Garcia , D., & Tsur, M. (2021). Information Design in Competitive Insurance Markets. *Journal of Economic Theory*, *191*, [105160]. https://doi.org/10.1016/j.jet.2020.105160

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC-ND

Link to published version (if available):
10.1016/j.jet.2020.105160

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at https://www.sciencedirect.com/science/article/pii/S0022053120301538 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 191 (2021) 105160
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Information design in competitive insurance markets

Daniel Garcia [a,b], Matan Tsur [c,*,1]

[a] University of Vienna, Austria
[b] CESifo, Germany
[c] University of Bristol, UK

Received 7 February 2019; final version received 9 November 2020; accepted 12 November 2020
Available online 18 November 2020

Abstract

This paper characterizes the optimal information structure in competitive insurance markets with adverse selection. We consider a regulator that assigns ratings to individuals according to their expected costs. Insurers observe these ratings and compete 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-cost types with high-cost types negative assortatively. We provide a simple algorithm that yields the optimal system and examine implications for government regulations of insurance markets.
© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

JEL classification: D81; D60; D86; I13

Keywords: Insurance markets; Adverse selection; Information design

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 question is how

* Corresponding author.
E-mail addresses: daniel.garcia@univie.ac.at (D. Garcia), matan.tsur@bristol.ac.uk (M. Tsur).
[1] We are thankful for the contributions of Roee Teper to this work. We would also like to thank the Editor, Alessandro Pavan, and three referees for their comments and suggestions. We also thank Anton Kolotilin, Alessandro Lizzeri, and Doron Ravid, for their feedback. Financial support from the Austrian Science Fund: P 30922 is gratefully acknowledged.

https://doi.org/10.1016/j.jet.2020.105160
0022-0531/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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 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 characteristics, alleviating adverse selection.

The tension between ex-ante insurance and ex-post participation appears to have substantial welfare consequences.[2] 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.

We consider a market where risk-averse agents buy insurance policies from risk-neutral sellers, as in Akerlof (1970). Each agent privately knows her type, specifying a distribution over her medical expenses and her willingness-to-pay for insurance. Insurers offer full coverage insurance contracts 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 type. Insurers can differentiate agents only by their rating, and thus the rating system determines the information structure in the market. The ratings induce a competitive equilibrium whereby each agent will be able to purchase insurance at a price equal to the average cost within her rating group, conditional on participation.

The regulator designs a rating system in order to maximize the expected social welfare given the market structure. In our baseline model, we assume that the regulator has access to the same information as market participants. This assumption is reasonable in health insurance markets. In Section 4 we discuss several extensions that accommodate imperfect information.

Our main result is that in any such market, there exists a unique optimal rating system. A simple algorithm implements the optimal assignment. In the first step, the agents with the lowest cost are pooled together with a group of agents with the highest cost, up to the point where the average cost of the risk pool equals the willingness-to-pay of the low-cost agents. In the next step, the agents with the second-lowest-cost are pooled together with a group of unmatched agents with the highest-cost types, and so on. The algorithm ends in at most as many steps as the number of types.

We show that the resulting system satisfies three intuitive properties which are both necessary and sufficient to characterize the optimal rating system. First, any optimal rating system induces a market outcome in which all agents trade. Otherwise, revealing the type of those agents who are excluded would lead to a Pareto improvement. We refer to this property as *no exclusion*. Second, any optimal rating system induces a market outcome whereby low-risk agents face a price equal to their willingness-to-pay. We refer to this property as *no rents at the top*. This property is also intuitive since cross-subsidizing across risk types improves ex-ante insurance. If every agent that receives a certain rating is willing to pay more than the equilibrium price, then the regulator should increase the share of high-risk agents in this rating, reducing price dispersion.

[2] See, e.g., Handel et al. (2015) for health insurance; Hendel and Lizzeri (2003) for life insurance; and Finkelstein et al. (2005) for long-term care insurance markets.

These two properties determine the distribution of prices across ratings but do not inform about their composition. The third property, *negative assortative pooling,* postulates that relatively low-cost types cross-subsidize relatively high-cost types. Negative assortative pooling improves ex-ante insurance because it minimizes the proportion of agents assigned to better ratings, reducing price dispersion. These three properties uniquely identify the optimal rating system.

In Section 4 we show that negative assortative assignment is optimal in more general environments, including when the social planner can set taxes and subsidies, when there is imperfect information, and even if insurers can offer partial insurance contracts. The model also provides several insights to assess existing policies. For instance, our results suggest that restricting the use of genetic information is more likely to increase welfare in markets where participation is higher. Similarly, as individuals obtain access to more accurate information about their health, the optimal policy requires prices to be more sensitive to the available data (e.g., progressively lifting restrictions on GINA).

While our main motivation concerns health insurance markets, our results apply to many other settings where the motive for trade is risk sharing, prices are determined competitively, and a social planner can either provide new information or restrict the information that agents can use. These features arise, for instance, when constructing stress tests for banks or designing the information that online labor platforms disclose (Fisman and Luca (2016)).

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, which provides a motive for information disclosure.

There is a small but influential literature on the value of information in competitive markets with adverse selection (see, e.g., Levin (2001), and more recently, Goldstein and Leitner (2015) and Bar-Isaac et al. (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.

This difference has significant implications. In Akerlof's (1970) adverse-selection market, Levin (2001) shows that the efficiency of trade may not increase monotonically as more information is revealed. For instance, new public information, which affects the beliefs of both buyers and sellers, may reduce the gains from trade. However, under quite general conditions, making *any* private signal public always increases the amount of trade. Furthermore, publicly releasing all private information always leads to an efficient outcome. In an insurance market, by contrast, new public information segments the market and thereby affects how the agents share risk (or which agents are pooled together under the same contract). Making private information public creates new opportunities for trade at a wider range of prices, but it also restricts risk sharing. As a consequence, full disclosure is never optimal.

Most related to our paper, Handel et al. (2015) quantitatively study the effect of price discrimination by simulating health insurance exchanges (markets). They focus on two regimes, health-based pricing and community rating, which in our model correspond to a rating system that reveals complete information and no information, respectively. The present paper characterizes the constrained efficient discrimination policy, which generally lies between these extreme policies. Section 3 provides a quantitative illustration.

The present work is also related to the literature on information design and strategic persuasion (Aumann and Maschler (1995); Kamenica and Gentzkow (2011); Bergemann and Morris (2016)). Closest to ours is Bergemann et al. (2015) who study price discrimination by a monopolist. In our model, instead, competition determines prices, total welfare depends on price dispersion, and there is adverse selection.

The present paper is also related to the literature on information design in screening problems (see, e.g., Calzolari and Pavan (2006), and more recently, Dworczak (2017)). In these papers, the designer must elicit private information from the agents and then choose how much to disclose to the market. The present paper abstracts from the elicitation problem and focuses on competitive insurance markets.

Our main result also contributes to a broader literature studying optimal arrangements in risk-sharing environments. Chiappori and Reny (2016) analyze a model of one-to-one matching between two populations of agents. Each agent faces an exogenous income risk, and two agents can match to share this risk. They show that when the agents are ranked by their risk aversion, negative assortative matching is the (generically) unique stable outcome. The present paper considers a very different setting – a competitive market with risk-neutral insurers and risk-averse agents – and shows that the efficient outcome has the same structure. This suggests that in the context of risk-sharing problems, the negative assortative assignment is not only intuitively appealing, but it is also efficient in a larger class of problems.[3]

2. Model

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

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 (see Section 4 for further discussion). The following assumption simplifies the exposition.

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. Assumption 1 holds, for instance, if the distributions can be ordered by FOSD. All results remain true without this Assumption (see Online Appendix).

[3] Negative assortative pooling may arise also in markets with bundled insurance: insurance purchased by a (heterogeneous) group of individuals. Most prominent is the case of family insurance. For instance, Nguyen (2018) shows that bundled insurance helps to mitigate adverse selection in health insurance in Vietnam.

Information. Each agent knows her type. Insurers know only the prior distribution. A regulator designs a rating system which reveals public information about agents' types.[4] For a given set Z, we let $\Delta(Z)$ denote the set of probability distributions over Z.

Definition. A *rating system*, (S, σ), is a set of public signals, or ratings, S and a mapping $\sigma : \Theta \to \Delta(S)$ that assigns a probability distribution over signals to each type $i = 1, 2, \ldots, N$.

Let Σ be the collection of all rating systems. The subset $\Sigma^M \subset \Sigma$ denotes the rating systems in which the cardinality of S is at most M. That is, $(\sigma, S) \in \Sigma^M$ implies that $S = (s_0, ..., s_K)$, $K < M$. Let $\sigma_{ji} = \Pr_\sigma(s_j \mid i)$ denote the probability of rating s_j conditional on type i, and $\sum_{j=0}^{K} \sigma_{ji} = 1$ for all $i \in \Theta$. We define a *risk pool* associated with a rating $s \in S$ as the posterior distribution of types among the population receiving the rating s.

Timing. At the ex-ante stage, the regulator designs a rating system. At the interim stage, agents privately learn their types. Then, public ratings are realized according to the designed system and agents' types. At the ex-post stage, trade occurs, the outcome of the lottery f_i is realized, and consumption takes place.

Market mechanism. Insurers compete over prices and offer insurance contracts conditional on the information that they observe. We focus on the minimum price that achieves the most efficient allocation. Hence, the price associated with signal s_j satisfies

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

In words, $A(t)$ is the set of types willing to accept price t. The equilibrium price of risk pool j equals the 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 such that if $i \in A(t_j)$ and $i' \in A(t_j)$ then $i'' \in A(t_j)$ for all $i < i'' < i'$. This property guarantees that the price is well defined.

Remark. In the case that a rating system generates multiple equilibria, we focus on the minimal price equilibrium, which Pareto dominates any other equilibria. Moreover, the efficient equilibrium can be approximated by another rating system that implements a unique competitive equilibrium.

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 agents with Pareto weights given by the prior distribution (ex-ante Pareto). As we discuss in Section 4, all results follow for any Social Welfare Function that satisfies (interim) Pareto optimality and a preference for Mean-Preserving Contractions of the price distribution (risk-aversion).[5] We restrict attention to rating systems that use countable signals, and hence the optimal rating system solves the following problem[6]:

[4] Equivalently, all information is publicly available but the regulator restricts the information that insurers can use in their contracts.

[5] Interim Pareto is a very weak criterion in our setup. For example, if the ex-ante probability of each type is sufficiently small, any allocation achieving full trade is interim Pareto.

[6] If S is uncountable, replace the second sum with an integral, and proceed to Lemma 1.

$$\sup_{(S,\sigma)\in\Sigma}\sum_{i=1}^{N}\mu_i\sum_{j=0}^{|S|-1}\sigma_{ji}\Big(u(w-t_j)1_{t_j\le\phi_i}+U_i 1_{t_j>\phi_i}\Big) \quad (1)$$
$$s.t.\ t_j=\min_t E\big(\theta\mid i\in A(t),s_j\big)$$

Lemma 1. *An optimal rating system exists. Furthermore, it is without loss of generality to consider rating systems $(S,\sigma)\in\Sigma^N$.*

A formal proof is given in the Appendix. The Lemma follows from two observations. First, if two rating systems induce the same allocation, it is without loss of generality to focus on the rating system with fewer signals. For example, if two signals induce the same price, we can construct another rating system in which both signals are merged and achieve the same allocation. Second, if some type is the healthiest type who participates in two different signals, then merging them into a single signal induces a mean-preserving contraction of the price distribution.[7]

3. Optimal rating system

The first crucial observation is that it is without loss of generality to consider rating systems that implement an allocation where all types are insured. We refer to this property as *no exclusion*.

Lemma 2. *An optimal rating system satisfies no exclusion.*

Proof. Let (S,σ) be optimal. By Lemma 1, we can assume that $(S,\sigma)\in\Sigma^N$. Suppose that $\sigma_{ji}>0$ and $\phi_i<t_j$ for some type $i\in\Theta$, so that type i does not buy insurance following signal s_j. We can construct another rating system, $(\hat{S},\hat{\sigma})\in\Sigma^{N+1}$, that strictly improves. We set $\hat{\sigma}_{ki'}=\sigma_{ki'}$ for all i', $k\ne j$; $\hat{\sigma}_{ji'}=\sigma_{ji'}$ for $i'\ne i$; $\hat{\sigma}_{ji}=0$ and $\hat{\sigma}_{(N+1)i}=\sigma_{ji}$. Notice that under $(\hat{S},\hat{\sigma})$ type i is strictly better-off and all other types are indifferent because the prices associated with all other signals remain the same. Finally, by Lemma 1, there must exist some $(S',\sigma')\in\Sigma^N$ achieving a (weakly) better allocation than $(\hat{S},\hat{\sigma})$. Hence, (S,σ) is strictly worse than (S',σ'), a contradiction. □

The proof is intuitive, if an agent receives signal s_j and does not trade, then assigning this agent to a new signal instead of s_j is a Pareto improvement, because no other price is affected and this agent is strictly better off.

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

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

It will be convenient to keep track of the highest equilibrium price, which we denote by t_0, and its corresponding signal s_0. From Lemmas 1 and 2, we can write the regulator's maximization problem as

[7] As one referee kindly suggested, this result may also be established using an argument based on Caratheodory's Theorem, and is akin to a number of results in the literature (e.g. Lemma 1 in Bergemann et al. (2015)).

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

The regulator chooses a rating system to maximize the ex-ante expected utility, subject to the interim participation and break-even constraints. Our main result shows that there exists a unique optimal rating system that is the outcome of a simple algorithm.

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

***Step** a_m. If $E_{\mu^m}(\theta) \le \phi_m$, then set $\sigma_{0i} = 1 - \sum_{j=1}^{m-1} \sigma_{ji}, \forall i \in \Theta$ and stop. Otherwise, create signal s_m with $\sigma_{mm} = 1$, and $\sigma_{mi} > 0$ **only if** $\forall m > i$, $\sum_{j=1}^{m} \sigma_{jl} = 1$, such that $t_m = E_{\mu^m}(\theta|s_m) = \phi_m$. Proceed to Step b_m.*

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

$$\mu_i^{m+1} = \frac{\mu_i^m(1 - \sigma_{mi})}{\sum_{l=1}^{N} \mu_l^m(1 - \sigma_{ml})},$$

increase m by one (that is, $m = m + 1$), and proceed to Step a_m.

The outcome of the algorithm is a negative assortative assignment. The agents with cost type 1 and a group of agents with the highest cost types are assigned to rating s_1 such that the average cost of the risk pool is ϕ_1. The agents with cost type 2 and a group of agents with the next highest cost types are assigned to rating s_2 such that the average cost of the risk pool is ϕ_2, and so on.

To see how the algorithm achieves this allocation, Step a_1 checks whether the agents are willing to participate at a price equal to the average cost of the population. If $\phi_1 \ge E_\mu(\theta)$, the entire population is assigned to signal s_0 and the algorithm stops. Competition drives the price down to the average cost, $t_0 = E_\mu(\theta)$, and the first-best outcome is achieved. Furthermore, in this case, any rating system that reveals information creates price dispersion, reducing ex-ante welfare. If $E_\mu(\theta) > \phi_1$, agents of type 1 will not participate at this price. In this case, agents of type 1 and a group of sickest agents are assigned to signal s_1, up to the point where the average cost of the risk pool is ϕ_1. The algorithm moves to Step b_1, where the agents in signal s_1 are removed, and the process continues with the residual population.

Fig. 1 illustrates the case where $\Theta = \{1, 2, 3, 4, 5\}$ and the first-best outcome is not feasible, $E_\mu(\theta) > \phi_1$. In Step a_1, the agents of type 1 are assigned to rating s_1 in addition to all the agents of type 5 and a fraction of the agents of type 4. Competition drives the price down to the average cost of the risk pool, $t_1 = \phi_1$. Notice that the insurers expected profit from type 1 agents cross-subsidize the expected losses from type 4 and type 5 agents (the areas A4 and A5 in Fig. 1):

$$\mu_1(t_1 - \theta_1) = \underbrace{\mu_5(\theta_5 - t_1)}_{A_5} + \underbrace{\mu_4\sigma_{14}(\theta_4 - t_1)}_{A_4}.$$

The algorithm moves to Step b_1, which removes the agents in signal s_1 from the population. The process continues with the residual population with the distribution of types given by,

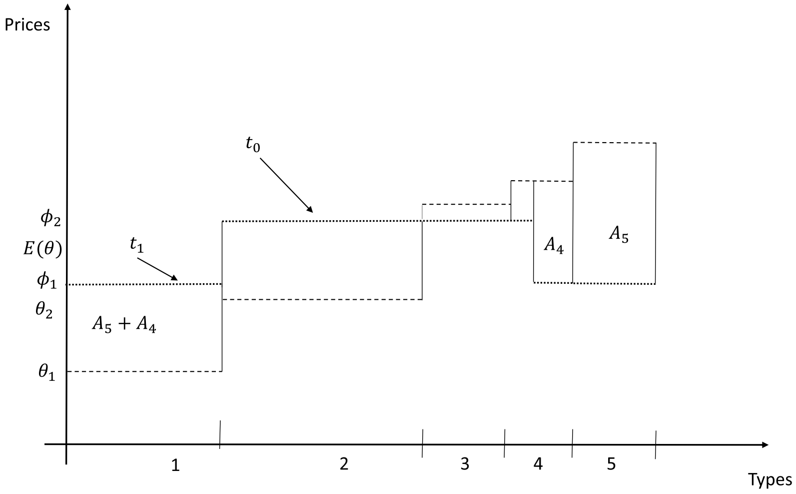

Fig. 1. The optimal rating system.

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

In Step a_2, the algorithm checks whether type 2 agents participate at a price equal to the average cost of the residual population, $E_{\mu^2}(\theta) = \sum_{i=1}^{5} \mu_i^2 \theta_i$. Notice that Fig. 1 depicts a case where $E_{\mu^2}(\theta) \leq \phi_2$. Therefore, the entire residual population is assigned to signal s_0, the associated price is $t_0 = E_{\mu^2}(\theta)$, and the algorithm stops.

The proof of Theorem 1 proceeds in two steps. First, we identify three properties that are necessary and jointly sufficient to characterize the optimal rating system. Second, we show that the output of the algorithm is the unique rating system that satisfies these properties.

The first property is *no exclusion*, which we previously discussed.

Definition. A rating system (S, σ) satisfies *no rents at the top* if whenever $i = \min\{l : \sigma_{jl} > 0\}$ and $t_j < t_0$, then $t_j = \phi_i$.

In other words, the participation constraint of the healthiest type receiving a certain rating with positive probability is binding, except for the rating s_0.

Lemma 3. *An optimal rating system satisfies no rents at the top.*

Proof. Suppose, for a contradiction, that there exists an optimal rating system (S, 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, $(\hat{S}, \hat{\sigma})$: $\hat{\sigma}_{jl} = \sigma_{jl} + (1 - \beta)\sigma_{0l}$, $\hat{\sigma}_{0l} = \beta\sigma_{0l}$, $\hat{\sigma}_{j'l} = \sigma_{j'l}$ 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 $\hat{t}_0 = t_0$, $\sum_{i=1}^{N} \sigma_{0i}\mu_i > \sum_{i=1}^{N} \hat{\sigma}_{0i}\mu_i$,

$$\sum_{i=1}^{N} \sigma_{0i}\mu_i t_0 + \sum_{i=1}^{N} \sigma_{ji}\mu_i t_j = \sum_{i=1}^{N} \hat{\sigma}_{0i}\mu_i \hat{t}_0 + \sum_{i=1}^{N} \hat{\sigma}_{ji}\mu_i \hat{t}_j,$$

and $t_j < \hat{t}_j < t_0$. It follows that the distribution of prices under $(\hat{S}, \hat{\sigma})$ is a mean-preserving contraction of that under (S, σ). □

The idea of the proof is that if all agents receive positive rents in a certain rating, which is not the worst, then we can "move" some of the sickest agents from the worst rating to the better one without violating the participation constraints (because the average price equals the average cost of the participating agents).

Definition. A rating system (S, σ) satisfies *negative assortative pooling*, 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 $\min\{k \neq i : \sigma_{jk} > 0\} \geq \max\{k : \sigma_{j'k} > 0\}$.

This property 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. In other words, negative assortative pooling requires that every type in a better pool, aside of the healthiest, is sicker than any type in a worse pool (where the ranking of pools is determined by the cost of its healthiest type).

Lemma 4. *The optimal rating system satisfies negative assortative pooling.*

Proof. The proof relies on the following claim.

Claim 1. *Let $(S, \sigma) \in \Sigma^N$ be an optimal rating system. If $t_0 > t_j > \phi_i$, and $\sigma_{ji} > 0$, then $\theta_i \geq t_0$ (with strict inequality unless $\sigma_{0k} = 0$ for all $k \neq i$).*

The proof of Claim 1 is given in the Appendix. Claim 1 states that if type i is assigned to signal s_j, except s_0, with positive probability and receives rents in s_j, then type i would be cross-subsidized in any other risk pool. From Claim 1 and Lemma 3 we know that for every signal s_j (except s_0), there exists a unique healthiest type i for which $t_j = \phi_i$, $\sigma_{ji} = 1$, and for all other types $l \neq i$, if $\sigma_{jl} > 0$ then $t_j < \theta_l$.

Assume the rating system does not satisfy negative assortative pooling. That is, there exist two signals j, j' with $t_j < t_{j'}$ and two types $l < l'$, who are not the healthiest types in either of these signals; and it holds that $\sigma_{jl} > 0$ and $\sigma_{j'l'} > 0$. We can improve this allocation by exchanging only types l and l' as follows.

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

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

The key observation is that by Claim 1, we have $\theta_{l'} > \theta_l \geq t_0 \geq t_{j'} > t_j$ and, thus, $G > 0$. Moreover, $\theta_l > t_{j'}$ because if $t_0 = t_{j'}$, then since there are at least two types in the support of s_0, then by Claim 1 $\theta_l > t_0$. Consider then $(S, \hat{\sigma})$ such that $\hat{\sigma}_{ki} = \sigma_{ki}$, for all $i \neq l, l'$ and for all 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

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

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 price distribution under (S, σ) is a mean-preserving spread of that under $(S, \hat{\sigma})$. □

The proof is based on two ideas. First, by Claim 1, we know that in each risk pool, except s_0, there exists a unique type that cross-subsidizes all other types in the pool. Second, if negative assortative pooling fails, there exists a swap between two cross-subsidized types in two different pools, such that the lowest of the two prices is kept constant and the highest drops. In the resulting allocation no participation constraint is violated, and price dispersion is reduced.

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. Furthermore, the optimal rating system is unique.*

Proof. Lemmas 2-4 establish the necessary conditions. For sufficiency, we show that there exists a unique rating system satisfying all three properties. Suppose that (S, σ) and $(\hat{S}, \hat{\sigma})$ 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 $t_j = \phi_i$. Since each of the first j prices must correspond to a unique type, it must be that $t_j = \phi_i$ and $i \geq j$. If $i > j$, there must exist type $i' < i$ and $\phi_{i'} < t_j = \phi_i$. This is only possible if either $\sigma_{i'j'} > 0$ for $j' < j$ which contradicts negative assortative pooling or $\sigma_{i'j'} > 0$ for $j' > j$ which contradicts no trade $(t_{j'} > \phi_{i'})$.
Furthermore, by negative assortative pooling, $\sigma_{ji} = \hat{\sigma}_{ji}, \forall j \leq k$. On the one hand, we have that:

$$\sum_{i=1}^{N} \mu_i \sigma_{0i} t_0 = \sum_{i=1}^{N} \mu_i \hat{\sigma}_{0i} \hat{t}_0 + \sum_{j \geq k+1} \sum_{i=1}^{N} \mu_i \hat{\sigma}_{ji} \hat{t}_j.$$

On the other hand, we have that:

$$\sum_{i=1}^{N} \mu_i \hat{\sigma}_{0i} \hat{t}_0 + \sum_{j \geq k+1} \sum_{i=1}^{N} \mu_i \hat{\sigma}_{ji} \hat{t}_j = \sum_{i=1}^{N} \mu_i \hat{\sigma}_{0i} \hat{t}_0 + \sum_{j \geq k+1} \sum_{i=1}^{N} \mu_i \hat{\sigma}_{ji} \phi_j$$
$$> \phi_{k+1} \left(\sum_{i=1}^{N} \mu_i \hat{\sigma}_{0i} + \sum_{j \geq k+1} \sum_{i=1}^{N} \mu_i \hat{\sigma}_{ji} \right) = \phi_{k+1} \sum_{i=1}^{N} \mu_i \sigma_{0i} \geq \sum_{i=1}^{N} \mu_i \sigma_{0i} t_0,$$

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 has a strictly positive probability; the next follows by definition, and the last inequality follows from the fact that σ is feasible and satisfies the participation constraints. Hence, we have a contradiction. □

All that is left to show now is that the output of the algorithm satisfies these properties.

Proof of Theorem 1. The algorithm yields a unique rating system in at most N steps. By construction, the rating system satisfies no exclusion (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

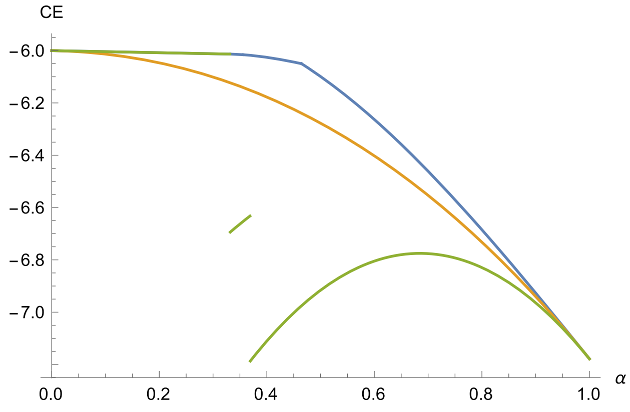

Fig. 2. Ex-ante Certainty Equivalent as a function of α for different regimes. The green line represents no information, the orange line represents full information and the blue line is the optimal rating system. (For interpretation of the colors in the figure, the reader is referred to the web version of this article.)

pooling (if for some j, $\sum_{k\leq j}\sigma_{ki} < 1$, $\sigma_{ji'} = 0$ for all $j < i' < i$). By Theorem 2, the algorithm yields the unique optimal rating system. □

3.1. Illustrative example

To illustrate the construction of the optimal rating system, let us consider the setting of Handel et al. (2015). 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. Following our baseline model, we assume that the two policies correspond to full coverage and no coverage.[8] Therefore, $\theta_i = E(x|\epsilon_i)$ and $\phi_i = E(x \mid \epsilon_i) + \gamma(1-\alpha)^2 Var(\epsilon_a) \equiv \theta_i + \Delta$.

The parameter α measures how well private information predicts future costs. Fig. 2 depicts the ex-ante welfare as a function of α under three rating systems.

1. **Full information** (health-based pricing). Each type pays her actuarially fair price $t_i = \theta_i$. 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 ($\alpha = 1$).
2. **No information** (community rating). The price equals the average cost of the participating agents. If $\alpha < 0.33$, then we have that $\theta_1 + \Delta \geq E_\mu(\theta)$ and all agents participate at the price $t = E_\mu(\theta)$. Otherwise, type 1 does not participate and we either have partial unraveling (if $\theta_2 + \Delta \geq E_\mu(\theta \mid i \neq 1)$) or complete unraveling. The green line depicts the (ex-ante) certainty equivalent under no information.
3. **Optimal rating system**. There are four possible configurations, depending on the value of Δ.
   - If $\Delta < \Delta_1$, no information is optimal.
   - If $\Delta \in [\Delta_1, \Delta_2)$, there are two ratings: type 1, type 3, and a fraction of type 2 receive rating s_1 such that $t_1 = \theta_1 + \Delta$. The rest of type 2 receive rating s_0 and pay $t_0 = \theta_2$.
   - If $\Delta \in [\Delta_2, \Delta_3)$, there are two ratings: type 1 and a fraction of type 3 receive rating s_1 such that $t_1 = \theta_1 + \Delta$. The remaining types receive rating s_0 and pay $t_0 \in [\theta_2, \theta_2 + \Delta)$.

8 See Appendix 1.5 for a discussion.

- If $\Delta \geq \Delta_3$, there are three different ratings: type 1 and a fraction of type 3 receive rating s_1 such that $t_1 = \theta_1 + \Delta$. Type 2 and a fraction of type 3 receive rating s_2 such that $t_2 = \theta_2 + \Delta$. The remaining types receive rating s_0 and pay $t_0 = \theta_3$.

The blue line depicts the (ex-ante) certainty equivalent under the optimal rating system.

This example also suggests that there is a clear relationship between the level of idiosyncratic risk and the efficiency of the market under the optimal rating system. More risky environments increase the wedge between the expected cost and the willingness-to-pay, which allows the regulator to cross-subsidize across types more efficiently. For example, consider two environments, M_1 and M_2, with associated distributions of medical expenses f_i^1 and f_i^2, and suppose that f_i^2 is a mean-preserving spread of f_i^i for all $i \in \Theta$. It follows that with no information, the set of types who trade is (weakly) larger in environment M_2 because every agent is willing to pay more for insurance, but their expected costs are the same. In addition, the optimal rating system in environment M_2 requires (weakly) fewer ratings and achieves higher expected welfare.[9] As we show in the Online Appendix, similar comparative statics emerge in the case of asymmetric information.

From a policy perspective, these findings suggest that regulations that limit price discrimination, such as GINA, should be less strict in markets with limited participation. Likewise, they imply that as individuals have access to more accurate information about their health, the optimal regulation requires prices to be more sensitive to the available data (e.g., progressively lifting restrictions on GINA).

4. Discussion

We have analyzed the problem faced by a regulator that provides information to insurers, and this information enables them to offer different contracts to individuals with different characteristics. We showed that the properties of no exclusion, no rents at the top, and negative assortative pooling uniquely characterize the optimal rating system. In this section, we discuss several extensions of the model and show that our main result, negative assortative pooling, continues to hold.

Partial insurance. We have assumed that insurers offer full coverage contracts. If insurers can offer partial insurance, they can cream-skim healthy agents with a cheaper contract that provides lower coverage. Furthermore, if every quantity-price pair contract is feasible, as in Rothschild and Stiglitz (1976), the only equilibrium outcome, when it exists, is fully separating. In this scenario, there are no gains from building risk pools which will unravel by competitive screening. Therefore, the optimal rating system perfectly reveals each type to eliminate the inefficiencies associated with screening.

In the Online Appendix, we analyze a market where insurers can offer contracts with partial coverage and the contract space is restricted. We first consider a market where all the feasible contracts provide the same (partial) level of coverage. The equilibrium is similar to the case of full insurance, but now the agents face residual risk, creating a wedge between prices and consumption. As a result, a mean preserving contraction of the price distribution need not imply a

[9] In each iteration of the optimal algorithm, the residual population under M_2 is a subset of that under M_1, and, therefore, the algorithm must stop in fewer steps. Further, the optimal rating under M_1 is feasible under M_2, so the latter must achieve higher welfare.

mean preserving contraction of the distribution of consumption profiles, and, hence, may not be desirable to the regulator. Nevertheless, under some technical assumptions on the utility function and the distribution of risk, we show that the optimal rating system still satisfies negative assortative pooling (see Proposition 1 in the Online Appendix).[10]

We then extend the analysis to an environment in which insurers can offer contracts with one of two exogenously given coverage levels. The contracts differ in the coverage level and prices are determined competitively.[11] As in Handel et al. (2015), we consider a Riley equilibrium (see Riley (1979)), which always exists, and the equilibrium allocation necessarily involves some cross-subsidization across types. The agents' participation constraints are now endogenous, and the regulator has to make sure that insurers cannot profitably cream-skim the agents in the high-coverage contract. We first show that under the optimal rating system, all agents assigned to a rating choose the same contract (quantity-price). We then show that under the same conditions of Proposition 1, negative assortative pooling holds within each coverage level.

Together, these two results demonstrate that negative assortative pooling is the key feature of an optimal rating system in a broad class of contractual environments, provided that there is competition and the equilibrium is not fully separating.

Information asymmetries. We have assumed that the regulator has access to the same information as the agents. We contend that this assumption is reasonable in insurance markets in which the availability of big data (medical records and perhaps genetic data), machine learning, and scientific innovations in predictive medicine are improving the accuracy of risk estimates.[12] Nevertheless, our analysis can be adapted and applied to different informational environments.

First, the regulator may be able to better predict the medical costs of different illnesses than the agents. In this case, as we show in Proposition 5 in the Online Appendix, a simple modification of the algorithm provided in Theorem 1 delivers the optimal rating system. This rating system satisfies no exclusion, no rents at the top, and negative assortative pooling, whereby agents with the lowest willingness-to-pay (based on their prior) are pooled with the agents that have the highest expected costs (based on the regulator's information).

Second, if agents have better information about their medical expenses, the optimal rating system need not satisfy no exclusion (see Example B.1 in Section 1.4 of the Online Appendix). To make some progress, we assume that the regulator knows the expected cost of each type but does not know their willingness-to-pay. As we show in Proposition 6 in the Online Appendix, the optimal rating system still satisfies negative assortative pooling and (a version of) no rents at the top.

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

[10] The optimal rating system satisfies no exclusion, but need not satisfy no rents at the top.

[11] For instance, in the exchanges set up by the ACA, health insurance plans appear in 4 metal categories, Bronze, Silver, Gold, and Platinum, differing in the degree of coverage.

[12] We abstract from the problem of soliciting private information from the agents and then making this information available to the market. This issue has been studied in Calzolari and Pavan (2006), and more recently, in Dworczak (2017). Notice that if the information is verifiable, every agent would be willing to reveal their private information and participate in the mechanism. The case of unverifiable information is left for future work.

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 satisfy no exclusion is Pareto dominated, and thus the optimal rating system under any social welfare function satisfies no exclusion (Lemma 2). 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 all agents have the same wealth level at the ex-ante stage. An alternative assumption is that individuals' preferences satisfy CARA and health costs are uncorrelated with wealth. In such an environment, if the regulator cannot discriminate based on wealth, any rating system that satisfies no exclusion with lower price dispersion is welfare-improving, and our results go through unchanged.[13]

In practice, health insurance policy is often used as a safety net with the (implicit) aim of reducing inequality and poverty, and in many cases insurance subsidies are means-tested. In Section 1.5 of the Online Appendix, we consider an extension in which the regulator observes both costs and wealth levels and designs the rating system to cross-subsidize in both dimensions. We show that the optimal rating system satisfies no exclusion and gives no rents to a subset of healthiest agents.

Market structure. If the regulator faces a monopoly, prices are determined by individual willingness-to-pay (rather than expected cost). Bergemann et al. (2015) study market segmentation (rating systems) without adverse selection and with risk-neutral buyers. They show that any feasible and individually-rational payoff vector is an equilibrium outcome for some information structure. Adverse selection introduces an additional role for information (guaranteeing efficient trade), while risk-aversion generates a strict ranking over information structures and a trade-off for the regulator: reducing seller's surplus may require a more disperse distribution of risk. We leave these issues for future research.

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.[14] A natural question to ask is how the optimal policy combines information design and fiscal policies.

[13] Notice that CARA ensures that the willingness-to-pay does not depend on wealth and, thus, the regulator knows each agent's ϕ_i. Independence between health and wealth implies that a mean-preserving contraction is welfare improving. In fact, the same logic applies if lower medical costs are associated with higher wealth.

[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.

To address this question, suppose that, as in our model, the regulator designs a rating system, and then competition determines the prices of the contracts offered to each 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 planner optimally 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 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 (in the sense that 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 (see Section 1.2). Proposition 3 shows that there exists a threshold level such that if the tax system is more efficient than this threshold, then the regulator should only use taxes, while if it is less efficient, the properties presented in Theorem 2 characterize the optimal rating system.

Appendix A. Omitted proofs

Proof of Lemma 1. Since multiple rating systems may achieve the same allocation, we focus on the one with the minimum number of signals. Therefore, it is without loss of generality to consider only rating systems $(S, \sigma) \in \Sigma$ in which each signal induces a different price. To see this notice that if two signals implement the same price, we can always merge them into one signal without changing the allocation.

Let us now define the auxiliary problem

$$\sup_{(S,\sigma)\in\Sigma^N} \sum_{i=1}^{N} \mu_i \sum_{j=0}^{N-1} \sigma_{ji} \Big(u(w - t_j) 1_{t_j \le \phi_i} + U_i 1_{t_j > \phi_i} \Big) \tag{2}$$
$$s.t.\ t_j = \min_t E_j \big(\theta \mid i \in A(t) \big)$$

Notice that a solution to (2) exists because Σ^N is compact and the payoff function is upper-hemi-continuous.[15] We will now argue that the values of problem (2) and problem (1) are equal. In particular, we show that for any rating system $(S, \sigma) \in \Sigma \setminus \Sigma^N$, there exists another rating system $(S', \sigma') \in \Sigma^N$ that achieves an allocation with a weakly higher value.

Let us first assume that $| S | < \infty$. The first observation is that there must exist a type $i \in \Theta$ such that i is the healthiest type in the support of (at least) two different signals, s_k and s_j, who is willing to trade. Formally, $i = \min\{i' : i' \in A(t_k) \wedge \{\sigma_{ki'} > 0\}\} = \min\{i' : i' \in A(t_j) \wedge \{\sigma_{ji'} > 0\}\}$. This is a direct consequence of the pigeonhole principle because there are only N types and at least $N + 1$ signals.

Without loss of generality, assume that $t_j > t_k$. Consider then $(S', \sigma') \in \Sigma^{|S|-1}$ with $\sigma'_{li'} = \sigma_{li'}$ for all $l \neq k, j$, $i' \in \Theta$, $\sigma'_{ki} = \sigma_{ki'} + \sigma_{ji'}$. In words, the new rating system merges signals k, j

[15] Notice that $E_j(\theta \mid i \in A(t))$ is u.h.c. in σ and the expected utility function is continuous.

into a single signal. Since i is the healthiest type in the support of this new signal s'_k, for all $i' : \sigma'_{ki'} > 0$, we have that $\phi_{i'} \geq \phi_i$. But since i was willing to participate in both signals, every type in the support of s'_k who traded before (in either s_k or s_j) is willing to trade now. It follows that $t_k \geq t'_k \geq t_j$. Because the average price always equals the average cost of the agents who participate, the allocation under (S', σ') implements the same average price with lower price dispersion and, therefore, it is a mean-preserving contraction of the allocation under (S, σ).

We finally consider the case where $| S |= \infty$. For each type i consider all the signals s_k such that $i = \min\{i' : i' \in A(t_k) \wedge \{\sigma_{ki'} > 0\}\}$. Let S_i be the collection of all such signals and let $\sigma_j(i) = \sum_{k \in S_i} \sigma_{kj}$. Let (S', σ') with $S' = \{s_1, ..., s_N\}$ and $\sigma'_{ij} = \sigma_j(i)$ for all i and all j. By construction, the probability of trade of each type is the same in both (S, σ) and (S', σ') and $t'_i = \sum_{k \in S_i} \sum_{j \geq i} \frac{\sigma_{kj}\mu_j}{\sum_{j \geq i} \sigma_{kj}\mu_j} \theta_j$. Hence, (S', σ') induces a mean-preserving contraction of the distribution under (S, σ). □

Proof of Claim 1. 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_{N+1}. There are four cases to consider.

Case 1: $t_0 > \theta_i > t_j$. We construct rating system $(\hat{S}, \hat{\sigma}) \in \Sigma^{N+1}$ 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. $\hat{\sigma}_{(N+1)l} = \gamma\sigma_{0l}$ for $l \neq i$ and $\hat{\sigma}_{(N+1)i} = \delta\sigma_{ji} + \gamma\sigma_{0i}$;
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 $N + 1$ (Steps 2 and 3); we move type i from rating j (Step 1) to rating $N + 1$ (Step 3); and we keep everyone else in the same rating (Step 4). By construction $t_0 = \hat{t}_0$ and,

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

For any $\delta > 0$, there exists some $\gamma(\delta)$ such that for all $\gamma < \gamma(\delta)$, we have that $\theta_i < \hat{t}_{N+1} \leq \phi_i$. 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 $t_0 = \hat{t}_0 > \hat{t}_{N+1} > t_j = \hat{t}_j$ and for all $k \neq j$, $\hat{t}_k = t_k$. Since the price distributions under (S, σ) and $(\hat{S}, \hat{\sigma})$ have the same mean, the former induces a mean-preserving contraction on the price distribution. Moreover, by Lemma 1, there must exist $(S', \sigma') \in \Sigma^N$ that outperforms $(\hat{S}, \hat{\sigma})$.

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 $(S, \hat{\sigma})$ 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 $(\hat{S}, \hat{\sigma})$ 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 $(\hat{S}, \hat{\sigma}) \in \Sigma^{N+1}$ with $\hat{\sigma}_{0l} = (1-\gamma)\sigma_{0l}$ for all l, $\hat{\sigma}_{ji} = (1-\delta)\sigma_{ji}$, $\hat{\sigma}_{ji} = (1-\delta')\sigma_{ji'}$, $\hat{\sigma}_{(N+1)i} = \delta\sigma_{ji} + \gamma\sigma_{0i}$, $\hat{\sigma}_{(N+1)i'} = \delta'\sigma_{ji'} + \gamma\sigma_{0i'}$. We can choose δ and δ' such that $\hat{t}_j = t_j$ and for γ sufficiently small, $\min\{\phi_i, t_0\} > t_{N+1} > t_j$, which implies that $(\hat{S}, \hat{\sigma})$ induces a mean-preserving contraction of the price distribution (and, therefore, there must exist some $(S', \sigma') \in \Sigma^N$ that strictly improves on (S, σ)). 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. □

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jet.2020.105160.

References

Akerlof, G.A., 1970. The market for "lemons": quality uncertainty and the market mechanism. Q. J. Econ. 84 (3), 488–500.

Aumann, R.J., Maschler, M., 1995. Repeated Games with Incomplete Information. MIT Press.

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

Bergemann, D., Brooks, B., Morris, S., 2015. The limits of price discrimination. Am. Econ. Rev. 105 (3), 921–957.

Bergemann, D., Morris, S., 2016. Information design, Bayesian persuasion, and Bayes correlated equilibrium. Am. Econ. Rev. 106 (5), 586–591.

Calzolari, G., Pavan, A., 2006. On the optimality of privacy in sequential contracting. J. Econ. Theory 130 (1), 168–204.

Chiappori, P.-A., Reny, P.J., 2016. Matching to share risk. Theor. Econ. 11 (1), 227–251.

Dworczak, P., 2017. Mechanism design with aftermarkets: cutoff mechanisms. Available at SSRN 2859206.

Epstein, L.G., Segal, U., 1992. Quadratic social welfare functions. J. Polit. Econ. 100 (4), 691–712.

Finkelstein, A., McGarry, K., Sufi, A., 2005. Dynamic inefficiencies in insurance markets: evidence from long-term care insurance. Am. Econ. Rev. 95 (2), 224–228.

Fisman, R., Luca, M., 2016. Fixing discrimination in online marketplaces. Harv. Bus. Rev. 94 (12), 88–95.

Goldstein, I., Leitner, Y., 2015. Stress Tests and Information Disclosure. Discussion paper.

Handel, B., Hendel, I., Whinston, M.D., 2015. Equilibria in health exchanges: adverse selection versus reclassification risk. Econometrica 83 (4), 1261–1313.

Hendel, I., Lizzeri, A., 2003. The role of commitment in dynamic contracts: evidence from life insurance. Q. J. Econ. 118 (1), 299–328.

Hirshleifer, J., 1971. The private and social value of information and the reward to inventive activity. Am. Econ. Rev. 61 (4), 561–574.

Kamenica, E., Gentzkow, M., 2011. Bayesian persuasion. Am. Econ. Rev. 101 (6), 2590–2615.

Levin, J., 2001. Information and the market for lemons. Rand J. Econ. 32 (4), 657–666.

Nguyen, A., 2018. Household Bundling to Reduce Adverse Selection: Application to Social Health Insurance. Working Paper, (2018-28). Becker Friedman Institute for Research in Economics.

Rawls, J., 2009. A Theory of Justice. Harvard University Press.

Riley, J.G., 1979. Informational equilibrium. Econometrica 47 (2), 331–359.
Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Q. J. Econ., 629–649.
Schlee, E.E., 2001. The value of information in efficient risk-sharing arrangements. Am. Econ. Rev. 91 (3), 509–524.
Sen, A., 1977. On weights and measures: informational constraints in social welfare analysis. Econometrica, 1539–1572.