l}$ for all $l \in \Theta$ and $j' \neq 0, j$. Thus, for β large enough, $\phi_i \geq \hat{t}_j$. To see that this is an improvement, notice that the only change in the allocation pertains to ratings s_0 and s_j. By construction, notice that $\hat{t}_0 = t_0$, $\sum_i \sigma_{0i} > \sum_i \hat{\sigma}_{0i}$, $\sum_i \sigma_{0i} t_0 + \sum_i \sigma_{ji} t_j = \sum_i \hat{\sigma}_{0i} \hat{t}_0 + \sum_i \hat{\sigma}_{ji} \hat{t}_j$, and $t_j > \hat{t}_j > t_0$. It follows that the distribution of prices under $\hat{\sigma}$ is a mean-preserving contraction of that under σ.

To prove claim (2), we proceed by contradiction and assume that $\sigma_{ji} > 0$, $\theta_i \leq t_0$, $t_j < \phi_i$, and i is not the unique type that has positive probability of receiving the worst signal. We construct a welfare improving rating system $\hat{\sigma}$ with an additional signal denoted by s_{M+1}. There are four cases to consider.

Case 1: $t_0 > \theta_i > t_j$. We construct rating system $\hat{\sigma}$ in 4 Steps:

1. $\hat{\sigma}_{0l} = (1-\gamma-\lambda)\sigma_{0l}$ for all $l \in \Theta$ and $1 > \gamma + \lambda > 0$ and $\gamma, \lambda \geq 0$;

2. $\hat{\sigma}_{jl} = \sigma_{jl} + \lambda\sigma_{0l}$ for all $l \neq i$ and $\hat{\sigma}_{ji} = (1-\delta)\sigma_{ji} + \lambda\sigma_{0i}$ for some $\delta \geq 0$;

3. $\hat{\sigma}_{(M+1)l} = \gamma\sigma_{0l}$ for $l \neq i$ and $\hat{\sigma}_{(M+1)i} = \delta\sigma_{ji} + \gamma\sigma_{0i}$.

4. For all $k \neq j, 0$ we have $\sigma_{kl} = \hat{\sigma}_{kl}$.

In words, we move a representative sample of those types who were in rating 0 (Step 1) and distribute them to ratings j and $M+1$ (Steps 2 and 3); we move type i from rating j (Step 1) to rating $M+1$ (Step 3); and we keep everyone else in the same rating (Step 4). By construction $t_0 = \hat{t}_0$ and,

$$\hat{t}_{M+1} = \frac{\sum_{l=1}^{N} \mu_l \hat{\sigma}_{(M+1)l}\theta_l}{\sum_{l=1}^{N} \mu_l \hat{\sigma}_{(M+1)l}} = \frac{\mu_i \delta \sigma_{ji}\theta_i + t_0 \sum_{l=1}^{N} \gamma \mu_l \sigma_{0l}}{\mu_i \delta \sigma_{ji} + \sum_{l=1}^{N} \gamma \mu_l \sigma_{0l}} \in (\theta_i, t_0)$$

For any $\delta > 0$, there exists some $\gamma(\delta)$ such that for all $\gamma < \gamma(\delta)$, we have that $\theta_i < \hat{t}_{M+1} \leq \phi_i$. The equilibrium price $\hat{t}_j$ satisfies

$$\hat{t}_j = \frac{\sum_{l=1}^{N} \mu_l \hat{\sigma}_{jl}\theta_l}{\sum_{l=1}^{N} \mu_l \hat{\sigma}_{jl}} = \frac{\sum_{l=1}^{N} \mu_l (\sigma_{jl} + \lambda \sigma_{0l})\theta_l - \mu_i \delta \sigma_{ji}\theta_i}{\sum_{l=1}^{N} \mu_l (\sigma_{jl} + \lambda \sigma_{0l}) - \mu_i \delta \sigma_{ji}}.$$

Since $\theta_i \geq t_j$, for any $\delta > 0$ sufficiently small, there exists some $\lambda(\delta)$ such that $\hat{t}_j = t_j$. Therefore, there exist combinations of $(\delta, \lambda, \gamma)$ such that $t_0 = \hat{t}_0 > \hat{t}_{M+1} > t_j = \hat{t}_j$ and for all $k \neq j$, $\hat{t}_k = t_k$. Since the price distributions under σ and $\hat{\sigma}$ have the same mean, $\hat{\sigma}$ induces a mean-preserving contraction of σ.

Case 2: $t_0 = \theta_i > t_j$. Then, either i is the unique type in rating s_0 and we are done, or there must exist some other type i' with $\sigma_{0i'} > 0$ and $\theta_{i'} > \theta_i$. We construct rating system $\hat{\sigma}$ with $\hat{\sigma}_{ji} = (1-\delta)\sigma_{ji}$ and $\sigma_{ji'} = \sigma_{ji'} + \gamma \sigma_{0i'}$, $\hat{\sigma}_{0i} = \sigma_{0i} + \delta \sigma_{ji}$ and $\hat{\sigma}_{0i'} = \delta \sigma_{0i'}$. In words, we move i from rating s_j to s_0 and i' from s_0 to s_j. Since $\theta_{i'} > \theta_i > t_j$, we can choose combinations of parameters (λ, δ) such that $\hat{t}_j = t_j$ and, therefore, $\hat{t}_0 < t_0$, which also leads to a mean-preserving contraction of the price distribution.

Case 3: $\theta_i = t_j$. We construct a rating system $\hat{\sigma}$ as in Case 1, only $\lambda = 0$.

Case 4: $t_j > \theta_i$. In this case, moving type $i \in \Theta$ from the support of signal s_j leads to an increase in its price, so we cannot simply replace type i with types from the support of s_0. However, in such a case there must be an additional type $i' \in \Theta$ such that $\sigma_{ji'} > 0$ with $\theta_{i'} > t_j$ (for otherwise the average cost of agents in s_j cannot be above θ_i). We construct a welfare-improving rating system $\hat{\sigma}$, with $\hat{\sigma}_{0l} = (1-\gamma)\sigma_{0l}$ for all l, $\hat{\sigma}_{ji} = (1-\delta)\sigma_{ji}$, $\hat{\sigma}_{ji} = (1-\delta')\sigma_{ji'}$, $\hat{\sigma}_{(M+1)i} = \delta \sigma_{ji} + \gamma \sigma_{0i}$, $\hat{\sigma}_{(M+1)i'} = \delta' \sigma_{ji'} + \gamma \sigma_{0i'}$. We can choose δ and δ' such that $\hat{t}_j = t_j$ and for γ sufficiently small, $\min\{\phi_i, t_0\} > t_{M+1} > t_j$, which implies that $\hat{\sigma}$ induces a mean-preserving contraction of the price distribution. In other words, we construct a virtual type which is a convex combination of type i and i' that has an average cost of t_j and proceed as in Case 3. $\square$

As a result of Lemma 1, under the optimal rating system, for every signal s_j (except s_0), there exists a unique healthiest type i for which $t_j = \phi_i$ (Condition 1 of Lemma 1), $\sigma_{ji} = 1$ (Condition 2 of Lemma 1), and for all other types $l \neq i$, if $\sigma_{jl} > 0$ then $t_j < \theta_l$. Therefore, we relabel the signals so that $t_j = \phi_i$ iff $i = j$. Under this relabeling, negative assortative pooling is equivalent to the following condition: if there exists a type l and signal s_j, with $j < l$ such that $\sum_{1 \leq k \leq j} \sigma_{kl} < 1$, then $\sigma_{jl'} = 0$ for all types $j < l' < l$. Notice that if $i < j$, then type i cannot be in the support of rating s_j.

Lemma 2. The optimal rating system satisfies negative assortative pooling.

Proof. Assume that the statement is not true, so that there exists a type $l \in \Theta$ and signal s_j, with $j > l$ such that $\sum_{1 \leq k \leq j} \sigma_{kl} < 1$, but for some type $l > l' > j$, $\sigma_{jl'} > 0$. Again, it must hold that there exists some other signal s_j, with $j < j' < l$ (or $s_j = s_0$) such that $\sigma_{j'l} > 0$. By definition, $t_j = \frac{\sum_{i=1}^{N} \mu_i \sigma_{ji} \theta_i}{\sum_{i=1}^{N} \mu_i \sigma_{ji}}$, and let

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

The key observation is that by the second condition of Lemma 1, we have that $\Delta > 0$ because $\theta_l > \theta_{l'} > t_j$. Consider then $\hat{\sigma}$ such that $\hat{\sigma}_{ki} = \sigma_{ki}$, for all $i \neq l, l'$ and for all ratings s_k; $\hat{\sigma}_{kl} = \sigma_{kl}$ and $\hat{\sigma}_{kl'} = \sigma_{kl'}$ for all $k \neq j, j'$; we set $\hat{\sigma}_{jl}$ and $\hat{\sigma}_{jl'}$ such that

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

where $\hat{\sigma}_{jl} > \sigma_{jl}$ and $\hat{\sigma}_{jl'} < \sigma_{jl'}$; and finally, we set $\hat{\sigma}_{j'l} = 1 - \sum_{k \neq j'} \hat{\sigma}_{kl}$ and $\sigma_{j'l'} = 1 - \sum_{k \neq j'} \hat{\sigma}_{kl'}$. Therefore, by construction, we have that $\hat{t}_j = t_j > \hat{t}_{j'} > t_{j'}$; and the average price under σ is a mean-preserving spread of that under $\hat{\sigma}$.$\square$

Proof of Theorem 2. By Lemmata 0, 1 and 2, it follows that no exclusion, no rents at the top, and negative assortative pooling are necessary properties of an optimal rating system. We now show that there exists a unique rating system satisfying all three properties. Suppose that σ and $\hat{\sigma}$ satisfy these three properties and induce the prices $(t_1, t_2, ..., t_k, t_0)$ and $(\hat{t}_1, \hat{t}_2, ..., \hat{t}_k, \hat{t}_{k+1}, ..., \hat{t}_l, \hat{t}_0)$, respectively. First, if prices are in ascending order, then $t_j = \phi_j$ for all $j \leq k$, and $\hat{t}_j = \phi_j$ for all $j \leq l$. To see this, observe that by no rents it must be that for $j \leq k$ there exists $i \in \Theta$ such that $t_j = \phi_i$. If $i < j$, we have a contradiction because then there exists a price $j' < j$ such that $t_{j'} \notin \phi_1, ..., \phi_N$, (by the pigeonhole principle). If $i > j$, we have a contradiction since there exists type $i' < i$ and $\phi_{i'} < t_j = \phi_i$ which implies that either $\sigma_{i'j'} > 0$ for $j' < j$ which contradicts negative assortative pooling or $\sigma_{i'j'} > 0$ for $j' > j$ which contradicts full trade $(t_{j'} > \phi_{i'})$. Moreover, by negative assortative pooling, $\sigma_{ji} = \hat{\sigma}_{ji}, \forall j \leq k$.

On the one hand, we have that:

$$\Pr_{\sigma}(s_0)t_0 = \Pr_{\hat{\sigma}}(s_0)\hat{t}_0 + \sum_{j \geq k+1} \Pr_{\hat{\sigma}}(s_j)\hat{t}_j.$$

On the other hand, we have that:

$$\Pr_{\hat{\sigma}}(s_0)\hat{t}_0 + \sum_{j \geq K+1} \Pr_{\hat{\sigma}}(s_j)\hat{t}_j = \Pr_{\hat{\sigma}}(s_0)\hat{t}_0 + \sum_{j \geq k+1} \Pr_{\hat{\sigma}}(s_j)\phi_j$$
$$> \phi_{k+1} \left(\sum_{j \geq k+1} \Pr_{\hat{\sigma}}(s_j) + \Pr_{\hat{\sigma}}(s_0) \right) = \phi_{k+1} \Pr_{\sigma}(s_0) \geq \Pr_{\sigma}(s_0)t_0,$$

where the first equality follows from the above; the inequality follows by the ordering of the willingness-to-pay and the fact that $\hat{t}_0$ is the highest price, and at least one of those signals have a 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 satisfies the participation constraints. Hence, we have a contradiction. □

Proof of Theorem 1. The algorithm yields a unique rating system in at most N steps. By construction it satisfies full trade (if $\phi_i < t_0$, $\sigma_{ii} = 1$ and $t_i = \phi_i$, else $\phi_i > t_0 \geq t_j$ for all j); no rents at the top (if $t_j < t_0$, then $t_j = \phi_j = \min\{i' : \sigma_{ii'} > 0\}$); and negative assortative pooling (if for some j, $\sum_{k \leq j} \sigma_{ki} < 1$, $\sigma_{ji'} = 0$ for all $j < i' < i$). By Theorem 2, there exists a unique such rating system, and, therefore, the algorithm yields the optimal rating system.

B Online Appendix

B.1 Proof of Proposition 1

Proof. The proof applies the same arguments that were used to characterize the optimal rating system in Section 3. The only difference is that we must keep track of the interim beliefs. We begin with several definitions. Given a rating system σ, let $\mu^j_{il}(\sigma)$ be the probability that an agent of type i who observes signal s_j attaches to the event that she belongs to type l. Notice that the partition structure of the prior information implies that for all $i, i' \in P_k$, $\mu^j_{il}(\sigma) = \mu^j_{i'l}(\sigma) = \frac{\mu_l \sigma_{jl}}{\sum_{l' \in P_k} \mu_{l'} \sigma_{jl'}}$ if $l \in P_k$, and $\mu^j_{il} = \mu^j_{i'l} = 0$ otherwise. The interim belief $\mu^j_k(\sigma)$ is the posterior distribution over the type space of an agent of type $i \in P_k$ who receives signal s_j. We define $\phi^j_k(\sigma)$ to be the willingness-to-pay of an agent who has belief $\mu^j_k(\sigma)$. Whenever there is no confusion, we suppress σ.

Since the regulator has superior information, she can induce full trade and the optimal rating system satisfies no exclusion (see Lemma 0 for the formal argument). We first show that an optimal rating system σ satisfies no rents at the top with respect to the interim beliefs:

$$t_j = \min\{\phi^j_k(\sigma) : \sigma_{ji} > 0 \text{ for some } i \in P_k\}, \forall j \neq 0.$$

Suppose that there exists a type $l \in P_k$ and signal s_j such that $\sigma_{jl} > 0$, and $t_j < \phi^j_k \leq \phi^j_{k'}$ for all k' such that $\sigma_{j'i} > 0$ with $i \in P_{k'}$. In other words, types in P_k are the healthiest in rating j and receive rents. We can construct a welfare-improving rating system $\hat{\sigma}$ as in Lemma 1: we "move" a small fraction $\epsilon > 0$ of the mass in the signal associated with the highest price to signal s_j such that $\phi^j_k(\hat{\sigma}) > \hat{t}_j$. Observe that such an ϵ exists because if $\{l \in P_k : \sigma_{0l} > 0\} = \emptyset$, then the interim beliefs satisfy $\mu^j_k(\hat{\sigma}) = \mu^j_k(\sigma)$ and $\phi^j_k(\hat{\sigma}) = \phi^j_k(\sigma)$; and if there exists some $l \in P_k$, with $\sigma_{0l} > 0$, then $\phi^j_k(\hat{\sigma}) > \phi^j_k(\sigma)$.

We now prove that under an optimal rating system, for every signal $s_j \neq s_0$ there exists a unique element of the partition, P_k, such that $\sigma_{ji} > 0$ for some $i \in P_k$ and $E(\theta \mid s_j, i \in P_k) \equiv \theta^j_k < t_0$. Suppose not, then there exist $i \in P_k$ and $i' \in P_{k'}$ such that $\sigma_{ji} > 0$, $\sigma_{ji'} > 0$, and $\theta^j_{k'} < \theta^j_k < t_0$. Let $i_0 \equiv \min\{l : \sigma_{0l}\} \notin P_k \cup P_{k'}$. We can construct a welfare improving rating system $\hat{\sigma}$ as in

Lemma 1, using a new rating s_{M+1}. In the case that $\theta_k^j \geq t_t$, we move a portion of agents of type i_0 from s_0 and distribute them to signals s_j and s_{M+1}; we move a representative sample of the types in P_k from signal s_j to signal S_{M+1} ; and we keep everyone else in the same rating (Case 1 in the proof of Lemma 1). The proportions are such that $t_j = \hat{t}_j$, $\hat{t}_{M+1} = \phi_k^{M+1}(\hat{\sigma})$. Importantly, no participation constraint is violated because $i_0 \notin P_k \cup P_{k'}$, and hence $\phi_{k'}^j(\hat{\sigma}) = \phi_{k'}^j(\sigma)$ and $\phi_k^{M+1}(\hat{\sigma}) = \phi_k^j(\sigma)$. For the case that $\theta_k^j < t_j$, see Case 4 in the proof of Lemma 1.

We say that a rating system satisfies *Property** if the healthiest type in the support of two different signals belongs to two different elements of the partition. That is, if $i = \min\{l : \sigma_{jl} > 0\}$ and $i \in P_k$, then $\forall j' \neq j : \min\{l : \sigma_{j'l} > 0\} \notin P_k$.

We now show that Property* is necessary. Given the element P_k, let $S_k = \{s_j : t_j = \phi_k^j(\sigma)\}$ be the set of all signals s_j for which the healthiest type is in P_k. If there are more than two signals in S_k, let $\hat{\sigma}$ be an alternative rating system whereby all the signals in S_k are merged. That is, $\hat{\sigma}_{(M+1)i} = \sum_{j \in S_k} \sigma_{ji}$ and $\hat{\sigma}_{ji} = \sigma_{ji}$ for $j \notin S_k$. We now show that $\hat{\sigma}$ satisfies the participation constraints. Let $\mu_i^k = \frac{\mu_i}{\sum_{l \in P_k} \mu_l}$ and we have that for all $j \in S_k$,

$$u(w - \phi_k^j(\sigma)) = \sum_{i \in P_k} \Pr_{\sigma}(i | s_j, i \in P_k) U_i = \sum_{i \in P_k} \left(\frac{\sigma_{ji} \mu_i^k}{\sum_{l \in P_k} \sigma_{jl} \mu_l^k} \right) U_i. \quad (1)$$

Likewise,

$$u(w - \phi_k^{M+1}(\hat{\sigma})) = \sum_{i \in P_k} \Pr_{\hat{\sigma}}(i | i \in P_k, s_{M+1}) U_i$$

$$= \sum_{i \in P_k} \left(\frac{\sum_{j \in S_k} \sigma_{ji} \mu_i^k}{\sum_{l \in P_k} \sum_{j \in S_k} \sigma_{jl} \mu_l^k} \right) U_i = \sum_{j \in S_k} \sum_{i \in P_k} \left(\frac{\sigma_{ji} \mu_i^k}{\sum_{j \in S_k} \sum_{l \in P_k} \sigma_{jl} \mu_l^k} \right) U_i.$$

Therefore, if we take (1) and sum over the signals in S_k weighted by the probability that agents of each type $l \in P_k$ receive them, we get:

$$u(w - \phi_k^{M+1}(\hat{\sigma})) = \sum_{j \in S_k} \frac{\sum_{l \in P_k} \sigma_{jl} \mu_l^k}{\sum_{j \in S_k} \sum_{l \in P_k} \sigma_{jl} \mu_l^k} u(w - \phi_k^j(\sigma)),$$

and Jensen's inequality implies that

$$\phi_k^{M+1}(\hat{\sigma}) \geq \sum_{j \in S_k} \frac{\sum_{l \in P_k} \sigma_{jl}\mu_l}{\sum_{j \in S_k} \sum_{l \in P_k} \sigma_{jl}\mu_l} \phi_k^j(\sigma). \tag{2}$$

For each signal $s_j \in S_k$, since only types in element k are such that $\theta_l \leq t_j = \phi_k^j(\sigma)$, we have that

$$\sum_{l \in P_k} \sigma_{jl}\mu_l(\phi_k^j(\sigma) - \theta_l) = \sum_{l \notin P_k} \sigma_{jl}\mu_l(\theta_l - \phi_k^j(\sigma)),$$

and summing over the signals yields

$$\sum_{j \in S_k} \sum_{l \in P_k} \sigma_{jl}\mu_l(\phi_k^j(\sigma) - \theta_l) = \sum_{j \in S_k} \sum_{l \notin P_k} \sigma_{jl}\mu_l(\theta_l - \phi_k^j(\sigma)).$$

Rewriting (2), and using that

$$\sum_{j \in S_k} \sum_{l \in P_k} \sigma_{jl}\mu_l = \sum_{l \in P_k} (\sum_{j \in S_k} \sigma_{jl})\mu_l = \sum_{l \in P_k} \hat{\sigma}_{(M+1)l}\mu_l,$$

we get

$$\sum_{j \in S_k} \sum_{l \in P_k} \sigma_{jl}\mu_l(\phi_k^j(\sigma) - \theta_l) \leq \sum_{l \in P_k} \hat{\sigma}_{(M+1)l}\mu_l(\phi_k^{M+1}(\hat{\sigma}) - \theta_l).$$

As a result,

$$\sum_{l \in P_k} \hat{\sigma}_{(M+1)l}\mu_l(\phi_k^{M+1}(\hat{\sigma}) - \theta_l) \geq \sum_{l \notin P_k} \hat{\sigma}_{(M+1)l}\mu_l(\theta_l - \phi_k^{M+1}(\hat{\sigma})),$$

from the formula for the equilibrium price, we obtain

$$\sum_l \hat{\sigma}_{(M+1)l}\mu_l \hat{t}_{M+1} = \sum_l \hat{\sigma}_{(M+1)l}\mu_l\theta_l \leq \sum_l \hat{\sigma}_{(M+1)l}\mu_l\phi_k^{M+1}(\hat{\sigma}),$$

and, therefore, $\phi_k^{M+1}(\hat{\sigma}) \geq \hat{t}_{M+1}$. As a result, $\hat{\sigma}$ is feasible and yields a mean-preserving contraction of the price distribution.

In the next step we establish negative assortative pooling. Assume towards a contradiction that there exist types $i < i'$ and signals $s_j, s_{j'}$ where $i = \min\{l :$

$\sigma_{jl} > 0\}$ and $i' = \min\{l : \sigma_{j'l} > 0\}$, but there exist types $l' > l > i$ such that $\sigma_{jl} > 0$ and $\sigma_{j'l'} > 0$. First, it follows from Property* that $i \in P_k$ and $i' \in P_{k'}$ and $k \neq k'$. Second, from no rents at the top, $t_j = \phi_k^j < t_{j'} = \phi_{k'}^{j'} \leq t_0$. Third, since $\theta_{l'} > \theta_l > t_0 > t_j$, we can construct a welfare improving rating system $\hat{\sigma}$ by moving a fraction of type l to signal $s_{j'}$ and a fraction of type l' to signal s_j such that $\hat{t}_j = t_j$ does not change and $\hat{t}_{j'} < t_j$ (the construction is identical to the proof of Lemma 2). Observe that $l' \notin P_k$ because $l' > i' \in P_{k'}$, and thus $\phi_k^j(\hat{\sigma}) = \phi_k^j(\sigma) \geq \hat{t}_j$. If $l \notin P_{k'}$, then $\phi_{k'}^{j'}(\hat{\sigma}) = \phi_{k'}^{j'}(\sigma) > \hat{t}_{j'}$; and if $l \in P_{k'}$, then since $\theta_l > t_0 \geq t_{j'} \geq E(\theta | s_{j'}, i \in P_{k'})$, $\phi_{k'}^{j'}(\hat{\sigma}) \geq \phi_{k'}^{j'}(\sigma)$. Thus, no participation constraint is violated.

Finally, there exists a unique rating system that satisfies no exclusion, Property*, no rents at the top with respect to the interim beliefs, and negative assortative pooling. To see this, suppose that σ and $\hat{\sigma}$ satisfy these properties and induce the prices $(t_1, t_2, ..., t_j, t_0)$ and $(\hat{t}_1, \hat{t}_2, ..., \hat{t}_j, \hat{t}_{j+1}, ..., \hat{t}_{j'}, \hat{t}_0)$, respectively. Let us assume that the prices are in ascending order and that all signals have a strictly positive support. We will show that for all $l < j$: $t_l = \hat{t}_l$; $\sigma_{li} = \hat{\sigma}_{li}, \forall i \in \Theta$; and $\sigma_{il} = \hat{\sigma}_{il} = 1$, $\forall i \in P_l$. By induction, for $l = 1$ it follows from no exclusion and Property* that $\sigma_{1i} = \hat{\sigma}_{1i} = 1 \; \forall i \in P_1$, and from no rents at the top, $\hat{t}_1 = t_1$. Negative assortative pooling and no exclusion imply that $\sigma_{1i} = \hat{\sigma}_{1i}, \forall i \in \Theta$. Take $l > 1$, by induction, we have that $\sigma_{l'i} = \hat{\sigma}_{l'i}$ for all $l' < l$ and $i \in P_l$. Therefore, if there exists $i \in P_l$ such that $\sigma_{l'i} = \hat{\sigma}_{l'i} > 0$ then we are done (because by negative assortative pooling, all the types are exhausted: $\sigma_{ki} = \hat{\sigma}_{ki} = 0, \forall i \in P_k$ and $k > l$). Therefore, we assume that $\sigma_{l'i} = \hat{\sigma}_{l'i} = 0$ for $l' < l$. No exclusion and Property* imply $\sigma_{li} = \hat{\sigma}_{li} = 1 \; \forall i \in P_l$, and no rents at the top further implies $t_l = \hat{t}_l$. It follows that $\sigma_{li} = \hat{\sigma}_{li}$ for all $i \in \Theta$ (because otherwise, negative assortative pooling or the induction hypothesis is violated).

We can use the same argument as in Theorem 2 to show that $\hat{\sigma}_{(j+1)i} = \ldots = \hat{\sigma}_{j'i} = 0$. It is straightforward to check that the outcome of the algorithm satisfies these properties. $\square$

B.2 Direct Intervention

We have assumed throughout Sections 2 and 3 that the regulator can only influence the market outcomes by controlling the information used to price contracts. In health insurance markets, taxes and subsidies are also used to affect the outcomes. For example, the Affordable Care Act specifies a broad redistributive scheme across contract pools (the so-called risk-corridor), compensating insurers with excessive costs.[14] This section extends our model to address the question of how regulators optimally combine information design and direct intervention policies.

We consider the insurance market defined in Section 2 and let the regulator design the public rating system σ and also set taxes and subsidies on the risk pools. An agent who receives signal s_j and purchases insurance at price t_j, will consume $c_j = w - t_j(1 + \tau_j) + b_j$, where τ_j is the tax rate and b_j is the per-capita subsidy of contract pool j. We assume that redistribution is costly so that a certain fraction of total revenue is lost. The budget-balance condition is $\sum_{j=1}^{M} \sum_{i=1}^{N} \mu_i \sigma_{ji}(\alpha t_j \tau_j - b_j) = 0$, where $\alpha \in [0, 1]$ measures the efficiency of the tax system.[15]

Competition between insurers drives the price of each risk-pool to the average cost, $t_j = \min_t E_j\{\theta : i \in A(t)\}$. The same argument used to prove Lemma 0 applies here, and hence we restrict attention to rating systems that implement full trade. The regulator's problem is to design the rating system σ and the tax policy $\{\tau_i, b_i\}_{i=1}^{M}$ to maximize ex ante welfare:

[14] 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.

[15] There is a large literature on the inefficiencies associated with the implementation of risk-corridors in Medicare. For a recent review see Geruso and Layton (2017)

$$
\begin{aligned}
&\max_{\sigma,\tau,b} \sum_{i=1}^{N} \mu_i \sum_{j=1}^{M} \sigma_{ji} u(c_j) \\
&s.t.\ c_j = w - (1+\tau_j)t_j + b_j \\
&\qquad t_j = E_j(\theta) \\
&\qquad (1+\tau_j)t_j - b_j \leq \phi_i, \forall i : \sigma_{ij} > 0 \\
&\qquad \sum_{j=1}^{M} \sum_{i=1}^{N} \mu_i \sigma_{ji} (\alpha \tau_j t_j - b_j) = 0
\end{aligned}
$$

To gain intuition, let us compare two redistribution policies. In the benchmark model, we had $\tau_j = b_j = 0$ for all j and redistribution was achieved by creating diverse risk-pools design. The optimal rating system is characterized by the properties of Theorem 2. We shall refer to the resulting allocation as the No-Taxation (NT) allocation. Another relevant policy configuration uses taxation and subsidies as the only tool for redistribution. We shall refer to it as the Ramsey (R) allocation. In the Ramsey allocation, the rating system perfectly reveals each risk-type. Trade occurs at the actuarially fair price $t_j = \theta_j, \forall j$. An interval of the healthiest types $1, \ldots K$ are taxed so that$(1+\tau_j)\theta_j = \phi_j$ for $j = 1, \ldots K$. The tax revenues $B_0 = \alpha \sum_{j=1}^{K} \mu_j (\phi_j - \theta_j)$ are redistributed to the remaining $K+1, \ldots, N$ types in such a way that equates their consumption. The cutoff type K is set to maximize the agents that are subsidized.[16]

Observe that negative assortative pooling redistributes income by promising very high consumption profiles to an interval of sickest agents, while taxes and subsidies redistribute income directly to the sickest individuals. Therefore, in the case where the tax system is efficient, $\alpha = 1$, the Ramsey allocation has the same average consumption and lower variation in the consumption profiles. In the cases where the tax system is inefficient, $\alpha < 1$, the average consumption is lower under the Ramsey allocation. This trade-off drives the next Proposition.

[16] If types $1, \ldots, K$ are taxed and $c_0(K)$ is the consumption of the agents that are subsidized (the consumption level is equal at the bottom), then type $K+1$ is willing to participate at consumption level $c_0(K)$ and type K is not willing to participate at consumption level $c_0(K-1)$.

Proposition B1. There exists $1 > \alpha_1 > \alpha_2 > 0$ such that

1. If $\alpha > \alpha_1$ the Ramsey allocation is optimal
2. If $\alpha < \alpha_2$ the No-Taxation allocation is optimal
3. If $\alpha \in [\alpha_1, \alpha_2]$, there exists some j such that $\tau_j > 0$ and σ satisfies the properties of Theorem 2.

Before we present the proof of Proposition B1, we would like to point out that the relative efficiency of taxes depends very much on the distribution of types. Indeed, it can be shown that for every $\epsilon > 0$, there exists some μ such that $\alpha_1 > 1 - \epsilon$. Thus, information design is useful policy tool in a wide range of environments.

Proof. We begin by establishing that it is without loss of generality to consider schedules whereby $b_j = 0$ if $c_j > c_0$. To see this consider a policy (σ, τ, b) with $b_k > 0$,$t_k > \phi_k$ and $\phi_k > c_0$. Since $\tau_k = 0$, it follows that $\sum_i \mu_i \sigma_{ki}(\theta_i - b_k - \phi_k) = 0$. Consider now the following alternative scheme $(\hat{\sigma}, \tau, \hat{b})$, 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 since both tax revenue and the associated waste is the same in both policies, 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

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

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

$$\sum_i \mu_i(\sigma_{0i} + \beta\sigma_{ki})(\theta_i - \hat{b}_0 - (w - \hat{c}_0)) = \sum_i \mu_i \sigma_{0i}(\theta_i - b_0 - (w - c_0)) = 0.$$

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 $\alpha \geq \alpha_1$, where α_1 is the largest value of $\alpha \in (0,1)$ that solves

$$u(c_k) - u(c_0) = u'(c_0)(\alpha_1(\theta_N - \phi_k) - (\theta_N - (t_0 + b_0))).$$

We begin with the case in which the participation constraint of those agents who are assigned a consumption c_0 is slack (which will generically be true). Consider a policy (σ, τ, β) that implements the Ramsey allocation and let c_0 be the lowest consumption level. We perturb this policy so that a fraction of risk-types N are pooled with risk type k, and then adjust the taxes and subsidies to satisfy the participation constraints. Formally, we define the policy $(\hat{\sigma}, \hat{\tau}, \hat{b})$ such that $\hat{\sigma}_{NN} = (1-\beta)$, $\hat{\sigma}_{kN} = \beta$ for some k such that $w - \phi_k > c_0$ and $\hat{\sigma}_{ji} = \sigma_{ji}$ otherwise. The price of risk pool k increases, and hence the tax rate is adjusted so that,

$$\hat{t}_k(1 + \hat{\tau}_k) = \frac{\mu_k \theta_k + \beta \mu_N \theta_N}{\mu_k + \beta \mu_{\mathbf{N}}}(1 + \hat{\tau}_k) = \phi_k$$

The tax revenue is distributed to risk-types $K+1, \ldots, N-1$ and the fraction $(1-\beta)$ of risk-type N to equalize their consumption levels, as in the Ramsey allocation.[17] The lowest consumption level, $\hat{c}_0 = w - \hat{t}_0 + \hat{b}_0$ changes because (i) the per-capita subsidy, b_0, falls and (ii) the average cost of agents in the pool $\hat{t}_0$ also decreases. At the same time, a fraction β of risk-type N increase their consumption by $u(c_k) - u(c_0)$. Observe that for α sufficiently low, we have that $\hat{c}_0 \geq c_0$, and the resulting allocation is welfare improving. Otherwise, $c_0 > \hat{c}_0$, and approximating the change in welfare around $\beta = 0$, we have that this perturbation

[17] This construction assumes that in the Ramsey allocation, the participation constraint of type $K+1$ is slack so that a perturbation in consumption does not violate it (which is generically true).

yields an improvement if and only if

$$u(c_k) - u(c_0) > -u'(c_0)\frac{dc_0}{d\beta} = u'(c_0)(\alpha(\theta_N - \phi_k) - (\theta_N - (t_0 + b_0))).$$

The left-hand side represents the increment in utility for those who jump to a better pool. 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 represents the drop in subsidies needed to ensure that type k still participates. The second term measures the consumption premium of the sickest type when receiving the worst possible signal, which represents the marginal benefit for the remaining types when she gets excluded. If $\alpha = 1$, the right-hand side is exactly $u'(c_0)(c_k - c_0)$ and, therefore, concavity ensures that the inequality is satisfied. If $\alpha = 0$, the right-hand side becomes negative since the lowest type is better than the mean type in the worst pool, so the inequality always holds. Notice then that for every $\alpha < \alpha_1$,any deviation from the Ramsey allocation is welfare-diminishing.

Suppose then that there is a type i who gets assigned c_0 with positive probability and whose participation constraint binds. In this case, any perturbation that marginally reduces c_0 violates the participation constraint of i and leads to a discrete drop in c_0.

Therefore, we only need to prove that if $\alpha < \alpha_1$ and the participation constraints of those agents who obtain c_0 in the R allocation are slack, there is no alternative (σ, τ, b) 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 1 and 2 must still hold fix an optimal allocation in which types l,l' with $\theta_l < \theta_{l'}$and signals s_j and $s_{j'}$ 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 s_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 a strict improvement. □

B.3 Private Information

We begin by describing a simple example that explains why the optimal rating system need not induce full trade.

Example B.1. Suppose there are two cost-types $\theta_1 < \theta_2$, each of them with probability 1/2. A fraction η of the low-cost agents have a willingness-to-pay of $\theta_1 + \Delta$ and a fraction $1 - \eta$ have a willingness-to–pay of $\theta_1 + 2\Delta$. First notice that it is suboptimal to induce two different prices for the low-cost type.[18] If $\theta_2 - \theta_1 < 2\Delta$, then the first best is feasible and the optimal rating system reveals no information. Else, if $\eta > \eta^*$, then the optimal rating system still implements full trade with $\sigma_{11} = 1$, $\sigma_{12} = \frac{\Delta}{\theta_2 - (\theta_1 + \Delta)}$ so that $t_1 = \theta_1 + \Delta$ and $t_0 = \theta_2$. If $\eta < \eta^*$, then the optimal rating system excludes the low-willingness-to-pay agents and sets $\sigma_{11} = 1$, $\sigma_{12} = \frac{2\Delta(1-\eta)}{\theta_2 - (\theta_1 + 2\Delta)}$, so that $t_1 = \theta_1 + 2\Delta$, $t_0 = \theta_2$ and those who do not trade obtain a lottery worth $\theta_1 + \Delta$. Notice that since $\frac{2\Delta}{\theta_2 - (\theta_1 + 2\Delta)} > \frac{\Delta}{\theta_2 - (\theta_1 + \Delta)}$, $\eta^* > 0$ and full trade is suboptimal in an open interval of parameters.

Proof of Corollary 2. Notice first that it is possible to achieve full trade by revealing the type of each agent (since $G_i(\theta_i) = 0$). But as Example B.1 shows, it need not be optimal to implement full trade. It is true, however, that an optimal rating system induces a positive measure of agents of each type to trade. Therefore, suppose that a certain agent of type i is the healthiest agent who receives signal s_j, then, in an optimal rating system, all agents of type i will receive signal s_j and they will trade if and only if their willingness-to-pay exceeds

[18] To see this notice that the ex-ante expected utility would be the weighted average of the expected utility of both signals and, generically, one is higher than the other.

a certain threshold $x_{ij} = t_j = E_j(\theta \mid \phi_i \leq x_{ij})$. The first part follows from the same argument as in the case with no private information and the second by our assumptions on preferences.
Hence, we can restrict attention to allocations with at most $N + 1$ signals and such that the agent with the lowest willingness-to-pay who trades in each signal (except perhaps the worst signal) receives no rents.

Now, consider any rating system that violates negative assortative pooling with respect to the information that the regulator has for a given rating s_j. By the argument we used in Lemma 2, there must exist a redistribution of agents from types whose expected cost is above the price t_j and that it induces the same price in pool j. Since t_j is constant, $x_{ij} = t_j$ and the set of agents of type i who trade in pool j does not change, then we can still apply the argument in Lemma 2 and show that this redistribution is welfare-improving and feasible. As a result, we have that the optimal rating system satisfies negative assortative pooling in the sense that if the healthiest agent in rating j is healthier than the healthiest agent in rating j', then every other agent who has positive probability in rating j is sicker than any agent in rating j'.

B.4 Monotonic Payoffs

We finally consider the problem of a regulator who maximizes ex-ante expected utility subject to the participation constraints and the restriction that if $i > l$ then $\sum_j \sigma_{ji} u(w - t_j) \leq \sum_j \sigma_{jl} u(w - t_j)$. The following result shows that there exists an optimal algorithm that satisfies a modified version of the properties derived in Theorem 2.

Proposition B.2. There is an optimal monotonic test with the following features:

1. All types $i \geq l$, with $\phi_l \geq E(\theta \mid \theta \geq \theta_l) = t_0$ are treated equally ($\sigma_{ji} = \sigma_{jN}$)

2. Types $l < i \leq k$ obtain the same expected utility as the sickest type, $\sum_j \sigma_{ji} u(w - t_j) = \sum_j \sigma_{jN} u(w - t_j)$ but $\sigma_{ji} > 0$ iff $j \geq i$.

3. Types $i > k$ are such that $\sigma_{ii} = 1$ and $u(w - t_i) > \sum_j \sigma_{j(i+1)} u(w - t_j)$.

Proof. Part 3 follows directly by Theorem 2 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 $\bar{u}_0 = \max_{j \le k} \sum \sigma_{ji} u(w - t_j)$. Monotonicity requires that $\bar{u}_0 = u_k$, but optimality requires that $u_k = \min_{j \le k} \sum \sigma_{ji} u(w - t_j)$. Hence, $u_i = \bar{u}_0$ for all $i \le k$. 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 $i \le l$ where l is the highest type such that $\sum_{i \le l} \mu_i \theta_i \le \sum_{i \le l} \mu_i \phi_l$ follows from Theorem 2 with the additional monotonicity constraint since the optimal profile has a decreasing expected utility. □

Example B.2. Consider the CARA case with a uniform prior over $\Theta = \{0, 6, 7, 8\}$ with $\phi_1 = 5$, $\phi_2 = 6$ and $\phi_3 = 8$.

- Without the monotonicity restriction, the optimal rating system creates two pools: pool 1 consists of the entire population of types $1, 3, 4$: $\sigma_{11} = \sigma_{13} = \sigma_{14} = 1$; and pool 0 consists of the entire population of type 2, and thus $\sigma_{22} = 1$. Observe that types 3 and 4 receive a higher payoff than type 2.

- A rating system that is optimal under the monotonicity restriction creates 3 pools: pool 1 contains the entire population of type 1 and a mixture of the other population: $\hat{\sigma}_{11} = 1$, $\hat{\sigma}_{13} = \hat{\sigma}_{14} = 0.997$, $\hat{\sigma}_{12} = 0.011$; pool 2 contains the residual population of type 2, $\hat{\sigma}_{22} = 1 - \sigma_{12}$; and pool 0 consists of the residual population of types 3 and 4, $\hat{\sigma}_{03} = \hat{\sigma}_{04} = 1 - \hat{\sigma}_{13}$. Observe that the expected payoffs of types $2, 3, 4$ are equal and the resulting allocation is a mean-preserving spread of the unconstrained optimal allocation.