Market Structure and Adverse Selection*

Dakang Huang† and Christopher Sandmann‡

February 2024

Abstract

This paper presents a unified perspective on multi-contracting in competitive markets afflicted by adverse selection. We subsume the two polar cases of exclusive and nonexclusive competition in the literature by introducing the concept of a *market structure*, i.e., a trading rule that specifies the subset of sellers with whom buyers can jointly trade. Our analysis reveals that modifying the market structure alone can alleviate perceived inefficiencies in market-based mechanisms. Normative results single out the "1+1" market structure, where buyers can trade with at most one seller from each of two subgroups. We prove that if adverse selection is severe, adopting the "1+1" market structure Pareto-improves upon the initial exclusive equilibrium allocation in an unregulated competitive market. When requiring, in addition, that all purchases include a contract exceeding a minimum quantity, the resulting equilibrium allocation is second-best efficient.

Keywords adverse selection, competitive markets, common values, nonexclusive competition

JEL D82, D86, G22, I13

*We are very grateful for the guidance, support and invaluable suggestions from Andrea Attar, Thomas Mariotti and François Salanié. We also thank Eloisa Campioni, Daniel Gottlieb, Piero Gottardi, Humberto Moreira and Chew Soo Hong as well as audiences at Stony Brook GTF 2022, Warwick Icebreaking 2022, City Theory 2023, Oxford SETF 2023, LSE, Mannheim, Essex, Toulouse, Peking University, Renmin University, Wuhan University, CMGTA 2023, AMES 2023 Beijing, AMES 2023 Singapore, ESEM 2023, EAYE 2023, PSE-Science Po and SAET 2024. Financial support from Takuro Yamashita's ERC grant 714693 is gratefully acknowledged.

†Southwestern University of Finance and Economics dakang.huang.c@gmail.com

‡London School of Economics c.sandmann@lse.ac.uk

1 Introduction

Since Akerlof (1970) first showed that markets afflicted by adverse selection[1] can unravel, economists have widely emphasized the need to regulate and in some instances outright replace decentralized exchange with a mechanism of centralized allocation.[2] Insurance markets are a particular case in point. Adverse selection prevails in that high levels of coverage are exceedingly sought by buyers with greater innate risk. Competitive allocations are seen to provide less insurance than would be optimal even when taking into account resource and information constraints. In response, many countries (e.g., UK, Brazil, India) have sought to bypass adversely selected markets by establishing publicly funded health systems.

We here argue that the presumption that market-based mechanisms are inefficient is premature and largely premised on the assumption that competition is *exclusive*: buyers observe the many price-quantity contracts posted by competing sellers; they can, however, buy coverage from a single insurance provider only. In this paradigm, based on Rothschild and Stiglitz (1976),[3] quantity distortions are significant (see Gottlieb and Moreira (2023)) because incumbent sellers fear cream-skimming—competitors targeting low-risk buyers while leaving high-risk buyers with incumbent sellers.

More recently, Attar et al. (2011, 2014, 2021, 2022) dispensed with the assumption that sellers propose exclusive contracts. They show that if buyers can purchase arbitrarily many contracts from different sellers, a paradigm referred to as *nonexclusive* competition, quantity distortions of low-risk buyers can be alleviated while risk premia on high-risk buyers are lower. Stiglitz et al. (2020); Asriyan and Vanasco (2023) echo these findings in related settings. Despite these desirable properties, normative policy prescriptions remain elusive. Exclusive and nonexclusive equilibria are not always Pareto-ranked; pure strategy equilibria do not always exist.[4]

In this paper, we adopt a unified perspective on multi-contracting. We view *exclusive* and *nonexclusive* competition as the two polar cases within the much larger set of *market structures*. A market structure specifies which sellers a buyer can jointly trade with. Examples include market structures where policy makers cap the number of policies that buyers can

[1]A market is adversely selected if the least desirable informed trading partners are also those most eager to trade.

[2]Low-risk agents must be incentivized or compelled to cross-subsidize high-risk agents through subsidies or mandates Einav et al. (2010); costly pre-market trades by the planner must change the prevailing risk profile of the market Tirole (2012), or costly verification must first restore the symmetry of risk via risk adjustments for trade to resume Glazer and McGuire (2006).

[3]A version of their model allowing for withdrawal of contracts was concurrently studied by Wilson (1977).

[4]Nonexistence of a pure-strategy equilibrium under exclusive competition is well-explored since Rothschild and Stiglitz (1976). See Mimra and Wambach (2014) for a survey. Azevedo and Gottlieb (2017) propose a perturbation-based equilibrium concept that restores equilibrium existence. The Pareto-efficient allocation among all zero profit separating allocations (cf. Riley (1979); Engers and Fernandez (1987)) prevails. Ania et al. (2002) establish this finding in an evolutionary analysis where sellers only locally deviate from incumbent contracts. This prediction is not anonymous, however, if pooling Pareto dominates the separating allocation (cf. Wilson (1977); Miyazaki (1977); Netzer and Scheuer (2014)). In a recent contribution, Farinha Luz (2017) proved existence of a unique mixed strategy equilibrium and characterized it. Attar et al. (2014) show that under nonexclusive competition a pure-strategy equilibrium typically fails to exist when sellers post menus.

hold, or an industry association that prohibits multiple contracting only within the subset of member firms. In practice, the market structure is a deliberate choice that usually falls under the purview of legislative or executive regulation. US health exchanges are an example of institutionalized exclusive competition; life insurance and annuity markets serve as examples of purely nonexclusive competition.[5] In light of the size of these markets, understanding how the regulator's choice of the market structure affects the competitive equilibrium, and consequently welfare, seems imperative.

Our game-theoretic analysis builds on the canonical adverse selection model in a competitive insurance market (refer to Rothschild and Stiglitz (1976); Wilson (1977) for results on exclusive competition and Attar et al. (2014, 2021) for results on nonexclusive competition). In this model, sellers compete by offering price-quantity contracts. Insurance demand stems from high- and low-risk buyers who are indistinguishable to sellers. Exclusive trading, by design, is bound to be inefficient. As demonstrated by Crocker and Snow (1985), the set of second-best efficient (i.e. incentive compatible and budget balanced) allocations involves both pooling and separating contracts. Yet, simultaneous pooling and separation was previously thought of as unattainable via a market mechanism. Rothschild and Stiglitz (1976) show that pooling is deterred under exclusive competition, whereas Attar et al. (2014) show that low-risk separation is deterred under nonexclusive competition. A preliminary result (see Claim 1) generalizes the set of market structures under which pooling can never occur: when sellers can offer exclusive contracts, cream-skimming deviations that uniquely target low-risk buyers destabilize any equilibrium involving pooling. Consequently, the unique equilibrium candidate for any competitive partially exclusive market structure is the separating Rothschild-Stiglitz allocation which often entails excessive rationing of low-risk buyers. To possibly improve welfare, emphasis must be placed on market structures that prohibit sellers from offering (what can be viewed as anti-competitive) exclusive contracts.

On normative grounds, our analysis singles out the "1+1" market structure. Here, sellers are divided into two subgroups, and buyers can trade with at most one seller from each subgroup. In effect, trade is exclusive within groups and nonexclusive across groups. Interestingly, this market structure largely replicates the public-private two-tier health insurance system in France, with the one difference that the public *securité sociale* is now offered by a competitive private health insurance market.[6] This structure suggests an equilibrium in which all buyers purchase the same pooling contract in group 1, and different separating contracts in group 2.

What are the advantages of restricting nonexclusive competition? More specifically, why

[5]The US Affordable Care Act explicitly requires qualified health plans offered on individual or group health exchanges to limit expected out-of-pocket expenses to at most 40% (cf. Affordable Care Act, Sec. 1302(d)), making it impossible to offer top-up plans.

[6]In France, all individuals subscribe to the public *securité sociale*—a pooling contract—and top-up to reduce co-payments by purchasing additional private health insurance called *mutuelle*. Recent work by Einav and Finkelstein (2023) endorses features of this system also for adoption in the United States. Relative to these policy ideas, our work shows that by regulating exclusivity covenants, a regulator can achieve the same allocation via private insurance markets.

stipulate that contracts within the same group cannot be purchased jointly? In a fully nonexclusive market, any contract with a sufficiently low price per unit of coverage is liable to attract high-risk buyers. In particular, this happens when inactive sellers *pivot* on existing, more attractively priced low-risk contracts. By offering complementary coverage, the combined low-risk and pivoting contracts replicate the initial coverage of the high-risk type but at a lower price. The implication is that "Pooling + Separation" may be technically feasible, but does not occur in equilibrium because any contract solely targeting low-risk buyers will inadvertently also attract high-risk buyers, thereby increasing its cost. The economic virtue of the "1+1" market structure is that it shields sellers serving low-risk buyers from pivoting deviations. Unlike under fully nonexclusive competition, the "1+1" market structure imposes that buyers have only two buy options. If a high-risk buyer wishes to purchase a contract that pivots on the low-risk contract in group 2, he must purchase said contract in group 1. This entails foregoing the contract previously bought in group 1. If this contract is pooling, it is more attractively priced than any profitable pivoting contract. In effect, pivoting deviations will fail to attract high-risk buyers and break even if pooling is sufficiently large.

Our first main result formulates necessary conditions that must hold in equilibrium under any given *never exclusive* market structure (see Theorem 1).[7] The derivation of this set of candidate equilibria relies on probing active trades with three classes of unilateral seller deviations: undercutting (as in Bertrand competition), pivoting (cf. Attar et al. (2011, 2014)), and efficiency-improvements (buyers trading the most desirable quantity at a given marginal price). The deviations pinpoint a set of "Pooling + Separating" aggregate trades where, in addition to pooling, low-risk buyer types almost always trade a separating contract. For fixed unit prices, equilibrium candidates are not unique but described by a continuum: if pooling coverage is greater under candidate 1 than 2, individuals typically purchase less separating coverage under candidate 1 than 2 (see Proposition 2).

From a welfare perspective, *never exclusive* equilibrium candidates qualitatively resemble the second-best efficient allocation: both characterize a set of "Pooling + Separating" allocations. However, an important distinction remains. Unlike under second-best efficiency, an increase in low-risk coverage, even though advantageous to low-risk buyers, will not entice high-risk buyers to opt for the more attractively priced lesser-coverage low-risk contract. Put differently, incentive constraints remain slack. However undesirable from a welfare perspective, this equilibrium condition must persist; otherwise, inactive sellers could attract low-risk buyers by proposing an alternative pooling contract with less coverage.

Our second main result (see Theorem 2) shows that any allocation that satisfies the necessary conditions identified by Theorem 1 can be decentralized as an equilibrium under the "1+1" market structure. A low type separating contract prevails in equilibrium. Our result establishes, moreover, that there is no benefit in adopting any *never exclusive* market structure different

[7]The set of *never exclusive* market structure encompasses the "1+1" market structure by imposing the minimal constraint that no seller contract is exclusive (which cancels the threat of pivoting).

from the "1+1" market structure. For the result to hold we require, as in Attar et al. (2022), the additional *flatter curvature assumption* that imposes restrictions on the curvature of indifference curves across types. As in their work, equilibrium existence hinges on so-called latent contracts, i.e., contracts that are not traded actively in equilibrium but that play a role to deter cream-skimming deviations uniquely targeting low-risk buyers. Analogously, we identify a principal latent contract that can be derived via a thought experiment. Suppose that some inactive sellers were to offer a cream-skimming contract uniquely targeting low-risk buyers. Since low-risk buyers cross-subsidize high-risk buyers when purchasing the pooling contract, cream-skimming can exploit potential gains from trade (to the detriment of sellers offering pooling contracts). The flatter curvature assumption posits that there always exists an additional, latent contract so that the latent contract and the cream-skimming contract taken together achieve greater high type utility than the high type's initial allocation. This ensures that large cream-skimming deviations are never profitable. Reassuringly, Attar et al. (2022) show that quadratic utility, as derived from constant-absolute-risk-aversion (CARA) preferences with normally distributed health shocks (as in Einav et al. (2013); Azevedo and Gottlieb (2017)), satisfies the flatter curvature assumption provided that low-risk buyers are weakly more risk-averse than high-risk buyers.

We wish to stress a normative implication of both results: the choice of market structure can be seen as a regulatory optimization problem given some utilitarian welfare objective. Our existence result suggests that the regulator's problem reduces to the binary choice between the fully exclusive "1 or 1" and the never exclusive "1+1" market structure. A comparison between equilibria under both market structures is therefore required. We show that "Pooling + Separating" allocations that occur under the "1+1" market structure Pareto dominate the Rothschild-Stiglitz allocation when adverse selection and therefore rationing of low-risk buyers is severe. By contrast, the Rothschild-Stiglitz allocation never Pareto dominates because high-risk buyers benefit from the cross-subsidies implied by pooling contracts. A positive reading, by contrast, suggests that the emergence of the "1+1" structure can also be understood as an informal industry agreement. Specifically, we show (see Proposition 3) that for any "Pooling + Separating" allocation, and once trade has taken place, no seller can deviate from the pre-assigned identities and propose profitable additional trades to the buyers. We refer to this stability concept as serendipitous-aftermarket-proofness in that the aftermarket occurs unexpectedly. The Rothschild-Stiglitz allocation, by contrast, need not be serendipitous-aftermarket-proof, even if it is an equilibrium.

Our third main result furthers this normative insight. We prove that when coupling the "1+1" market structure with the requirement that all purchases include a group 1 contract exceeding a minimum quantity, the ensuing equilibrium allocation need not only Pareto-improve upon the exclusive allocation but is always second-best efficient.[8] Minimum quantity require-

[8]We emphasize the distinction between minimum quantity regulations and an insurance mandate, particularly in light of the legal ruling in Texas v. United States, No. 4:18-cv-00167-O (N.D. Tex. Dec. 14, 2018), which invalidated the insurance mandate under the Affordable Care Act. Unlike a mandate, minimum quantity

ments can be beneficial even if the required coverage equals the amount of coverage offered in group 1 prior to the quantity regulation. The reason is that a minimum quantity requirement cancels all cream-skimming deviations in group 1. Absent the threat of cream-skimming, sellers in group 2 can provide greater low-risk coverage without concurrently creating incentives for inactive sellers to offer less pooling coverage in group 1. The main constraint for minimum quantity regulations to be beneficial is that they are robust to pivoting deviations, i.e., they must be sufficiently large. Largeness ensures that high-risk buyers prefer the group 1 pooling contract and the group 2 high-risk separating contract over any profitable high-risk pivoting contract in group 2 combined with the group 1 low-risk separating contract. A small group 1 quantity requirement is inconsequential by contrast.

We would like to note that our result exhibits features akin to a second welfare theorem. Any second-best efficient allocation that satisfies large pooling can be decentralized as an equilibrium if regulators impose the "1+1" market structure and a minimum quantity regulation. Changes in the minimum quantity move the equilibrium allocation along the Pareto frontier. A greater minimum quantity benefits high-risk buyers to the detriment of low-risk buyers. In addition to enhancing consumer welfare, introducing a minimum quantity regulation further stabilizes the equilibrium: The regulation dispenses with the need of sustaining an allocation with latent contracts. Where previously latent contracts deterred cream-skimming deviations, the minimum quantity regulation now outright bans cream-skimming deviations, thus obviating the need to issue latent contracts to begin with.

Related Literature

Our paper relates to the literature on adverse selection in competitive markets. The dominant framework is exclusive competition, introduced by Rothschild and Stiglitz (1976) with an important contribution by Azevedo and Gottlieb (2017). A recent, burgeoning literature considers nonexclusive competition instead. An early paper is Pauly (1974) who restricts attention to linear contracts.[9] The JHG allocations was derived by Jaynes (1978), Hellwig (1988), Glosten (1994). Our paper most closely relates to Attar et al. (2022), who share our focus on single contracts as opposed to menus, and more broadly the research agenda on nonexclusive competition in markets afflicted by adverse selection (see Attar et al. (2011)).[10] Attar et al. (2014) singled out the JHG allocation as the unique equilibrium candidate in competitive insurance markets, later shown to be the unique equilibrium allocation under a sequential auction mechanism (see Attar et al. (2021)) and single contracting (see Attar et al. (2022)). In further work, Stiglitz et al. (2020) consider disclosure rules that sustain the JHG allocation when sellers

requirements pertain to contract characteristics on the supply side, akin to current regulations still in effect in U.S. health insurance marketplaces (e.g., bronze, silver, gold, or platinum plans). In contrast to a mandate, buyers are only affected if they choose to purchase insurance.

[9]This restriction arises naturally in a Walrasian equilibrium where the agents' trades cannot be monitored as in Bisin and Gottardi (1999, 2003). Much of the literature instead focuses on competition in contracts where sellers compete in price and the extent of coverage.

[10]See also the earlier and related common agency literature that employed mechanism design tools, cf. Biais et al. (2000).

post menus and Asriyan and Vanasco (2023); Donaldson et al. (2020) prove how nonexclusive securitization of asset cash flows gives rise to the JHG allocation.[11,12]

A key assumption of our baseline model is that each seller offers only a single contract. This rules out deviations involving a loss-making contract to high-risk buyers in group 2 and a cream-skimming deviation to low-risk buyers in group 1. These deviations are at the heart of the nonexistence of a pure-strategy equilibrium result in Attar et al. (2014). Introducing regulation that penalizes issuers of loss-making contracts would restore the competitive environment analyzed here—the set of equilibrium allocations would be unchanged—while allowing sellers to offer a menu of contracts. Penalties on loss-making contracts are advocated for by Attar et al. (2022) and loosely resemble cost-sharing mechanisms that pool and redistribute costs among sellers of a standardized basic-coverage contract.[13] The complementary analysis of the unregulated menu pricing game is found in Huang (2022).[14]

It is a priori not clear whether multiple contracting is pro-competitive or anti-competitive. On the one hand, and as Attar et al. (2011) argue, nonexclusive competition allows for more possible deviations for sellers, in particular pivoting on other sellers' contracts. On the other hand, sellers have more tools at their disposal to block deviations through the issuance of latent contracts. [15]

There is an extensive empirical literature that tests whether adverse selection occurs (see Chiappori and Salanie (2000); Chiappori et al. (2006)).[16] As pointed out by Attar et al. (2022), multiple contracting or nonexclusive competition likely reverses the predictions on the correlation between risk and coverage on a *per-contract* basis. Mimra and Waibel (2021) conduct

[11]In Asriyan and Vanasco (2023), securities must be fully backed by the asset cash flows, thus allowing the security originator to effectively commit to exclusive trading when pledging the entire cash flow at a given level of return. In Donaldson et al. (2020), the security originator can instead restore exclusivity of trades in a priori nonexclusive markets by pledging collateral. In their model, collateral claims, unlike cashflows, cannot be diluted.

[12]Auster et al. (2022), building on Guerrieri et al. (2010), consider an otherwise standard directed search model where agents can apply for several contracts. In relation to the literature on nonexclusive competition, they show that agents' indirect utilities conditional on already made applications can feature a reversal of single-crossing.

[13]Countries using such schemes include Australia, Germany, Ireland, the Netherlands, Slovenia and Switzerland. Glazer and McGuire (2006) offer a more detailed description of risk-adjustments and offer a discussion of the "sickness fund" system in Germany.

[14]An equilibrium only exists under a perfect translation property of indifference curves (e.g., quadratic utility with identical risk preferences) which is stronger than requiring the flatter curvature assumption.

[15]Latent contracts were first introduced in the context of competing mechanisms in the seminal paper by Peters (2001). Since, latent contracts have mainly appeared under multiple contracting in the literature on moral hazard. Hellwig (1983) argues that latent contracts can deter entry into the insurance market when agents' effort decisions are not contractible. This can result in positive equilibrium profits. In Attar et al. (2019) latent contracts collectively sustain the monopoly profit for the sellers. Ours is the first paper on multiple contracting in the context of adverse selection where latent contracts allow some sellers to sustain positive profits.

[16]Finkelstein and Poterba (2004) study the UK annuities market and find a positive correlation between buyers' ex-post risk (i.e., longevity) and the coverage level purchased. Bauer et al. (2020) find evidence for adverse selection in the secondary market for settlements of life insurance policies. Other studies point towards the absence of adverse selection. Cawley and Philipson (1999) finds no evidence that adverse selection exists in the primary life insurance market. Chiappori and Salanie (2000) use data on the French car insurance market on contracts and accidents and find no evidence for adverse selection, even when considering senior drivers only.

an experiment to assess whether the predictions of the theory on contracting under adverse selection hold up in a laboratory environment. They show that in the context of menu-pricing and both exclusive and nonexclusive competition the theory's predictions match the observed behavior. Their study design invites further investigation into the "1+1" market structure; it would equally be helpful to see whether double-deviations that are feasible under menu-pricing play as destructive a role in a laboratory environment as the theory suggests.

Our paper suggests that markets can perform better under segmentation—indeed, the "1+1" market structure proposes such segmentation for policies that could a priori be sold as a single policy. Up until now, similar positive results on market segmentation were only reported in the context of private values: Malamud and Rostek (2017) show that equilibrium utilities in a decentralized market can be strictly higher in the Pareto sense than in a centralized market with the same traders and assets. Chen and Duffie (2021) argue that fragmentation induces agents to trade more aggressively; any degree of fragmentation is welfare-superior to a centralized financial market. And Rostek and Yoon (2021) find that multiple trading protocols that clear independently can be designed to be at least as efficient as joint market clearing for all assets.

Finally, our welfare result complements a literature that studies market interventions under exclusive competition. Minimum coverage requirements are sometimes seen to be beneficial[17] while also aggravating adverse selection for levels of coverage exceeding said requirement (Azevedo and Gottlieb (2017)).[18] Welfare implications of market interventions in even partially nonexclusive insurance economies, by contrast, remain unexplored to the best of our knowledge.

This paper is organized as follows. Section 2 introduces the model and proposes the concept of a market structure. Section 3 partitions the set of competitive market structures into partially exclusive and never exclusive market structures that subsume exclusive and nonexclusive competition as two (polar) cases. Sections 4 and 5 present our main *positive* results: necessary conditions that any equilibrium candidate under a never exclusive competitive market structure must satisfy, and an equilibrium existence result. In particular, we prove the existence of "Pooling + Separating" equilibrium allocations that have never been studied before. Finally, Section 6 explores normative and positive implications. Here we prove our main *normative* result: imposing that competition—as governed by the "1+1" market structure —be partially nonexclusive and jointly introducing a minimum requirement on group 1 coverage results in an equilibrium that is second-best efficient. If adverse selection is severe, the equilibrium moreover Pareto-improves upon the separating exclusive equilibrium.

[17]See also Neudeck and Podczeck (1996), Encinosa (2001), McFadden et al. (2015) who build on the Grossman (1979) equilibrium concept.

[18]Veiga and Levy (2023) expand on the welfare implications of regulating contract characteristics in exclusive markets.

2 Set-up

We here introduce a model of strategic price-setting in a competitive market plagued by adverse selection. The common interpretation given to this model is that of an insurance economy in which buyers with an exogenously given high- and low-risk profile purchase coverage in exchange for a premium.

Our description of the economy (preferences, cost and the space of admissible contracts) is identical to that in Attar et al. (2021), Attar et al. (2022) and encompasses the classical set-up in Rothschild and Stiglitz (1976). We innovate in that we introduce the concept of a *market structure*, i.e., a trading rule that specifies the subsets of sellers whom the buyers can jointly with. The definition admits as special cases the market structures where each buyer can trade with at most one (exclusive competition) and with arbitrarily many (nonexclusive competition) sellers.

2.1 The Contracting Environment

We consider a finite set of sellers $\mathcal{K} = \{1, ..., K\}$ and a continuum of buyers that are characterized by their type $\theta \in \{L, H\}$. Sellers compete by proposing contracts $(q^k, t^k) \in \mathbb{R}^2_+$ specifying a quantity and a transfer. Types are non-contractible so that all buyers can select identical trades if they so wish.

Preferences. Buyers view sellers as perfect substitutes, so that their preferences can be represented over aggregate trades: Denote $M \subset \{1, ..., K\}$ a subset of sellers that buyer type θ trades with and $Q = \sum\limits_{k \in M} q^k$ and $T = \sum\limits_{k \in M} t^k$ the corresponding aggregate quantity and transfer. Preferences over aggregate trades are represented by a utility function $U_\theta(Q, T)$ that is increasing in Q and decreasing in T and that satisfies the following assumptions:

First, we impose regularity conditions so that the buyer's demand is well-behaved.[19]

Assumption 1 (quasi-concavity)**.** *$U_\theta(Q, T)$ is strictly quasi-concave, i.e. $\forall \alpha \in (0, 1)$, and $(Q_1, T_1) \neq (Q_2, T_2)$ it holds that $U_\theta(\alpha(Q_1, T_1) + (1 - \alpha)(Q_2, T_2)) > \min\{U_\theta(Q_1, T_1), U_\theta(Q_2, T_2)\}$.*

Assumption 2 (finite demand)**.** *$\arg\max\limits_{Q \geq 0} U_\theta(Q, Qc_x)$ is finite $\forall\, c_x > 0$ and $\theta \in \{L, H\}$.*

Notice that strict quasi-concavity implies that $\arg\max\limits_{Q \geq 0} U_\theta(Q, Qc_x)$ is a singleton.

Second, we assume that types are ordered so that high buyer types' demand exceeds low buyer types' demand:[20]

Assumption 3 (single-crossing)**.** *For all (Q, T) and (Q', T') so that $Q' > Q$ it holds that $U_L(Q', T') \geq U_L(Q, T) \quad \Rightarrow \quad U_H(Q', T') > U_H(Q, T)$.*

[19] In Section 5 we introduce further regularity conditions that ensure the existence of an equilibrium.

[20] Provided that utility is differentiable, this is equivalent to assuming that the slope of the indifference curve $\tau_H(Q, T) = -\frac{\partial_1 U_\theta(Q,T)}{\partial_2 U_\theta(Q,T)}$ is greater for higher types, i.e., $\tau_H(Q, T) > \tau_L(Q, T)$.

Cost. Trading a contract (q, t) with a buyer type θ earns the seller an expected profit $t - c_\theta q$. Here c_θ denotes the marginal cost of serving type θ. In line with a model of adverse selection, we assume that those buyer types most eager to trade, i.e., high types H, are also the most costly to serve.

Assumption 4 (Adverse Selection)**.** $c_H > c_L$.

Finally, denote m_H the proportion of type H and m_L the proportion of type L buyers so that the average marginal cost is $c = c_H m_H + c_L m_L$.

2.2 Market Structure

The key innovation of our framework is the concept of a market structure. A market structure specifies which sellers a buyer can jointly trade with.

Definition 1. *A market structure $\mathcal{M}$ is a (non-empty) collection of subsets of sellers with whom a buyer can jointly trade: $\mathcal{M} \subseteq \mathcal{P}(\{1, ..., K\}) \equiv \mathcal{P}(\{\textit{all sellers}\})$.*

The two polar cases considered in the literature are defined as follows:

Example 1. (i) *Exclusive competition* $\mathcal{M} = \big\{\emptyset, \{1\}, \{2\}, ..., \{K\}\big\}$.

(ii) *Nonexclusive competition :* $\mathcal{M} = \mathcal{P}(\{1, ..., K\})$

Here $\mathcal{P}(\{1, ..., K\})$ denotes the power set, i.e., the set of all subsets of $\{1, ..., K\}$.

In line with the two polar cases of exclusive and nonexclusive market structures, our focus is on competitive market structures. Therefore, we require that buyers can decline sellers' offers and that each seller is replaceable.

Definition 2. *A market structure $\mathcal{M}$ is competitive if*

- *buyers can trade with any subset of a feasible set of trading partners, i.e., for all $M \in \mathcal{M}$ and $j \in \mathcal{K}$, if $j \in M$, then also $M \setminus \{j\} \in \mathcal{M}$;*
- *each seller is twice replaceable, i.e. for all $M \in \mathcal{M}$ and $j \in \mathcal{K}$, if $j \in M$, then there exist distinct $k_1, k_2 \in \mathcal{K} \setminus M$ so that $M \cup \{k_1\} \setminus \{j\} \in \mathcal{M}$ and $M \cup \{k_2\} \setminus \{j\} \in \mathcal{M}$.*

The ability to decline offers allows buyers to play some sellers against others by threatening to accept only a subset of the offers they receive. Bernheim and Whinston (1986) call this arrangement delegated common agency. Less intuitively, our definition insists that sellers be twice replaceable. This requirement ensures that an active seller is always competing with an inactive seller, preserving undercutting incentives.[21]

[21] If seller h had only one replacement seller, ℓ, it is conceivable that seller ℓ also has only one replacement seller, h. This scenario would allow sellers h and ℓ to collude, with one seller profitably serving high types and the other profitably serving low types.

2.3 Equilibrium

One can imagine a benevolent regulator that decides on the jointly feasible trades before the market opens. For now we shall take the (competitive) market structure as given and focus our analysis on the ensuing equilibria. Welfare considerations that can inform the selection of different market structures will be discussed in Section 7.

The Simultaneous Move Game. We consider a competitive screening game in which firms compete by each posting a single contract.[22] Given a fixed market structure $\mathcal{M} \in \mathcal{P}(\{1, ..., K\})$, the game unfolds as follows:

- Stage 1: Each seller k proposes a contract $(q^k, t^k) \in \mathbb{R}^2_+$.
- Stage 2: Each buyer learns her type, selects some $M \in \mathcal{M}$ and derives utility $U_\theta\big(\sum_{k\in M} q^k, \sum_{k\in M} t^k\big)$.

Our equilibrium concept is standard:

Definition 3 (Equilibrium)**.** *A pure strategy perfect Bayesian equilibrium (PBE) specifies, for a fixed market structure $\mathcal{M}$, buyer strategies $S = (S_\theta)_{\theta\in\{L,H\}}$ and seller contracts $\mathcal{C} = (q^k, t^k)_{k\in\mathcal{K}}$. Optimality of individual choices and consistency of the sellers' beliefs entail that:*

- *Each buyer type's strategy S_θ selects a set of sellers $M \in \mathcal{M}$ that collectively offers the most advantageous contracts:*

$$S_\theta(\mathcal{C}') \in \arg\max_{M\in\mathcal{M}} U_\theta\Big(\sum_{k\in M} q^k, \sum_{k\in M} t^k\Big) \quad \forall \mathcal{C}' \in \mathbb{R}^{2K}_+.$$

- *Individual seller k offers a contract (q^k, t^k) that is profit-maximizing given the contracts offered by their competitors and anticipating the buyers' choices:*

$$(q^k, t^k) \in \arg\max_{(q'^k, t'^k)\in\mathbb{R}^2_+} \sum_{\theta\in\{L,H\}} \big[t'^k - c_\theta\, q'^k\big] m_\theta\, 1\Big\{k \in S_\theta\big(\{(q^\ell, t^\ell), (q'^k, t'^k)\}_{\ell\neq k}\big)\Big\}.$$

An immediate consequence of our focus on PBE is that along the equilibrium path we can distinguish between active sellers and inactive sellers. Active sellers propose contracts that are actively traded, inactive sellers may propose so called latent contracts. As we shall see, latent contracts play an important role to sustain an equilibrium (if it exists).

[22] Our focus on single contracts avoids the issue of some sellers offering loss-making contracts on the equilibrium path. Alternatively, we can consider a regulated insurance market (as discussed in Attar et al. (2022)) in which sellers face heavy fines if some contracts on their menu incur a loss. The restriction that contracts must not be loss-making is crucial in our environment. In a companion paper, Huang (2022) shows that when firms can post menus instead of single contracts, an equilibrium often fails to exist for certain parameter values, even when it exists in our set-up. The unique equilibrium allocation, if it exists, is the JHG allocation studied in the next section.

3 Partially and Never Exclusive Market Structures

As a precursor to our main results, we here introduce a partition of the set of competitive market structures into two disjoint subsets. We show that if the market structure permits exclusive trades, then the Rothschild-Stiglitz separating allocation is the unique equilibrium candidate allocation. Conversely, we show that if the market structure prohibits exclusive trades, then any admissible equilibrium candidate allocation is at least partially pooling. This serves as a precursor to our subsequent analysis which will be concerned with characterizing the set of non-separating equilibria when the market structure is never exclusive.

3.1 Partially Exclusive Market Structures

The most prominently discussed equilibrium candidate in the literature is the Rothschild-Stiglitz separating allocation. It is defined as follows:

Definition 4 (Rothschild-Stiglitz (RS))**.** *The RS allocation is the separating allocation* (Q_L^{RS}, T_L^{RS}) *and* (Q_H^{RS}, T_H^{RS}) *where, subject to* $U_H(Q_H^{RS}, T_H^{RS}) \geq U_H(Q_L^{RS}, T_L^{RS})$*,*

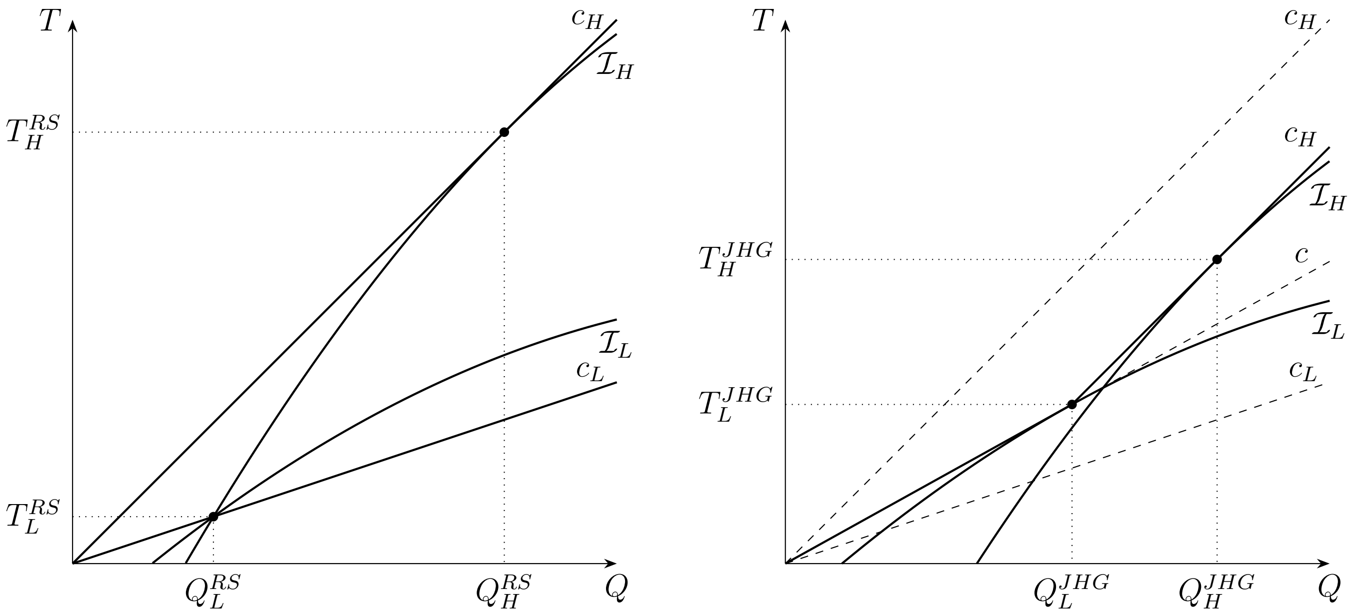
$$Q_H^{RS} = \arg\max_{Q_H \geq 0} U_H(Q_H, c_H Q_H), \quad T_H^{RS} = c_H Q_H^{RS}$$
$$Q_L^{RS} = \arg\max_{Q_L \geq 0} U_L(Q_L, c_L Q_L), \quad T_L^{RS} = c_L Q_H^{RS}.$$

This allocation usually entails rationing (as depicted in the left panel of Figure 1): low type buyers would like to purchase more quantity at the low unit price c_L offered. The sellers, by contrast, refuse to provide more coverage because they anticipate that doing so would also attract high type buyers who are more costly to serve. Rothschild and Stiglitz (1976) show that the RS separating allocation is the unique equilibrium candidate when the market structure is exclusive.

We extend this result and introduce the largest class of market structures for which the RS allocation is the unique equilibrium candidate. This class, rather than imposing that all sellers can offer exclusive trades, only requires that some seller can offer exclusive contracts. (Since each seller is twice replaceable, the presence of one exclusive seller implies that there are at least three sellers that offer exclusive contracts.) If a market structure satisfies this property, we say that the market structure is partially exclusive.

Definition 5. *A market structure* $\mathcal{M}$ *is partially exclusive if there exists a seller that trades exclusively, i.e.,* $\max_{M \in \mathcal{M}: k \in M} |M| = 1$ *for some* $k \in \mathcal{K}$*.*

The right to offer an exclusive contract gives sellers the ability to destabilize any partially pooling equilibrium via cream-skimming deviations, i.e., deviations that uniquely target low type (and low cost) buyers. It follows that any equilibrium allocation must be fully-separating. Efficiency arguments then imply that the separating equilibrium candidate is uniquely defined.



Exclusive Competition: RS

Nonexclusive Competition: JHG

Figure 1: Rothschild-Stiglitz (RS) and Jaynes-Hellwig-Glosten (JHG) allocation

Claim 1. *Posit Assumptions 3 and 4 . The RS separating allocation is the unique equilibrium candidate allocation under a partially exclusive and competitive market structure.*

3.2 Never Exclusive Market Structures

We now consider the complement to partially exclusive market structures: no seller can offer an exclusive contract that prohibits further trade with other sellers. If so, we say that the market structure is never exclusive.

Definition 6. *A market structure $\mathcal{M}$ is never-exclusive if no seller exclusively trades with buyers, i.e.,* $\max_{M\in\mathcal{M}:k\in M} |M| \neq 1$ *for all* $k \in \mathcal{K}$.

Nonexclusive competition is an example of a never exclusive market structure. Another example of a never exclusive market structure —indeed the key example —is the following:

Example 2 ("1+1" market structure)**.** *$\mathcal{M}$ is a "1+1" market structure if the sellers $\mathcal{K} = \{1, ..., K\}$ can be partitioned into two disjoint subgroups $\mathcal{K}_1$ and $\mathcal{K}_2$ so that buyers can never trade with two sellers from the same subgroup at the same time:*

$$\mathcal{M} = \big\{\{j, k\} : j \in \mathcal{K}_1 \cup \{\emptyset\}, \, k \in \mathcal{K}_2 \cup \{\emptyset\}\big\}.$$

In contrast to partially exclusive market structures, never exclusive market structures never admit fully-separating equilibria. This means that if a (non-trivial) equilibrium exists, some sellers must actively trade a pooling contract with both buyer types.

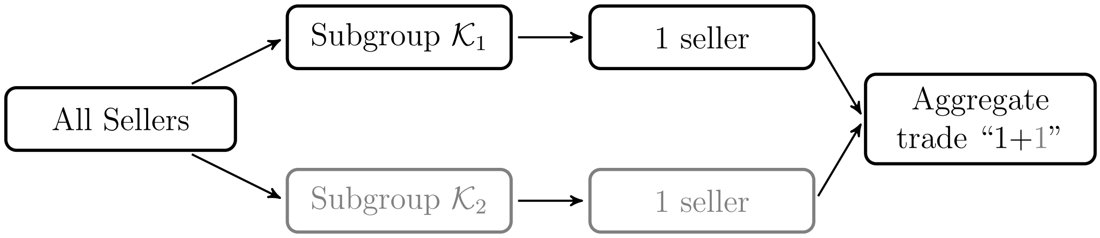


Figure 2: The "1+1" market structure. Buyers select at most one seller from each subgroup.

Claim 2. *Posit Assumptions 3 and 4. No allocation that is fully-separating can occur as an equilibrium allocation under a never exclusive and competitive market structure.*

This claim is due to a pivoting argument: Suppose that some seller exclusively trades with low type buyers. Since the separating contract has low unit cost, there always exists a profitable complementary contract so that both the separating and the complementary contract taken together match the high type's quantity allocation at a lower total price. In particular, following the introduction of the complementary contract, the initially separating contract is pooling and loss-making.

If not fully-separating, what will be the equilibrium? One candidate is given by the Jaynes-Hellwig-Glosten allocation. This allocation consists of two competitively priced layers: a basic pooling layer and an additional layer purchased only by high type buyers. In a series of recent contributions, Attar et al. (2022) show that the JHG allocation is the unique equilibrium candidate when the market structure is nonexclusive. As the next section shows, the JHG allocation remains a viable equilibrium candidate when considering the larger class of never exclusive market structures. The more surprising insight is that many other equilibrium candidates become viable, too.

Definition 7 (Jaynes-Hellwig-Glosten (JHG))**.** *The JHG allocation is the partially pooling allocation* (Q_L^{JHG}, T_L^{JHG}) *and* (Q_H^{JHG}, T_H^{JHG}) *where*

$$Q_L^{JHG} = \arg\max_{Q \geq 0} U_L(Q, cQ) \quad \text{and} \quad Q_H^{JHG} - Q_L^{JHG} = \arg\max_{Q \geq 0} U_H(Q_L^{JHG} + Q, T_L^{JHG} + c_H Q),$$

purchased at actuarially fair transfers $T_L^{JHG} = cQ_L^{JHG}$ *and* $T_H^{JHG} - T_H^{JHG} = c_H(Q_H^{JHG} - Q_L^{JHG})$.

4 Non-Fully-Separating Equilibrium Candidates

The preceding section shows: if an equilibrium exists, the ensuing allocation will be non-fully-separating if and only if the market structure is never exclusive. But what are the possible partially pooling allocations that can occur in equilibrium? A characterization extending beyond the example of the JHG allocation is still missing. We here consider arbitrary competitive (and in light of Claim 1 necessarily never exclusive) market structures. Our objective in this section is to present four necessary conditions that any non-fully-separating equilibrium allocation must satisfy.

An allocation consists of the sum of trades of the low type, (Q_L, T_L), and the sum of trades of the high type, (Q_H, T_H). Clearly, equilibrium allocations must be incentive compatible, that is the high (low) type prefers her allocation over the low (high) type's allocation.

Condition 1 (Incentive Compatibility)**.** *An allocation (Q_L, T_L) and (Q_H, T_H) is incentive compatible if $U_L(Q_L, T_L) \geq U_L(Q_H, T_H)$ and $U_H(Q_H, T_H) \geq U_H(Q_L, T_L)$.*

In general, thinking in terms of allocations is imprecise. Since high buyer types are more costly to serve than low types, we must distinguish between contracts purchased by low, high and both buyer types. Refer to these as the aggregate active trades $(q_L, t_L), (q_H, t_H)$ and (Q_P, T_P). Then $(Q_L, T_L) = (Q_P + q_L, T_P + t_L)$ and $(Q_H, T_H) = (Q_P + q_H, T_P + t_H)$.

4.1 Necessary Conditions

As is commonly the case in competitive equilibrium we identify the set of equilibrium candidates via possible seller and buyer one-shot deviations.

4.1.1 Undercutting and Pivoting

Much insight can be won by probing a candidate equilibrium with two kinds of seller deviations only: undercutting, i.e., sell the same quantity of an existing active contract for less, and pivoting, ask the buyer to combine an existing contract with a deviating contract without changing the aggregate quantity traded. Undercutting deviations are well-understood in the context of Bertrand competition. Pivoting deviations have been extensively explored in the context of non-exclusive competition but may in general be less well-known. What is important to observe is that neither deviation requires actual knowledge of the buyers' preferences. This makes necessary conditions derived from this class of deviations particularly robust.

Condition 2. *Active trades are single-seller separating and competitively priced if*

(i) *each component is competitively priced, i.e.,*

$$t^\ell \in q^\ell [c_L, c], \qquad t^h = q^h c_H \qquad \textit{and} \qquad t^p = q^p c$$

for all $\ell \in M_L \backslash M_H$ and $h \in M_H \backslash M_L$ and $p \in M_L \cap M_H$;

(ii) *the low type separating component, if non-zero, is actively traded by a single seller only, i.e., $|M_L \backslash M_H| = 1$.*

Proposition 1. *Posit Assumptions 1, 3 and 4. Active trades that occur in a non-fully-separating equilibrium under a competitive market structure are competitively priced and single-seller separating.*

The idea that all contracts must be competitively priced is reminiscent of Bertrand competition. What is concerning is that it does not apply to the low type contract. Serving the low type is potentially profitable. The reason is that by undercutting a contract exclusively targeting low type buyers one may also attract high type buyers, thereby greatly increasing the cost of said contract.

The key insight here is that sellers can be endogenously separated into two groups, $\mathcal{K}_1$ and $\mathcal{K}_2$, so that along the equilibrium path sellers in group one sell pooling contracts and sellers in group two sell separating contracts. Denote $(Q_P, T_P) = \sum_{k \in M_L \cap M_H} (q^k, t^k)$ the sum of active pooling contracts and $(q_L, t_L), (q_H, t_H)$ the (sum of) active separating contracts. The parentheses are warranted, for Proposition 1 asserts that (q_L, t_L) is in fact a single contract. Single-seller separation must occur because of possible pivoting deviations. To see why, suppose there were (at least) two sellers that exclusively trade with low type buyers, e.g., $q_L = q_L^1 + q_L^2$. Then possible undercutting deviations ensure that none of them makes a profit. Yet one of the two, say seller one, could pivot on seller two's contract traded in conjunction with the aggregate pooling trade (Q_P, T_P) and propose the quantity $Q_H - q_L^1 - Q_P$ at a unit price slightly exceeding c_H (and thereby be profitable). Due to (q_L^2, t_L^2) being priced competitively (at unit cost c_L), high type buyers must be strictly better off following this deviation.

Proposition 1 opens up the possibility of separation of low type buyers. Whether this can happen is a property of the prevailing market structure. (It does for some). A consequence of another pivoting deviation is however that the separating contract must be sufficiently small vis-à-vis the aggregate pooling quantity.

Condition 3 (large pooling)**.** *Active aggregate trades are largely pooling if*

$$T_H - T_L + T_P \leq c_H(Q_H - Q_L + Q_P).$$

Condition 3 prevents pivoting deviations that target the high type: complement type L's separating contract (q_L, t_L) with a pivoting contract promising quantity $(Q_H - q_L)$. Trading the pivoting contract is potentially attractive, because high type buyers benefit from the lower unit cost of the separating contract (q_L, t_L). Previously, they elected not to trade this contract because it forced them to forfeit some (if not all) of the contracts that made up the aggregate separating trade (q_H, t_H). Yet by complementing the low type's separating contract, high type buyers can now purchase the same aggregate quantity as before. Condition 3 ensures that rendering this deviation incentive compatible is too costly to the seller. Indeed, the pivoting contract comprises quantity $Q_H - q_L = Q_H - Q_L + Q_P$ which includes the pooling segment Q_P. The cost of this segment is higher than before, because the pivoting unlike the pooling contract exclusively targets high type buyers.

Lemma 1. *Posit Assumptions 1, 2, 3 and 4. Any aggregate active trades that occur in a non-fully-separating equilibrium under a competitive market structure are largely pooling.*

Due to competitive pricing asserted by Proposition 1, the lower bound on the pooling quantity can be expressed more explicitly. Competitive pricing, e.g., Condition 2, implies that $T_P = cQ_P$ and $T_H - T_P = c_H(Q_H - Q_P)$. Then Condition 3 re-writes as follows:

$$c_H q_L - t_L \leq (c_H - c)Q_P.$$

4.1.2 Efficiency-improving Deviations

Undercutting and pivoting deviations only require the seller to observe which trades are active. They do not require the seller to know the buyers' preferences. We now consider properties of any equilibrium that must hold due to possible efficiency-improving deviations.

Condition 4 (conditional efficiency). *Active aggregate trades are conditionally efficient if* $Q_P \in \arg\max_{Q\geq 0} U_L(Q + q_L, Qc + t_H)$, *and* $q_H \in \arg\max_{q\geq 0} U_H(Q_P + q, T_P + qc_H)$ *whenever* $q_H > 0$. *If instead* $q_H = q_L = 0$, *it must hold that* $\max_{q\geq 0} U_H(1/2Q_P + q, 1/2Q_P c + qc_H) \leq U_H(Q_P, Q_P c)$.

Of course, assuming that indifference curves are continuously differentiable, this is equivalent to requiring that the slope of the indifference curve satisfies $\tau_L(Q_P + q_L, T_P + t_L) = c$ whenever $Q_P > 0$ and $\tau_H(Q_P + q_H, T_P + t_H) = c_H$ whenever $q_H > 0$.

Condition 4 is motivated as follows. Fix as a unit price the lower tail expectation, i.e., the expected unit cost conditional on all buyer types greater than oneself purchasing the same contract. If a buyer wanted to trade at this price to deviate from the candidate allocation, the desire to trade would not go away if the unit price were slightly less favorable to him. Since this price is profitable to the sellers, we should expect that any such gain from trade will be exploited in equilibrium.[23]

Lemma 2. *Posit Assumptions 1, 2, 3 and 4. Any aggregate active trades that occur in a non-fully-separating equilibrium under a competitive market structure are conditionally efficient.*

Review of the Necessary Conditions

We now take stock and summarize in a single theorem the findings of Proposition 1, Lemma 1 and Lemma 2.

Theorem 1. *Posit Assumptions 1, 2, 3 and 4. Active trades that occur in a non-fully-separating equilibrium are single-seller separating. Moreover, aggregate active trades satisfy Conditions 1, 2, 3 and 4, i.e., are incentive compatible, competitively priced, largely pooling and conditionally efficient.*

[23] The case where $Q_H = Q_P$, yet $\max_{q\geq 0} U_H(Q_P + q, Q_P c + qc_H) > U_H(Q_P, Q_P c)$ is pathological. It allows for situations in which the high type would like to purchase additional coverage at unit price c_H but cannot, because he has exhausted his purchasing options in both groups by choosing pooling contracts. Efficiency would suggest that only one group should offer (the entire) pooling contract $(Q_P, Q_P c)$, to liberate the high buyer type's option of purchasing a separating contract in the other group. This could only arise due to a coordination failure among sellers. We view such coordination failure as implausible: profitable deviations exist if two sellers from both groups could simultaneously deviate.

4.2 Existence of an Equilibrium Candidate

How can we ensure that aggregate active trades satisfying Conditions 1, 2, 3 and 4 do exist? First, observe that the set of equilibrium candidates is non-empty: the JHG allocation is an admissible candidate. Do there exist other? And how large a space do they comprise?

If we fix the low type contract's unit price $c_x \in [c_L, c)$, the equilibrium candidate is uniquely determined by the pooling quantity Q_P. T_P follows from competitive pricing, and $(q_H, q_H c_H)$ and $(q_L, q_L c_x)$ are uniquely determined by conditional efficiency. Visually, one may walk along the separating c_H and c_x-unit cost lines emanating from the pooling allocation until one finds the separating trades that satisfy conditional efficiency.

Uniqueness of q_H and q_L, however, is conditional on the competitively priced pooling allocation $(Q_P, Q_P c)$. And as we now shall see, many pooling quantities Q_P are conceivable. For every low type unit price $c_x \in [c_L, c)$ there exists an open set of equilibrium candidates. This stands in contrast to the unique equilibrium candidates that have been identified for exclusive and nonexclusive market structures. To show this result, we must slightly strengthen our assumptions: we require that utility is strictly quasi-concave and twice continuously differentiable.

Assumption 5. *The utility function of buyers $U_\theta(Q, T)$ is twice continuously and differentiable.*

Proposition 2 (Existence of a continuum of equilibrium candidates)**.** *Posit Assumptions 1, 2, 3, 4 and 5. Then for all $c_x \in [c_L, c]$ there exist $\underline{Q}_x, \overline{Q}_x : \frac{1}{2}Q_L^{JHG} \leq \underline{Q}_x \leq Q_L^{JHG} \leq \overline{Q}_x$ so that for all $Q_P \in [\underline{Q}_x, \overline{Q}_x]$ there exists unique q_L, q_H and associated aggregate trades $(q_L, q_L c_x), (q_H, q_H c_H)$ and $(Q_P, Q_P c)$ that satisfy Conditions 1, 2, 3 and 4. If, moreover, $Q_H^{JHG} > Q_L^{JHG}$, then the set of equilibrium candidates is a continuum, i.e., $\overline{Q}_x > \underline{Q}_x$.*

Note that typically (but not always) $\overline{Q}_x = Q_L^{JHG}$. Figure 3 illustrates the set of candidate allocations that are consistent with Conditions 1, 2, 3 and 4 when utility is quasi-linear.

4.3 Welfare Comparison

A wealth of equilibrium candidates is sometimes viewed with suspicion. Here it should not. Within the set of "Pooling + Separating" equilibrium candidates, no allocation is typically more plausible than another. Indeed, in most cases, high and low types Pareto-rank the set of candidates in opposite directions: high-type buyers prefer greater pooling over less, whereas for low types, increased pooling is generally accompanied by a reduction in the separating quantity, making low-risk buyers worse off overall.[24] What is evident is that high-risk buyers are always

[24] One can construct an exceptional example where, contrary to intuition, the greatest admissible pooling allocation involves low-risk buyers purchasing a separating contract on top. In such a case, trivially, the greatest admissible pooling allocation Pareto-improves upon the JHG allocation. The necessary and sufficient condition for this to occur is that there exists a $c_x > c_L$ for which the directional $(1, c_x)$-derivative of the low-risk indifference curve is zero, i.e., $D_Q \mathcal{I}_L(Q_L^{JHG}, T_L^{JHG}) + D_T \mathcal{I}_L(Q_L^{JHG}, T_L^{JHG}) c_x = 0$. If satisfied, there exists a JHG-utility-surpassing low-risk indifference curve whose slope for some (Q, T) above the low-risk separating zero-profit line emanating from the JHG allocation is c.

better off under any "Pooling + Separating" equilibrium allocation than under the Rothschild-Stiglitz allocation resulting from exclusive competition. Similar to the previous comparison, this is due to cross-subsidies from low to high-risk buyers. Moreover, in situations where adverse selection is severe and the Rothschild-Stiglitz allocation leads to significant rationing of low-risk types, the low type benefits from purchasing a greater coverage amount at a higher unit price in a "Pooling + Separating" equilibrium.

Example 3 (Quasi-Linear Utility). *Consider quasi-linear utility, i.e., preferences $U_\theta(Q,T) = U_\theta(Q) - T$ that are linear in transfers for both types. Here the amount of transfers paid does not affect preferences over coverage and is an appropriate assumption when T is small, e.g., because rare tail risks are being insured. Then the aggregate quantities Q_L and Q_H are uniquely determined due to conditional efficiency. And pooling has a purely distributional effect: more pooling benefits high-risk buyers as it lowers their premia at the expense of raised low-risk premia.*

5 Existence of Non-Fully-Separating Equilibria

The existence of a pure strategy Bayesian equilibrium is a thorny issue. As Rothschild-Stiglitz and Attar-Mariotti-Salanié show, pure strategy equilibria may fail to exist, and do so for unrelated reasons. In partially exclusive markets, pooling deviations can unravel separating allocations when the proportion of low type buyers is sufficiently large. In nonexclusive markets, Attar et al. (2022) demonstrate that the existence of a PBE can be guaranteed by imposing a further curvature assumption on the buyers' utility functions.

In this section, we establish a sweeping existence result. Maintaining the same curvature assumption as in Attar et al. (2022), we show that any set of aggregate trades that satisfy our necessary conditions can occur in an equilibrium under the "1+1" market structure. In particular, it follows from here that from a regulator's point of view this simplest never exclusive market structure is sufficient to implement any equilibrium allocation that occurs for some never exclusive and competitive market structure.

5.1 Cream-Skimming Deviations and Latent Contracts

What can destabilize an equilibrium allocation are cream-skimming deviations: a *cream-skimming deviation* is a contract (q', t') that (the more profitable) low type buyers find attractive, whereas high type buyers prefer their initial allocation (Q_H, T_H). In partially exclusive markets the possibility of cream-skimming deviations alone guarantees that any equilibrium candidate must be fully-separating. In never exclusive markets, by contrast, competing sellers have more tools at their disposal to "block" a cream-skimming deviation, i.e., render it attractive to (less profitable) high type buyers also. These tools are contracts that are not traded actively in equilibrium. The literature calls these *latent* contracts.

The purpose of the flatter curvature assumption (Assumption 6) is to identify a principal latent contract that can block large cream-skimming deviations, i.e., contracts (q', t') that (the more profitable) low type buyers find attractive on a stand-alone basis. The latent contract blocks such a cream-skimming deviation if, when combining contract (q', t') with the latent contract (q^ℓ, t^ℓ), also high type buyers find it advantageous to purchase contract (q', t'). If so, the unit cost of the deviating contract is c, and so it can never be at the same time profitable and attract the low type.[25]

5.2 Flatter Curvature Assumption

We will first introduce the flatter curvature assumption as a property that resembles increasing differences. This representation highlights why the flatter curvature assumption is sufficient to block cream-skimming deviations. We will then discuss geometric implications for the slope of translated indifference curves.[26]

Assumption 6 (flatter-curvature). *$U_L(q,t)$ is of flatter-curvature than $U_H(q,t)$ if for all $(Q_L, T_L), (Q_H, T_H)$ satisfying $U_H(Q_H, T_H) \geq U_H(Q_L, T_L)$ there exists (q^ℓ, t^ℓ) so that*

$$U_L(q', t') > U_L(Q_L, T_L) \quad \Rightarrow \quad U_H(q^\ell + q', t^\ell + t') > U_H(q^\ell + Q_L, t^\ell + T_L) \quad \textit{for all } (q', t')$$

where $U_H(q^\ell + Q_L, t^\ell + T_L) = U_H(Q_H, T_H)$.

It is readily apparent from here that the latent contract (q^ℓ, t^ℓ) does indeed block a large cream-skimming deviation (q', t'). As to the required latent contract, it is uniquely pinned down by a tangency condition: the high type's utility when trading $(Q_L + q^\ell, T_L + t^\ell)$ must lie on the on same indifference curve as (Q_H, T_H) and be tangent to average unit cost c. Since the indifference curve's slope is uniquely defined, so is the latent contract.

Lemma 3. *Posit Assumptions 5 and Condition 4. Then for given (Q_L, T_L) and (Q_H, T_H) the principal latent contract as given by Assumption 6 is unique, and for this contract it holds that*

$$\max_q U_H(Q_L + q^\ell + q, T_L + t^\ell + q\, c) = U_H(Q_L + q^\ell, T_L + t^\ell).$$

[25] The terminology principal latent contract is motivated by the fact that further latent contracts are required. Indeed, the presence of the principal latent contract invites the possibility of pivoting: complement the latent contract with a deviating contract proposed by an inactive seller. Such a deviation motivates further, derivative latent contracts that block the pivoting contract. And those derivative latent contracts invite even further pivoting deviations. Since, as we shall see, latent contracts are pricey, i.e., $t^\ell > q^\ell c$, this motivates a problem of finite regress only so that finitely many latent contracts suffice to sustain the equilibrium allocation.

[26] The importance of latent contracts in stabilizing an equilibrium is not a new insight. Assumptions 5 and 6, however presented in a new guise, are equivalent to Assumption C in Attar et al. (2022). They also provide a characterization in the context of the classical preferences over final wealth when types encode the probability of suffering a loss considered by Rothschild-Stiglitz Rothschild and Stiglitz (1976) (also more recently studied by Chade and Schlee (2012)): $U_\theta(Q, T) = p_\theta u_\theta(w - (1-Q)\ell - T) + (1 - p_\theta)u(w - T)$. More specifically, they show that if low type consumers are uniformly weakly more risk-averse, i.e., if $\min_w -\frac{v''_L(w)}{v'_L(w)} > \max_w -\frac{v''_H(w)}{v'_H(w)}$, then the flatter curvature assumption holds. CARA utility with a weakly greater coefficient of risk-aversion for low types is one common and admissible example.

Some readers may object to the flatter-curvature assumption on the grounds that it operates on an object that is endogenous to preferences, namely the set of incentive compatible allocations. This reading is perhaps misled by our presentation that emphasizes the assumption's connection to the question of equilibrium existence. To the contrary, a representation in terms of indifference curves shows that the flatter curvature assumption acts in the same domain as the more standard single-crossing assumption.

Lemma 4. *Posit Assumption 5 so that indifference curves $\mathcal{I}_\theta^{\bar{u}_\theta}(Q_\theta)$, as given by $U_\theta(Q_\theta, \mathcal{I}_\theta^{\bar{u}_\theta}(Q_\theta)) = \bar{u}_\theta$, are twice differentiable. Then Assumption 6 holds if for all $Q_H \geq Q_L$ and $\bar{u}_H, \bar{u}_L$*

$$\dot{\mathcal{I}}_H^{\bar{u}_H}(Q_H) = \dot{\mathcal{I}}_L^{\bar{u}_L}(Q_L) \qquad \Rightarrow \qquad \ddot{\mathcal{I}}_H^{\bar{u}_H}(Q_H) \geq \ddot{\mathcal{I}}_L^{\bar{u}_L}(Q_L).$$

Proof. Assumption 6 holds if

$$\begin{aligned} &U_L(Q_L + q, T_L + t) > U_L(Q_L + q, \mathcal{I}_L^{U_L(Q_L,T_L)}(Q_L + q)) \\ &\quad \Rightarrow \quad U_H(Q_H + q, T_H + t) > U_H(Q_H + q, \mathcal{I}_H^{U_H(Q_H,T_H)}(Q_H + q)). \end{aligned}$$

To see this, set $q^\ell = Q_H - Q_L$, $t^\ell = T_H - T_L$ and $q' = Q_L + q$, $t' = T_L + t$.) Equivalently, we require that for all $\bar{u}_L, \bar{u}_H$

$$\mathcal{I}_L^{\bar{u}_L}(Q_L) + t < \mathcal{I}_L^{\bar{u}_L}(Q_L + q) \quad \Rightarrow \quad \mathcal{I}_H^{\bar{u}_H}(Q_H) + t < \mathcal{I}_H^{\bar{u}_H}(Q_H + q).$$

Then set $\bar{u}_L = U_L(Q_L, T_L)$ and $\bar{u}_H = U_H(Q_H, T_H)$. Assuming differentiability of the utility functions, this is equivalent to assuming that for all quantities Q_H, Q_L and utility levels it holds that $\dot{\mathcal{I}}_H^{\bar{u}_H}(Q_H + q) \geq \dot{\mathcal{I}}_L^{\bar{u}_L}(Q_L + q)$ for $q > 0$ and $\dot{\mathcal{I}}_H^{\bar{u}_H}(Q_H + q) \leq \dot{\mathcal{I}}_L^{\bar{u}_L}(Q_L + q)$ for $q < 0$. If second-order derivatives exist, this is equivalent to requiring that $\ddot{\mathcal{I}}_H^{\bar{u}_H}(Q_H) \geq \ddot{\mathcal{I}}_L^{\bar{u}_L}(Q_L)$ when first-order derivatives coincide.[27] $\square$

Example 4 (CARA utility with exponential shocks)**.** *We follow Einav et al. (2013); Azevedo and Gottlieb (2017); Farinha Luz et al. (2023): insurance coverage $Q \in [0, 1]$ specifies the fraction of health expenditures that are reimbursed. Insurees face normally distributed health shocks. Expected utility is quasi-linear and equal to*

$$U_\theta(Q, T) = c_\theta Q - \frac{\gamma_\theta}{2}(1 - Q)^2 - T.$$

The parameter c_θ is the mean expected health shock and coincides with the seller's marginal cost. γ_θ rises in the variance of health shocks and the degree of risk aversion. Then the flatter

[27] Note that

$$\ddot{\mathcal{I}}_\theta^{\bar{u}_\theta}(Q) = \frac{-D_{11}^2 U_\theta(\cdot)(D_2 U_\theta(\cdot))^2 + 2D_{12}^2 U_\theta(\cdot) D_1 U_\theta(\cdot) D_2 U_\theta(\cdot) - D_{22}^2 U_\theta(\cdot)(D_1 U_\theta(\cdot))^2}{(D_2 U_\theta(\cdot))^3}.$$

In particular, if utility is quasi-linear, i.e., of the form $U_\theta(Q, T) = U_\theta(Q) - T$, the curvature of the indifference curve satisfies $\ddot{\mathcal{I}}_\theta^{\bar{u}_\theta}(Q_\theta) = \ddot{U}_\theta(Q)$.

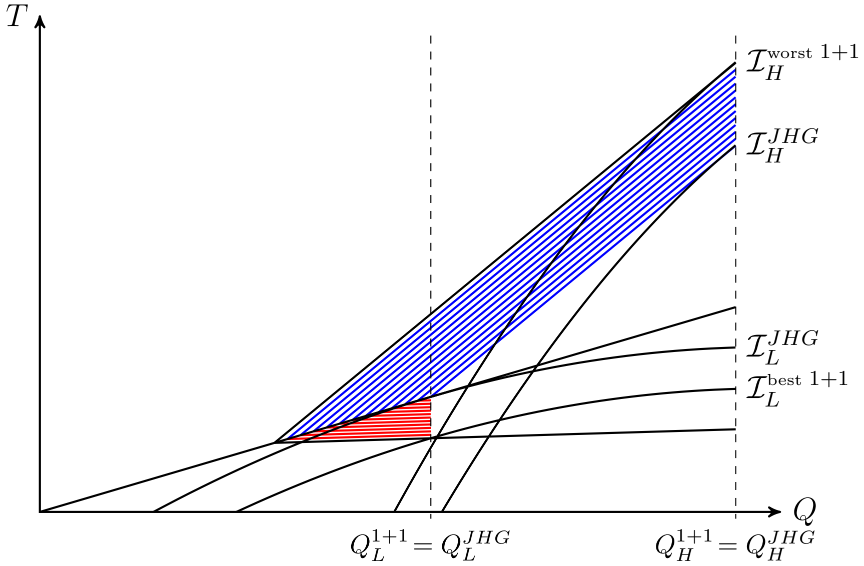


Figure 3: The "1+1" equilibria. The continuum of equilibrium allocations is drawn and located at the endpoints of the individual zero-profit lines. Partial efficiency dictates high and low-risk aggregate coverage. Under quasi-linear utility, aggregate coverage is not affected by the extent of pooling which exclusively determines the amount of cross-subsidies paid from low to high-risk buyers. Notice that all incentive constraints are slack.

curvature assumption holds iff $\gamma_H \leq \gamma_L$, meaning that high-risk buyers are either more risk-averse or indeed riskier (as measured by the variance of the normally distributed health shocks) or both.

5.3 The Equilibrium Existence Theorem

Equipped with the flatter curvature assumption, we can state our main existence result.

Theorem 2. *Posit Assumptions 1, 2, 3, 4 and 6. Any non-fully-separating aggregate active trades that satisfy Conditions 1, 2, 3 and 4 can occur in an equilibrium under the "1+1" market structure.*

This is a positive result: a PBE may fail to exist for many market structures. Yet under the stated assumptions a PBE always exists under the "1+1" market structure. Moreover, if a PBE exists for some competitive market structure, we need not look further than at the more familiar "1+1" or exlusive "1 or 1" market structures, for the same allocation can also occur as an equilibrium here.

Corollary 1. *Posit Assumptions 1, 2, 3, 4. Any equilibrium allocation that occurs under a never exclusive competitive market structure is also an equilibrium allocation under the "1+1" market structure. Any equilibrium allocation that occurs under a partially exclusive competitive market structure is also an equilibrium allocation under the "1 or 1" market structure.*

The first part of the corollary is an immediate consequence of Theorem 1 because our assumptions guarantee that any non-fully-separating equilibrium trades must satisfy Conditions

1, 2, 3 and 4 (see Proposition 1, Lemma 1 and Lemma 2). And the second part of the corollary is a reminder of Claim 1.

6 Positive Implications

In the introduction, we had emphasized two viewpoints that one can adopt vis-à-vis the concept of a market structure. We here pursue the positive viewpoint. In particular, we assess the plausibility of our equilibrium predictions and introduce an equilibrium refinement—serendipitous-aftermarket-proofness—in competitive markets.

Theorems 1 and 2 provided us with a set of aggregative active trades that occur in equilibrium. This set can be divided into two subsets: a continuum of allocations ("Pooling + Separation") and two isolated points (Rothschild-Stiglitz and full pooling). Is there a sense in which one prediction is more plausible than the other? We now propose an equilibrium refinement that will argue that "Pooling + Separating" allocations are more plausible than either a fully-pooling or a fully-separating allocation. This refinement is motivated by the conspicuous absence of dynamics from our model thus far. We now introduce dynamics, albeit in a very crude way. Fix an allocation (Q_L, T_L) and (Q_H, T_H) and define indirect utility functions

$$V_L(q,t) = U_L(Q_L + q, T_L + t) \qquad \text{and} \qquad V_H(q,t) = U_H(Q_H + q, T_H + t).$$

Our equilibrium refinement requires that conditional on the initial allocation any additional deviation contracts is loss-making. This implies that the market is in a form of rest. Future side-trading is impossible. What is crude about this definition is that ex-post trading opportunities are not anticipated.

Definition 8. *An allocation (Q_L, T_L) and (Q_H, T_H) is serendipitous-aftermarket-proof if in the aftermarket economy $((V_L, V_H), (c_L, c_H), (m_L, m_H))$ there do not exist profitable seller deviations.*

We view this definition as complementary to Hendren (2013). Whereas Hendren adopts an ex-ante perspective and asks when adverse selection shuts down the market, serendipitous-aftermarket-proofness adopts an ex-post perspective: once trading has taken place, will the market remain inactive going forward?

Proposition 3. *Under Assumptions 1, 2, 3, 4 and 6 it holds that:*

1. *Full pooling is serendipitous-aftermarket-proof if and only if full pooling is conditionally efficient, i.e.,* $0 = \arg\max_{q \geq 0} U_H(Q_P + q, T_P + qc_H)$ *and* $0 = \arg\max_{q} U_L(Q_P + Q, T_P + qc)$.
2. *Any "Pooling + Separating" allocation satisfying conditions 1, 2, 3 and 4 is serendipitous-aftermarket-proof.*

3. *The RS allocation is serendipitous-aftermarket-proof if and only if low type buyers do not wish to trade additional, competitively priced pooling contracts, i.e.,* $0 = \arg\max_{q \geq 0} U_L(Q_L + q, T_L + qc)$.

Proposition 3 further facilitates the viewpoint that market structures can emerge as an informal industry agreement between sellers. If so, serendipitous-aftermarket-proofness provides the analyst with a criterion that speaks to the plausibility of competitive allocations without further knowledge of the market structure.

To formalize this view, take a dynamic perspective and imagine that every period a new cohort of previously uninsured buyers purchases insurance policies. In a steady state, there will exist a huge population of policy holders. Some policy holders will exogenously exit from this population, e.g., due to death or expiry of existing policies. And there will be a continuous inflow of entrants who are yet to make their first purchase. Then we can ask: What would happen in such a dynamic market if the sellers had agreed, e.g., through the appropriate labeling of their policies in red or blue, or core and complementary policies, that buyers should only be able to trade according to the "1+1" market structure? Would any seller be able to approach (or rather poach) the existing policy holders and profitably offer an additional contract? Such a trading offer would be an unexpected, serendipitous opportunity for the buyer. Based on the belief that no further offers are to follow, Proposition 3 asserts that no such offer can be made if the initial allocation is serendipitous-aftermarket-proof, e.g., satisfies conditions 1, 2, 3 and 4. This gives credence to the view that the "1+1" market structure can endogenously arise as an industry agreement.

The prediction that the Rothschild-Stiglitz separating allocation emerge, on the other hand, is often implausible. Whenever an equilibrium fails to exist under an exclusive market structure, the Rothschild-Stiglitz allocation is not serendipitous-aftermarket-proof. Conversely, the Rothschild-Stiglitz allocation may occur in equilibrium under the exclusive market structure, yet not be serendiptious-aftermarket-proof.

7 Welfare

Under the normative view, a regulator can select any competitive market structure. Corollary 1 shows that this choice reduces to selecting among the partially exclusive "1+1" or the fully exclusive "1 or 1" market structure. As Figure 3 illustrates, the "1+1" market structure becomes all the more desirable if adverse selection is severe.

Our main welfare result goes beyond the mere selection of the market structure. It is motivated by the observation that equilibria under the "1+1" market structure fail an important efficiency benchmark: they are not Pareto-efficient among the set of allocations that satisfy incentive compatibility and break even for the sellers.

7.1 Incentive and Participation Efficiency

Recall that an allocation $(Q_L, T_L), (Q_H, T_H)$ is incentive compatible if

$$U_L(Q_L, T_L) \geq U_L(Q_H, T_H) \quad \text{and} \quad U_H(Q_H, T_H) \geq U_H(Q_L, T_L).$$

An allocation further satisfies participation constraints if

$$U_L(Q_L, T_L) \geq U_L(0, 0) \quad \text{and} \quad U_H(Q_H, T_H) \geq U_H(0, 0).$$

Finally, say that an allocation is feasible if

$$Q_H c_H m_H + Q_L c_L m_L \leq T_H m_H + T_L m_L.$$

As Bisin and Gottardi (2006) observe, incentive compatibility and feasibility are insurmountable obstacles to achieving greater efficiency for the planner and the market alike. Market allocations are further constrained by participation constraints. We here define the allocations that are Pareto efficient within the set of allocations that are feasible, incentive compatible and satisfy participation constraints. Following Bisin and Gottardi (2006), we call such an allocation incentive and participation efficient.

Definition 9. *An allocation is incentive and participation efficient if it satisfies participation constraints, is feasible and incentive compatible and if there does not exist another allocation* $(Q'_L, T'_L), (Q'_H, T'_H)$*, satisfying the same constraints, such that* $U_L(Q'_L, T'_L) \geq U_L(Q_L, T_L)$ *and* $U_H(Q'_H, T'_H) \geq U_H(Q_H, T_H)$ *with at least one inequality being strict.*

As argued earlier, allocations are imprecise in that they obfuscate the cross-subsidies across types that a planner may find desirable. Thus decompose any allocation $(Q_H, T_H), (Q_L, T_L)$ satisfying the aggregate resource constraint into pooling (Q_P, T_P) and separating trades $(q_L, t_L), (q_H, t_H)$ so that the sum of trades replicates the initial allocation,

$$Q_H = Q_P + q_H,\ T_H = T_P + t_H \quad \text{and} \quad Q_L = Q_P + q_L,\ T_L = T_P + t_L,$$

and each trade breaks even conditional on the buyers that trade it:

$$Q_P c = T_P \quad \text{and} \quad q_H c_H = t_H,\ q_L c_L = t_L.$$

This decomposition exists and is given by

$$Q_P = \frac{Q_H c_H - T_H}{c_H - c} = \frac{T_L - Q_L c_L}{c - c_L}, \quad q_H = \frac{T_H - c Q_H}{c_H - c} \ \text{ and } \ q_L = \frac{Q_L c - T_L}{c - c_L}.$$

This shows algebraically how greater pooling increases cross-subsidies from low to high-risk buyers. It offers, moreover, a path towards decentralization of incentive and participation efficient allocations.

7.2 Main Welfare Result

Incentive and participation efficiency is an important concept because it is a key demand that a regulator would impose on the market. The rationale behind this demand is the maxim that a constraint that does not affect a benevolent planner designing a menu of contracts must not constrain the market allocation either. To this, our preceding (positive) analysis conveys bad news. Deterred by the prospect of cream-skimming deviations, even partially nonexclusive market structures fail to be efficient. To restore efficiency, a stronger market intervention that goes beyond merely changing the market structure is required.

Our proposal is markedly simple and involves both a market structure and contract regulation. First, we propose the implementation of the "1+1" market structure, where buyers acquire one contract from each of two seller groups. By facilitating "Pooling + Separating" allocations, this market structures implements implicit transfers from low to high-risk buyers. Second, we suggest to expand upon the extent to which low-risk buyers can seek out additional coverage at actuarial fair low cost. To achieve this, we propose an additional minimum quantity requirement on group 1 contracts that emerge as pooling contracts in equilibrium. This requirement is designed to prevent group 1 sellers from exclusively attracting low-risk buyers through cream-skimming, where less coverage is offered at more attractive prices. Efficiency improves for both risk profiles even if the regulation merely replicates the present pooling quantity Q_P: In the absence of destabilizing cream-skimming deviations in group 1, there is scope for group 2 sellers to offer additional low cost coverage to low-risk types up to the incentive efficient level. Finally, we stipulate that group 2 contracts must be limit orders, allowing buyers to trade any fraction $(\alpha q, \alpha t)$ with $\alpha \in [0,1]$ of a contract (q,t) offered in group 2. Although inactive in equilibrium, the option to purchase partial coverage is required to sustain an equilibrium in the odd case where both types' incentive constraints are slack. The reason is that, as demonstrated in Theorem 2, latent contracts can deter group 1 deviations that promise coverage in excess of Q_P.

Our main welfare result is as follows:[28]

Theorem 3. *Posit Assumptions 1, 2, 3 and 4. Consider any incentive and participation efficient allocation* $(Q_L, T_L), (Q_H, T_H)$ *and let* $(Q_P, Q_P c), (q_H, q_H c_H), (q_L, q_L c_L)$ *be the aggregate trades that decentralize it. If aggregate trades satisfy large pooling, i.e., Condition 3 whereby* $q_L(c_H - c_L) \leq Q_P(c_H - c)$*, are partially efficient, i.e.,* $Q_P \geq \arg\max_{\tilde{Q}} U_L(\tilde{Q} + q_L, \tilde{Q}c + q_L c_L)$ *and are non-negative, i.e.,* $Q_P, q_H, q_L \geq 0$*, then this second-best allocation occurs as an equilibrium under the "1+1" market structure if the regulator imposes the minimum quantity requirement*

[28] Partial efficiency rules out pooling deviations that, as in the original Rothschild and Stiglitz (1976) model prevent the existence of a pure strategy equilibrium. If we imposed the flatter curvature assumption here and relied on latent contracts (as we do not do here) once more, then partial efficiency would be superfluous to the result.

that any buyer making a purchase must purchase a quantity weakly greater than Q_P in group 1 and group 2 contracts are limit orders.

We wish to emphasize that a minimum quantity requirement on group 1 contracts does not only Pareto improve upon the equilibrium under the unregulated "1+1" market structure, but also obviates the necessity of implausible latent contracts that are guaranteed to incur a loss were they ever mistakenly traded. Any contract offered in equilibrium will be actively traded.

Discussion: From a theoretical standpoint, expanding the regulation of group 1 contracts to entirely determine the level of group 1 coverage (where Q_P serves as both a lower and upper bound) might seem advantageous. This approach would eliminate altogether the possibility of deviations within group 1 that could disrupt an incentive and participation efficient allocation. Under this stronger regulatory intervention, Theorem 3 continues to hold, although without the necessity for group 2 offers to be limit orders. Market orders, as indicated in our positive analysis, would be sufficient to sustain the allocation. However, from a practical perspective, we do not find this approach desirable. Imposing an upper limit on group 1 coverage could have unfavorable implications. It would essentially prohibit the inclusion of new advanced treatments in basic plans, potentially resulting in long-term stagnation in the quality of care provided.

7.3 Utilitarian Welfare

We conclude this section by offering an interpretation of our main welfare result in terms of utilitarian welfare maximization.

Surely, if an allocation satisfies participation constraints and maximizes utilitarian welfare for some non-zero $\lambda = (\lambda_L, \lambda_H) \in \mathbb{R}^2_+$ $(\lambda_L + \lambda_H = 1)$ welfare weights,

$$\max_{(\tilde{Q}_H,\tilde{T}_H),(\tilde{Q}_L,\tilde{T}_L)} U_L(\tilde{Q}_L,\tilde{T}_L)\lambda_L + U_H(\tilde{Q}_H,\tilde{T}_H)\lambda_H \quad \text{s.t.} \quad \begin{cases} Q_H c_H m_H + Q_L c_L m_L = T_H m_H + T_L m_L \\ U_L(\tilde{Q}_L,\tilde{T}_L) \geq U_L(\tilde{Q}_H,\tilde{T}_H) \\ U_H(\tilde{Q}_H,\tilde{T}_H) \geq U_H(\tilde{Q}_L,\tilde{T}_L), \end{cases}$$

i.e., it is second-best efficient in the sense of Harris and Townsend (1981); Crocker and Snow (1985), then it is incentive and participation efficient. Although the reverse need not be true,[29] welfare weights greatly facilitate the interpretation of the implications of greater or lesser pooling. Specifically, we can identify changes in the degree of pooling along the Pareto-frontier with the identity of the buyers that stand to benefit. This highlights that the minimum quantity requirement serves as a policy tool whose level is ranked in opposite directions by low and high-risk individuals. Whereas high-risk individuals prefer greater minimum quantity requirements, low-risk individuals (along the incentive and participation efficient frontier) prefer lower

[29] Since utility is non-linear in the allocation, the set of incentive compatible allocations need not be convex. In the quasi-linear case, convexity (and hence equivalence of the two efficiency notions) would hold if $\ddot{U}_H(Q) \geq 0 \geq \ddot{U}_L(Q)$.

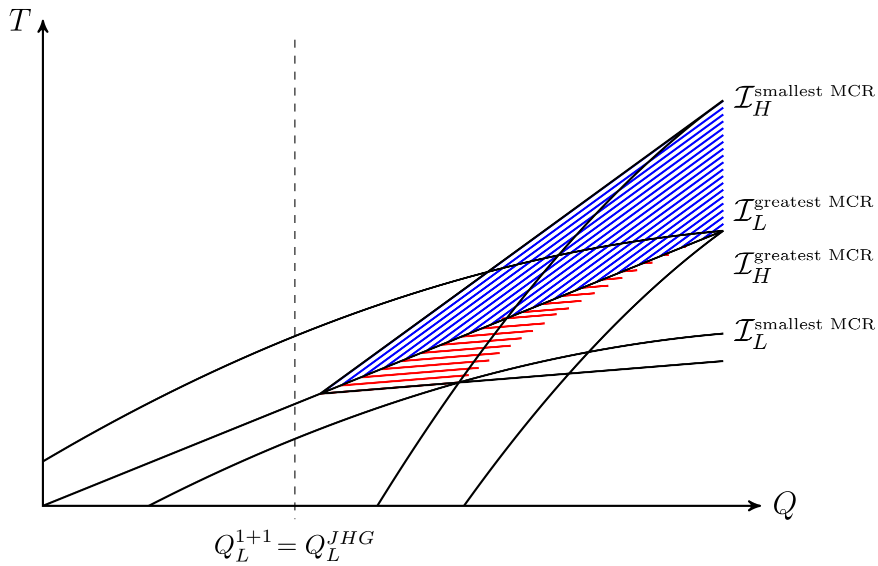


Figure 4: Efficient allocations attainable via group 1 coverage requirements. A continuum of allocations is drawn. Allocations are located at the endpoint of zero profit lines. The group 1 minimum coverage requirement (MCR) effectively dictates the extent of pooling coverage. This is located at the origin of individual zero profit lines. Since separating coverage is large under incentive efficiency, large pooling implies that the MCR exceeds the JHG level of coverage. Notice that incentive constraints bind and determine the extent of low-risk buyers' separating coverage.

minimum quantity requirements. Critically, a regulator can attain second-best efficiency for a wide range of welfare weights, although these tend to favor high-risk buyers. In particular, large pooling and partial efficiency raise the level of required pooling coverage and as such may constrain low-risk buyers' welfare. We define $\underline{L}$ the smallest bound so that Q_P satisfies large pooling and partial efficiency for all $\frac{\lambda_H}{\lambda_L} \geq \overline{L}$ and $\overline{L}$ the greatest bound so that q_L remains positive for all $\frac{\lambda_H}{\lambda_L} \leq \overline{L}$ and satisfies low-risk participation constraints.

Lemma 5. *If utility is quasi-linear, bounds $\underline{L}$ and $\overline{L}$ are well-defined.*

Critically, pooling coverage Q_P is bounded from below and rises in the relative welfare weight placed on high-risk buyers. Quasi-linearity further allows to deduce that equilibrium separating quantities, denoted $\mathbf{q}_L(Q_P)$ and $\mathbf{q}_H(Q_P)$, are non-increasing in Q_P. The low-risk buyer allocation $Q_P \mapsto Q_P + \mathbf{q}_L(Q_P)$ is increasing due to the relaxed incentive constraint.

Corollary 2. *Suppose that utility is quasi-linear and posit Assumption 1, 2, 3, 4. Then consider any second-best efficient allocation $(Q_L, T_L), (Q_H, T_H)$ and let $(Q_P, T_P), (q_H, t_H), (q_L, t_L)$ be the aggregate trades that decentralize it. If the associated welfare weights λ_L, λ_H are intermediate, i.e., $\frac{\lambda_H}{\lambda_L} \in [\underline{L}, \overline{L}]$, then this second-best allocation occurs as an equilibrium under the "1+1" market structure if the regulator imposes the minimum quantity requirement that any buyer making a purchase must purchase a quantity weakly greater than Q_P in group 1 and group 2 contracts are limit orders.*

8 Discussion of the Model with Multiple Types

An obvious limitation of our work—albeit common since the celebrated work by Rothschild and Stiglitz (1976)—is our focus on binary types. With more than two types, the "1+1" market structure would surely not encompass all non-fully separating equilibria for an arbitrary market structure. One could imagine an allocation where low, middle, and high-risk buyers can potentially buy three distinct contracts. Then consider pooling on contract $(q_{1,C}, t_{1,C})$. And further suppose that low-risk buyers purchase a further separating contract $(q_{2,S}, t_{2,S})$, whereas middle and high-risk buyers are assigned a "Pooling + Separating" allocation and wound up purchasing contracts $(q_{2,C}, t_{2,C}), (q_{3,H}, t_{3,H})$ and $(q_{2,C}, t_{2,C}), (q_{3,L}, t_{3,L})$ respectively instead. The "1+1" market structure can impossibly replicate this allocation, a third group would be required. Furthermore, with many types it is conceivable that moving beyond the "1+1" market structure is Pareto-improving: If there is pooling in group 1, yet separation in group 2, there is likely scope for further partial pooling among types $\{k, ..., N\}$ in group 2 that would analogously increase welfare as in the two-type model.

We would like to offer two comments on the model with multiple types. First we note that the "1+1" market structure admits a straightforward extension to multi-type models. The objective of this generalized market structure is to facilitate partial pooling among a subset of types. It can be visualized like the famous centipede game in game theory: Given N types that satisfy single-crossing, consider $2N-2$ groups so that for $k \in \{1, ..., N-2\}$ an agent that makes a purchase in group $2k$ is ineligible to make a purchase in any group $k' > 2k$. This invites an interpretation to think of groups $\{1, 3, 5, ..., 2N-3\}$ as continuation groups, for they allow the agent to make further purchases, and groups $\{2, 4, ..., 2N-2\}$ as stopping groups, for any purchase here forfeits the right to make further purchases. Observe that the induced market structure when only considering the final groups $2N-3$ and $2N-2$ corresponds to the "1+1" market structure studied in the main text. The centipede market structure naturally suggests an equilibrium in which types $\{k, ..., N\}$ (with k the lowest and N the highest risk type) all purchase the same pooling contracts in groups $\{1, 3, ..., 2k-1\}$, whereas type k then makes a purchase in group $\{2k\}$ and types $\{k+1, ..., N\}$ go on to make a purchase in group $\{2k+1\}$ instead. Existence would be a daunting issue, however. Minimum quantity requirements in continuation groups or a extensive rather than a simultaneous-move game form could possibly sidestep these.[30]

Our second comment is more practical. Extensive centipede-like market structures may work well in perfectly competitive markets. But this theory de-emphasizes search and informa-

[30] In an extensive game the regulator could stipulate that sellers start posting contracts in group $k+1$ only once contracts have been irreversibly posted in group k. A distinction would have to be made as to whether buyers make purchases right after contracts in group k have been posted, or only once contracts in all groups have been posted. Future work of ours is to address these issues. The spirit of the exercise should be clear, however: In line with Attar et al. (2021), a sequential auction-like mechanism that breaks the simultaneity of moves cancels many destabilizing deviations without altering the competitive character of the game. And where a simultaneous-move game must wrestle with equilibrium existence (requiring ever-more restrictive assumption akin to the flatter curvature assumption), the extensive-form would neatly bypass these issues.

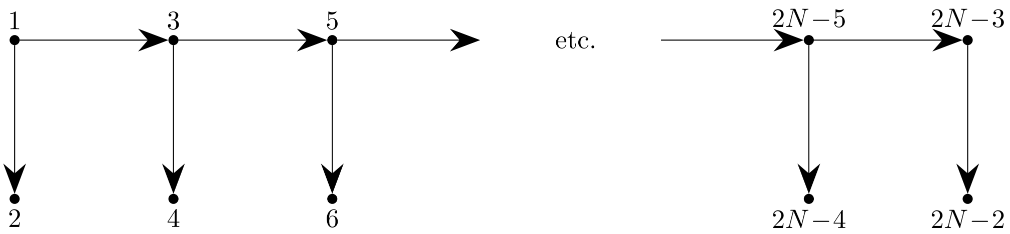


Figure 5: The centipede market structure

tion frictions. As the stakes of individual contracts become smaller due to the sheer number of contracts purchased, buyers' willingness to explore less established sellers will likely diminish. Reduced incentives for further search, disregarded in our model, will likely diminish the competitiveness of markets governed by more intricate market structures. A further reason to prioritize the "1+1" market structure over more intricate designs is rooted in political economy considerations. Indeed, the default option for policy makers when intervening in markets where adverse selection prevails is to propose an exclusive market structure. Regulation of contract characteristics in US healthcare markets ('platinum', 'gold', 'silver', or 'bronze') or Chilean post-retirement annuity contracts (where the pension law stipulates that every annuity contract purchased must provide a minimum pension requirement) are cases in point. In the spirit of policy incrementalism, the "1+1" market structure will well enhance flexibility over markets that are—without theoretical justification—to date designed to be fully exclusive. In light of this reality, experimenting with the simplest form of partially exclusive competition first, namely the "1+1" market structure, appears to be the more prudent design.

9 Conclusion

This paper revisits the canonical model of competitive markets plagued by adverse selection. We propose the concept of a market structure and show that under severe adverse selection a compromise between fully exclusive and nonexclusive competition results in Pareto-improving equilibrium allocations that have never been studied before. Our analysis singles out the "1+1" market structure whereby buyers can purchase a single contract from each of two groups of sellers.

As a key result we show that, despite the presence of adverse selection, a minimal regulation—imposing the "1+1" market structure coupled with a minimum quantity requirement in group 1—results in a second-best efficient equilibrium allocation. Since different risk types value greater minimum quantity requirements in opposite directions, changes in said required quantity result in moves along the Pareto-frontier. Thus our minimal regulation plays a similar redistributive role as do initial allocations in perfectly competitive markets without information asymmetries as highlighted by the second-welfare theorem.

Appendix

A Necessary Conditions: missing proofs

A.1 Proof of Claim 1

Proof. Fix a partially exclusive and competitive market structure and consider an equilibrium allocation (Q_L, T_L) and (Q_H, T_H).

Step 1. we show that $T_L = Q_L c_L$. First, $T_L \geq Q_L c_L$. Otherwise, there exists at least one seller who makes a negative profit. Then this seller can choose to be inactive instead. Second, $T_L \leq Q_L c_L$. Otherwise, suppose that $T_L > Q_L c_L + \epsilon$. Since the market structure is competitive and partially exclusive, there exist at least three exclusive sellers, i.e., sellers k for whom $\max_{M \in \mathcal{M}: k \in M} |M| = 1$. Then at least one of these three is inactive and makes zero profit on-path. A profitable deviation for an exclusive seller consists in offering a cream-skimming contract (q', t') so that $U_L(q', t') > U_L(Q_L, T_L)$ and $U_H(q', t') < U_H(Q_L, T_L)$. Following standard arguments (using single-crossing and continuity of the utility function), such a contract (q', t') always exists and can be chosen to be arbitrarily close to (Q_L, T_L). In effect, one can choose a (q', t') that is profitable conditional on trading with low type buyers only: $q' < Q_L + \frac{\epsilon}{2}$ and $t' > T_L - \frac{\epsilon}{2} > Q_L c_L + \frac{\epsilon}{2} > q' c_L + \frac{\epsilon}{2}$. And since $U_H(Q_L, T_L) \leq U_H(Q_H, T_H)$ due to incentive compatibility, the exclusive contract (q', t') only attracts low type buyers.

Step 2. we show that $T_H = Q_H c_H$. First, step 1 implies that in equilibrium no pooling contract can be actively traded. Since sellers serving high types cannot make a negative profit, it follows that $T_H \geq Q_H c_H$. And due to Bertrand's competition, the unit price for serving high type buyers must be smaller or equal to c_H. As a result, we have that $T_H = Q_H c_H$.

Step 3. we observe that the allocation must be efficient. This means that $Q_H = \arg\max_{Q \geq 0} U_H(Q, c_H Q)$, and $Q_L = \arg\max_{Q \geq 0} U_L(Q, c_L Q)$ subject to high type incentive compatibility, i.e., $U_H(Q_H, Q_H c_H) \geq U_H(Q_L, Q_L c_L)$. But this is the Rothschild-Stiglitz allocation so efficiency follows from their familiar arguments.

□

A.2 Proof of Claim 2

Proof. Denote (Q_H, T_H) and (Q_L, T_L) the equilibrium allocation for high and low type buyers. Then notice that any fully-separating equilibrium allocation must satisfy $T_H = c_H Q_H$ and $T_L \leq c Q_L$. This follows from probing the equilibrium candidate with undercutting deviations familiar from Bertrand's competition. (Actually, it holds that $T_L = c_L Q_L$, but we do not require this here.) The bounds on profit imply that there exists a seller k who actively trades a contract (q', t') with low type buyers such that $t' \leq c q'$. And since the market structure is never exclusive, there exists another seller j_0 who can jointly trade with seller k. If seller j_0 is inactive, denote $j = j_0$. If this seller is active instead, denote j seller j_0's inactive replacement, i.e., $j \notin M_L$, yet $M_L \cup \{j\} \backslash_{j_0} \in \mathcal{M}$. (Seller j_0's replacement j exists because the market structure is competitive.) Then seller j can propose a contract $(Q_H - q', T_H - t' - \epsilon)$. This deviation attracts high types buyers, because trading jointly with sellers k and j following j's deviation gives high type buyers strictly greater utility than the initial allocation which was preferred among all trades excluding seller j: $U_H(Q_H - q' + q', T_H - t' - \epsilon + t') \geq U_H(Q_H, T_H) = \max_{M \in \mathcal{M}: j \notin M} U_H(\sum_{i \in M} q^i, \sum_{i \in M} t^i)$. And attracting high type buyers suffices to render this deviation profitable; profit is $T_H - t' - \epsilon - c_H(Q_H - q') = c_H q' - t' - \epsilon \geq c_H q' - cq' - \epsilon$ which is positive for ϵ sufficiently small.

□

A.3 Proof of Proposition 1

Proof. If (Q_L, T_L) and (Q_H, T_H) is an equilibrium allocation, then there exists an equilibrium in which, first, sellers collectively offer the menu $\{(q^k, t^k)\}_{k \in \mathcal{K}}$ and, second, buyer types L and H trade with sellers M_L and M_H in $\mathcal{M}$ so that

$$(Q_L, T_L) = \sum_{k \in M_L} (q^k, t^k) \quad \text{and} \quad (Q_H, T_H) = \sum_{k \in M_H} (q^k, t^k).$$

Then define the pooling component $(Q_P, T_P) = \sum_{p \in M_L \cap M_H} (q^p, t^p)$ and the separating components $(q_L, t_L) = \sum_{\ell \in M_L \backslash M_H} (q^\ell, t^\ell)$ and $(q_H, t_H) = \sum_{h \in M_H \backslash M_L} (q^h, t^h)$ as in Condition 2. Item (i) trivially holds.

Next observe that seller j's profit is proportional to

$$t^j - c' q^j \qquad \text{where } c' = \begin{cases} c_L & \text{if } j \in M_L \backslash M_H \\ c_H & \text{if } j \in M_H \backslash M_L \\ c & \text{if } j \in M_H \cap M_L. \end{cases}$$

Since seller j can offer the null trade $(0, 0)$ instead, he cannot make a loss in equilibrium. It follows that the unit price of serving the low type θ is at least c_L, the unit price of serving the high type θ is at least c_H and the unit price of serving both types is at least c.

Then we prove that $t^j = q^j c_H$ for all $j \in M_H \setminus M_L$ by drawing on undercutting deviations as in Bertrand's competition. For there to be viable competitors that can undercut we will require that $t^p = q^p c$ for all $p \in M_L \cap M_H$. This is proven by drawing on a pivoting deviation instead.

Step 1: We show by contradiction that all sellers in $M_L \cap M_H$ make zero profit. Or, suppose that $j \in M_L \cap M_H$ made a positive profit. Since the market structure is competitive, there exists $k \in \mathcal{K} \setminus \{j\}$ so that $M_L \cup \{k\} \backslash \{j\} \in \mathcal{M}$. In particular, seller k does not actively trade in

$$\begin{cases} j \in M_L \cap M_H & \Rightarrow \quad \pi^j = 0 \text{ (step 1)} \\ j \in M_H \setminus M_L & \Rightarrow \quad \begin{cases} \pi^j = 0 & \text{if } |M_H \setminus M_L| > 1 \text{ (step 2)} \\ \pi^j = 0 & \text{if } |M_H \setminus M_L| = 1 \text{ (step 5)} \end{cases} \\ j \in M_L \setminus M_H & \Rightarrow \quad \begin{cases} t^j \leq q^j\, c & \text{if } |M_L \setminus M_H| > 1 \text{ (step 3)} \\ t^j \leq q^j\, c & \text{if } |M_L \setminus M_H| = 1 \text{ (step 6)} \end{cases} \\ |M_L \setminus M_H| = 1 & \text{(step 4).} \end{cases}$$

Figure 6: A roadmap of the proof of Proposition 1

equilibrium and makes zero profit. If so, seller k could propose the contract $(q^j, t^j - \epsilon)$ so that $t^j - \epsilon > q^j c$. In effect, the deviation is profitable conditional on trading with both buyer types.

Then denote $((\hat{q}^i, \hat{t}^i))_{i \in \mathcal{K}}$ the new menu of contracts following seller k's deviation. Clearly, due to monotonicity of transfers, for both $\theta \in \{L, H\}$

$$\max_{\substack{M \in \mathcal{M} \\ k \in M}} U_\theta(\sum_{i \in M} \hat{q}^i, \sum_{i \in M} \hat{t}^i) \geq U_\theta(\sum_{i \in M_\theta \cup \{k\} \setminus \{j\}} \hat{q}^i, \sum_{i \in M_\theta \cup \{k\} \setminus \{j\}} \hat{t}^i) = U_\theta(\sum_{i \in M_\theta} q^i, -\epsilon + \sum_{i \in M_\theta} t^i) > U_\theta(\sum_{i \in M_\theta} q^i, \sum_{i \in M_\theta} t^i)$$
$$= \max_{M \in \mathcal{M}} U_\theta(\sum_{i \in M} q^i, \sum_{i \in M} t^i) \geq \max_{\substack{M \in \mathcal{M} \\ k \notin M}} U_\theta(\sum_{i \in M} q^i, \sum_{i \in M} t^i) = \max_{\substack{M \in \mathcal{M} \\ k \notin M}} U_\theta(\sum_{i \in M} \hat{q}^i, \sum_{i \in M} \hat{t}^i).$$

So, following the deviation both buyer types trade with seller k, thereby rendering seller k's deviating contract $(q^j, t^j - \epsilon)$ strictly profitable.

Step 2: We show by contradiction that if there is more than one seller only serving high type buyers, then all those sellers make zero profit. Or, suppose that some seller $j \in M_H \backslash M_L$ made a positive profit and $|M_H \backslash M_L| > 1$. Then there exists another seller $k \in M_H \backslash (M_L \cup \{j\})$ that could deviate and offer $(q^j + q^k, t^j - \epsilon + t^k)$ instead where $t^j - \epsilon > q^j c_H$. Since (q^k, t^k) is weakly profitable, the deviation is strictly profitable conditional on continued trade with high type buyers (and even more profitable if it also attracts low type buyers).

To see that high type buyers will trade with seller k following the deviation, denote $((\hat{q}^i, \hat{t}^i))_{i \in \mathcal{K}}$ the new menu of contracts and note that

$$\max_{\substack{M \in \mathcal{M} \\ k \in M}} U_H(\sum_{i \in M} \hat{q}^i, \sum_{i \in M} \hat{t}^i) \geq U_H(\sum_{i \in M_H \backslash \{j\}} \hat{q}^i, \sum_{i \in M_H \backslash \{j\}} \hat{t}^i) = U_H(\sum_{i \in M_H} q^i, -\epsilon + \sum_{i \in M_H} t^i) > U_H(\sum_{i \in M_H} q^i, \sum_{i \in M_H} t^i)$$
$$= \max_{M \in \mathcal{M}} U_H(\sum_{i \in M} q^i, \sum_{i \in M} t^i) \geq \max_{\substack{M \in \mathcal{M} \\ k \notin M}} U_H(\sum_{i \in M} q^i, \sum_{i \in M} t^i) = \max_{\substack{M \in \mathcal{M} \\ k \notin M}} U_H(\sum_{i \in M} \hat{q}^i, \sum_{i \in M} \hat{t}^i).$$

Step 3: We claim that if there is more than one seller only serving low type buyers, then unit profit must be bounded by the cost of serving both types, i.e., $t^j \leq q^j\, c$ for all $j \in M_L \setminus M_H$. This employs essentially identical arguments as in Step 2 where the undercutting deviation comes from a seller $k \in M_L \setminus M_H$.

Step 4: We show by contradiction that there is at most one seller only serving low type buyers. This corresponds to item (ii) of Condition 2. Or, suppose that there exist distinct $j_1, j_2 \in M_L \backslash M_H$. Since the market structure is competitive, there exist distinct $k_1, k_2 \in \mathcal{M} \backslash M_L$ so that $M_L \cup \{k_1\} \setminus \{j_1\}$ and $M_L \cup \{k_2\} \setminus \{j_1\}$ in $\mathcal{M}$. Then either at least one of the two sellers k_1, k_2 does not belong to M_H, i.e., one of the two is an inactive trader that makes zero profit. Or they both belong to M_H (but not M_L) in which case they must also make zero profit due to step 2.[31]

Then consider the seller k_1 pivoting deviation: propose the contract $(q_H - q_L^2, t_H - t_L^2 - \epsilon)$ with $t_H - t_L^2 - c_H(q_H - q_L^2) > \epsilon$. By jointly trading with sellers in $(M_L \cap M_H) \cup \{j_2\}$ and the pivoting seller k_1, buyers trade the aggregate quantity $Q_H = Q_P + q_H$ at the lesser price $T_H - \epsilon = T_P + t_H - \epsilon$. Moreover, since $t_L^2 \le q_L^2\, c$, this deviation is profitable conditional on trading with high type buyers (and even more profitable if it also attracts low type buyers).

We verify that the deviation attracts high type buyers. Observe that $(M_L \cap M_H) \cup \{j_2\} \cup \{k_1\} \subset M_L \cup \{k_1\} \setminus \{j_1\}$. Since the market structure is competitive, $M_L \cup \{k_1\} \setminus \{j_1\} \in \mathcal{M}$ implies that buyers can drop trades and trade with sellers $(M_L \cap M_H) \cup \{j_2\} \cup \{k_1\} \in \mathcal{M}$. Then denote $((\hat{q}^i, \hat{t}^i))_{i\in\mathcal{K}}$ the new menu of contracts following seller k's deviation.

$$\max_{\substack{M\in\mathcal{M}\\ k_1\in M}} U_H\Big(\sum_{i\in M}\hat{q}^i, \sum_{i\in M}\hat{t}^i\Big) \ge U_H\Big(\sum_{i\in(M_L\cap M_H)\cup\{j_2\}\cup\{k_1\}}\hat{q}^i, \sum_{i\in(M_L\cap M_H)\cup\{j_2\}\cup\{k_1\}}\hat{t}^i\Big) = U_H\Big(\sum_{i\in M_H} q^i, -\epsilon + \sum_{i\in M_H} t^i\Big)$$
$$> U_H\Big(\sum_{i\in M_H} q^i, \sum_{i\in M_H} t^i\Big) = \max_{M\in\mathcal{M}} U_H\Big(\sum_{i\in M} q^i, \sum_{i\in M} t^i\Big) \ge \max_{\substack{M\in\mathcal{M}\\ k_1\notin M}} U_H\Big(\sum_{i\in M} q^i, \sum_{i\in M} t^i\Big) = \max_{\substack{M\in\mathcal{M}\\ k_1\notin M}} U_H\Big(\sum_{i\in M}\hat{q}^i, \sum_{i\in M}\hat{t}^i\Big).$$

And so high type buyers are strictly better off by trading with seller k_1 following the deviation.

Step 5: We show by contradiction that if there is exactly one seller only serving high type buyers, then this seller makes zero profit. Or, suppose that $|M_H \setminus M_L| = 1$ and seller $j \in M_H \backslash M_L$ made a positive profit. Since sellers are twice replaceable, there exist distinct sellers $k_1, k_2 \in \mathcal{K} \backslash M_H$ so that $M_H \cup \{k_1\} \backslash \{j\} \in \mathcal{M}$ and $M_H \cup \{k_2\} \backslash \{j\} \in \mathcal{M}$. Since $|M_L \backslash M_H| = 1$ due to step 4, at least one of the two, say k_1, must be an inactive seller in $\mathcal{K} \setminus (M_L \cup M_H)$. Then the deviating contract $(q^j, t^j - \epsilon)$ where $t^j - \epsilon > q^j\, c_H$ is strictly profitable for the inactive seller k_1 conditional on trading with high type buyers (and even more profitable if it also attracts low type buyers). And following analogous arguments as before, trading with sellers $M_H \cup \{k_1\} \backslash \{j\}$ is feasible because the market structure is competitive and gives strictly greater utility to high type buyers than not trading with seller k_1: $\max_{\substack{M\in\mathcal{M}\\ k_1\in M}} U_H(\sum_{i\in M}\hat{q}^i, \sum_{i\in M}\hat{t}^i) > \max_{\substack{M\in\mathcal{M}\\ k_1\in M}} U_H(\sum_{i\in M}\hat{q}^i, \sum_{i\notin M}\hat{t}^i)$.

Step 6: We show by contradiction that if there is exactly one seller only serving low type buyers, then this seller's unit price must be bounded by the marginal pooling cost c. Or, suppose that $|M_L \setminus M_H| = 1$ and seller $j \in M_L \setminus M_H$'s contract satisfies $t^j > q^j\, c$. Then, following symmetric arguments as in step 5, an inactive seller can propose an undercutting

[31] Step 4 and 5 are the only instances where we use that sellers are twice replaceable (cf. Definition 2). If sellers were only once replaceable, we could not rule out that seller $\ell \in M_L \setminus M_H$'s replacement is $h \in M_H \setminus M_L$ and seller H's replacement is ℓ and they both make a profit by serving the low and the high type respectively.

contract and attract low type buyers:Since sellers are twice replaceable, there exist distinct sellers $k_1, k_2 \in \mathcal{K} \setminus M_L$ so that $M_L \cup \{k_1\} \setminus \{j\} \in \mathcal{M}$ and $M_L \cup \{k_2\} \setminus \{j\} \in \mathcal{M}$. Since $|M_H \setminus M_L| = 1$ due to step 4, at least one of the two, say k_1, must be an inactive seller in $\mathcal{K} \setminus (M_L \cup M_H)$. Then the deviating contract $(q^j, t^j - \epsilon)$ where $t^j - \epsilon > q^j\, c$ is strictly profitable for the inactive seller k_1 conditional on trading with buyers of both types (and even more profitable if it only attracts low type buyers). Finally, trading with sellers $M_L \cup \{k_1\} \setminus \{j\}$ is feasible because the market structure is competitive. And, following analogous arguments as before low type buyers are better off when trading with seller k_1 following the undercutting deviation: $\max_{\substack{M\in\mathcal{M} \\ k_1\in M}} U_L(\sum_{i\in M} \hat{q}^i, \sum_{i\in M} \hat{t}^i) > \max_{\substack{M\in\mathcal{M} \\ k_1\in M}} U_L(\sum_{i\in M} \hat{q}^i, \sum_{i\notin M} \hat{t}^i)$. □

A.4 Proof of Lemma 2

Proof. (1) We first show that $Q_P \in \arg\max_q U_L(Q + q_L, Qc + t_L)$ whenever $Q_P > 0$. Otherwise, a seller k actively trading a pooling contract has a profitable deviation.

Suppose by contradiction that the pooling quantity $Q_P > 0$ satisfies $U_L(q_L + q', t_L + q'c) > U_L(Q_P + q_L, Q_P c + t_L)$ for some $q' > 0$. Then there exists $\epsilon > 0$ so that $U_L(q_L + q', t_L + q'c + \epsilon) > U_L(Q_L, T_L)$. And seller k trading a (competitively priced) pooling contract can profitably deviate by proposing the contract $(q', q'c + \epsilon)$ instead of $(q^k, q^k c)$. This deviation is profitable conditional on trading with both buyer types and even more profitable if it only attracts low buyer types. And low type buyers are strictly better off following seller k's deviation and will want to trade the deviating contract (possibly by dropping all other pooling contracts). To see this, denote $((\hat{q}^i, \hat{t}^i))_{i\in\mathcal{K}}$ the new menu of contracts following seller k's deviation. We find that

$$\max_{\substack{M\in\mathcal{M} \\ k\in M}} U_L(\sum_{i\in M} \hat{q}^i, \sum_{i\in M} \hat{t}^i) \geq U_L(q' + q_L, q'c + \epsilon + t_L)$$
$$> U_L(Q_L, T_L) = \max_{M\in\mathcal{M}} U_L(\sum_{i\in M} q^i, \sum_{i\in M} t^i) \geq \max_{\substack{M\in\mathcal{M} \\ k\notin M}} U_L(\sum_{i\in M} q^i, \sum_{i\in M} t^i) = \max_{\substack{M\in\mathcal{M} \\ k\notin M}} U_H(\sum_{i\in M} \hat{q}^i, \sum_{i\in M} \hat{t}^i).$$

(2) Second, we show using analogous arguments that $q_H \in \arg\max_q U_H(Q_P + q, T_P + q\, c_H)$ whenever $q_H > 0$. Otherwise, a seller h actively trading a separating contract with the high type has a profitable deviation.

Suppose by contradiction that there exists $q' > 0$ such that $U_H(Q_P + q', Q_P c + q' c_H) > U_H(Q_P + q_H, T_P + t_H)$. Then there exists $\epsilon > 0$ so that $U_H(Q_P + q', Q_P c + q' c_H + \epsilon) > U_H(Q_H, T_H)$. Then a seller h actively trading a (competitively priced) separating contract $(q^h, q^h c_H)$ can profitably deviate by proposing the contract $(q', q'\, c_H + \epsilon)$ instead of $(q^h, q^h c_H)$. This deviation is profitable conditional on trading only with high type buyers and even more profitable if it also attracts low buyer types. And high type buyers are strictly better off the following seller k's deviation and will want to trade the deviating contract (possibly by dropping all other separating contracts). To see this, denote $((\hat{q}^i, \hat{t}^i))_{i\in\mathcal{K}}$ the new menu of contracts following seller k's deviation. We find that

$$\max_{\substack{M\in\mathcal{M}\\k\in M}} U_H(\sum_{i\in M}\hat{q}^i,\sum_{i\in M}\hat{t}^i) \geq U_H(Q_P+q',T_P+q'\,c_H+\epsilon) > U_H(Q_H,T_H)$$
$$= \max_{M\in\mathcal{M}} U_H(\sum_{i\in M}q^i,\sum_{i\in M}t^i) \geq \max_{\substack{M\in\mathcal{M}\\k\notin M}} U_H(\sum_{i\in M}q^i,\sum_{i\in M}t^i) = \max_{\substack{M\in\mathcal{M}\\k\notin M}} U_H(\sum_{i\in M}\hat{q}^i,\sum_{i\in M}\hat{t}^i).$$

(3) Finally, to understand the qualifying condition when $q_H = q_L = 0$ (or equivalently $Q_H = Q_P$), observe that $\max_{q\geq 0} U_H(\frac{1}{2}Q_P + q, \frac{1}{2}Q_P c + qc_H)$ the utility from purchasing only half the pooling allocation and the most desired separating quantity at unit price c_H is equal to

$$\min_{\alpha\in[0,1]} \max\{\max_{q\geq 0} U_H(\alpha Q_P + q, \alpha Q_P c + qc_H); \max_{q\geq 0} U_H((1-\alpha)Q_P + q, (1-\alpha)Q_P c + qc_H)\}$$

due to quasi-concavity. Here α corresponds to the share of the pooling quantity purchased in group one, and $1-\alpha$ to the share purchased in group two. Or, whenever the high type purchases pooling contracts in both groups, the possibility of only purchasing one pooling contract and complementing it with the desired quantity at a unit cost c_H is least attractive when the pooling quantity offered in both groups is identical, i.e., $\alpha = \frac{1}{2}$.

□

A.5 Proof of Lemma 3

Proof. Define $\mathcal{T}_H(q), \mathcal{T}_L(q)$ so that $U_H(q, \mathcal{T}_H(q)) = U_H(Q_L + q^\ell, T_L + t^\ell)$ for all q in an open ball around $Q_L + q^\ell$ and $U_L(q, \mathcal{T}_L(q)) = U_L(Q_L, T_L)$ for all q in an open ball around Q_L. Since $U_H(q,t)$ is increasing in q, decreasing in t and continuously differentiable, $\mathcal{T}_H(q), \mathcal{T}_L(q)$ are well-defined and continuously differentiable due to the implicit function theorem, moreover increasing.

And by construction $U_j(q, \mathcal{T}_j(q))$ is constant in q so that differentiation yields $D_1U_j(q,\mathcal{T}_j(q)) + D_2U_j(q,\mathcal{T}_j(q))\mathcal{T}_j'(q) = 0$. Applying the first-order condition to low type conditional efficiency 4 then implies that $D_1U_L(Q_L,T_L) + D_2U_L(Q_L,T_L)c = 0$. Therefore $\mathcal{T}_j'(q) = c$.

Next, observe that for all ϵ in an open ball around zero it must hold that $U_L(Q_L+\epsilon, \mathcal{T}_L(Q_L+\epsilon) - \epsilon^2) > U_L(Q_L,T_L)$. The flatter-curvature assumption 6 thus implies that $U_H(Q_L + q^\ell + \epsilon, \mathcal{T}_L(Q_L+\epsilon) + t^\ell - \epsilon^2) > U_H(Q_L + q^{ell}, T_L + t^\ell) = U_H(Q_L + q^\ell + \epsilon, \mathcal{T}_H(Q_L + q^\ell + \epsilon))$. In effect, the function $\epsilon \mapsto U_H(Q_L + q^\ell + \epsilon, \mathcal{T}_L(Q_L+\epsilon) + t^\ell - \epsilon^2) - U_H(Q_L + q^\ell + \epsilon, \mathcal{T}_H(Q_L + q^\ell + \epsilon))$ attains a local minimum at $\epsilon = 0$. Whence differentiating with respect to ϵ and noting that $\mathcal{T}_L(Q_L) = T_L$ and $\mathcal{T}_H(Q_L + q^\ell) = T_L + t^\ell$ establishes that

$$\begin{aligned}0 =& D_1U_H(Q_L + q^\ell, \mathcal{T}_L(Q_L) + t^\ell) - D_1U_H(Q_L + q^\ell, \mathcal{T}_H(Q_L + q^\ell))\\ &+ D_2U_H(Q_L + q^\ell, \mathcal{T}_L(Q_L) + t^\ell)\mathcal{T}_L'(Q_L) - D_2U_H(Q_L + q^\ell, \mathcal{T}_H(Q_L + q^\ell))\mathcal{T}_H'(Q_L + q^\ell),\end{aligned}$$

that is to say that $\mathcal{T}_H'(Q_L + q^\ell) = \mathcal{T}_L'(Q_L) = c$.

Finally, if (q', t') were such that $U_H(Q_L + q^\ell + q', T_L + t^\ell + t') > U_H(Q_L + q^\ell, T_L + t^\ell) = U_H(Q_L+q^\ell+q', \mathcal{T}_H(Q_L+q^\ell+q'))$, it must be that $\mathcal{T}_H(Q_L+q^\ell+q') > T_L+t^\ell+t' = \mathcal{T}_H(Q_L+q^\ell)+t'$. Taking differences gives

$$\frac{\mathcal{T}_H(Q_L + q^\ell + q') - \mathcal{T}_H(Q_L + q^\ell)}{q'} > \frac{t'}{q'}.$$

It remains observed that the left-hand side is less than c. This holds because $U_H(q, t)$ is quasi-concave so that $\mathcal{T}_H(q)$ is concave. In conclusion, $t' < q' c$, so that $U_H(Q_L+q^\ell+q', T_L+t^\ell+q' c) \leq U_H(Q_L + q^\ell, T_L + t^\ell)$ as claimed. □

A.6 Proof of Proposition 2

Proof. Step 1: The JHG allocation $(Q_L^{JHG}, T_L^{JHG}) = (Q_P, T_P) = (Q_L, T_L), (Q_H^{JHG}, T_H^{JHG}) = (Q_P + q_H, T_P + t_H) = (Q_H, T_L)$ (see Definition 7) always satisfies conditions 1, 2, 3 and 4. By construction, it is single-seller separating, competitively priced, and conditionally efficient. Incentive compatibility follows from conditional efficiency as shown in Section 2. It remains to verify large pooling, i.e., $T_H - T_L + T_P \leq c_H(Q_H - Q_L + Q_P)$. This is always satisfied, for by construction $T_L = T_P$ and $Q_L = Q_P$, and $T_H = T_P + t_H = Q_P c + q_H c_H$.

Step 2: We claim that there exists a regular curve $\gamma(t)$ in an open ball around the JHG allocation (Q_L^{JHG}, T_L^{JHG}) so that $\gamma(0) = (Q_L^{JHG}, T_L^{JHG})$ and $\tau_L(\gamma(t)) = c$ for all $t \in (-1, 1)$.

Case 1: if $cD_{22}^2 U_L(Q_L^{JHG}, T_L^{JHG}) \neq -D_{12}^2 U_L(Q_L^{JHG}, T_L^{JHG})$, the existence of such a curve is a consequence of the implicit function theorem. To see this, define implicitly $\rho(Q)$ so that $\tau_L(Q, \rho(Q)) = c$. Clearly, $\rho(Q_L^{JHG}) = T_L^{JHG}$ by construction of the JHG allocation. We then show that $\rho(Q)$ is well-defined locally around (Q_L^{JHG}, T_L^{JHG}). To apply the implicit function theorem we require that τ is continuously differentiable and that $D_2\tau_L(Q, T) \neq 0$ when evaluated at $(Q, T) = (Q_L^{JHG}, T_L^{JHG})$. $\tau_L(Q, T)$ is continuously differentiable if $U_L(Q, T)$ is twice continuously differentiable. And $D_2\tau_L(Q, T) \neq 0$ at (Q_L^{JHG}, T_L^{JHG}) is equivalent to $D_{22}^2 U_L(Q, T) D_1 U_L(Q, T) \neq D_{12}^2 U_L(Q, T) D_2 U_L(Q, T)$. Then note that by construction of the JHG allocation $c = \tau_L(Q_L^{JHG}, T_L^{JHG}) = -\frac{D_1 U_L(Q_L^{JHG}, T_L^{JHG})}{D_2 U_L(Q_L^{JHG}, T_L^{JHG})}$ So, equivalently, $cD_{22}^2 U_L(Q_L^{JHG}, T_L^{JHG}) \neq -D_{12}^2 U_L(Q_L^{JHG}, T_L^{JHG})$ as we had assumed. And so $\gamma(t)$ is a parametrization of the graph of $\rho(Q)$.

Case 2: if $cD_{12}^2 U_L(Q_L^{JHG}, T_L^{JHG}) \neq -D_{11}^2 U_L(Q_L^{JHG}, T_L^{JHG})$, the existence of such a curve is a consequence of the implicit function theorem. To see this, define implicitly $\sigma(T)$ so that $\tau_L(\sigma(T), Q) = c$ and follow the same steps as before.

Case 3: if both $cD_{22}^2 U_L(Q_L^{JHG}, T_L^{JHG}) = -D_{12}^2 U_L(Q_L^{JHG}, T_L^{JHG})$ and $cD_{12}^2 U_L(Q_L^{JHG}, T_L^{JHG}) = -D_{11}^2 U_L(Q_L^{JHG}, T_L^{JHG})$, then $D_1\tau_L(Q_L^{JHG}, T_L^{JHG}) = D_2\tau_L(Q_L^{JHG}, T_L^{JHG}) = 0$. This means that there exists an open ball around (Q_L^{JHG}, T_L^{JHG}) so that $\tau_L(Q, T) = c$ for all (Q, T) inside.

Step 3: We observe that $Q_P \mapsto \max_q U_H(Q_P + q, Q_P\, c + q\, c_H)$ is continuous. This is a consequence of Berge's maximum theorem: $U_H(Q,T)$ is continuous, moreover $\arg\max_{Q\geq 0} U_H(Q+q, Qc+qc_H) \leq \max\{\arg\max_q U_H(q,qc), \arg\max_q U_H(q,qc_H)\}$ due to quasi-concavity of $U_H(Q,T)$.

Step 4: We show that $U_H(Q_H^{JHG}, T_H^{JHG}) > U_H(Q_L^{JHG}, T_L^{JHG})$. Or, suppose by contradiction that $U_H(Q_H^{JHG}, T_H^{JHG}) = U_H(Q_L^{JHG}, T_L^{JHG})$. Whence, due to strict quasi-concavity, it holds that $U_H(\lambda Q_L^{JHG} + (1-\lambda)Q_L^{JHG}, \lambda T_L^{JHG} + (1-\lambda)T_L^{JHG}) > \min\{U_H(Q_H^{JHG}, T_H^{JHG}), U_H(Q_L^{JHG}, T_L^{JHG})\}$ for all $\lambda \in (0,1)$. Yet, by construction of the JHG allocation, $U_H(Q_H^{JHG}, T_H^{JHG}) = \max_q U_H(Q_L^{JHG}+q, T_L^{JHG}+qc_H)$ and this establishes the desired contradiction.

Conclusion: Fix $\epsilon > 0$ so that $U_H(Q_H^{JHG}, T_H^{JHG}) > U_H(Q_L^{JHG}, T_L^{JHG}) + \epsilon$. Such an ϵ exists due to Step 4. Next, due to step 3, there exists δ_H so that $\max_q U_H(Q_P+q, Q_Pc+qc_H) > \max_q U_H(Q_L^{JHG}+q, Q_L^{JHG}c+qc_H) - \frac{\epsilon}{2}$ for all $Q_P : |Q_P - Q_L^{JHG}| < \delta_1$. And due to continuity of U_H there exists δ_L so that $U_H(Q_L^{JHG}, Q_L^{JHG}) + \frac{\epsilon}{2} > U_H(Q_P+q_L, Q_Pc+q_Lc_x)$ for all (Q_P, q_L) so that $0 < q_L < \delta_2$ and $|Q_P - Q_L^{JHG}| < \delta_2$. Finally, by construction of γ, i.e., Step 2, for all $\delta > 0$ there exists a low type allocation $(Q_P+q_L, Q_Pc+q_Lc_x)$ in the image of γ so that $|Q_P - Q_L^{JHG}| < \delta$ and $0 < q_L < \delta$. In particular, we can choose such (Q_P, q_L) for $\delta = \min\{\delta_1, \delta_2\}$. Then pick $q_H \in \arg\max_q U_H(Q_P+q, Q_Pc+qc_H)$.

Thus constructed aggregate trades $(Q_P, Q_Pc), (q_L, q_Lc_x)$ and (q_H, q_Hc_H) are incentive compatible, competitively priced and conditionally efficient. For δ sufficiently small they moreover satisfy large pooling, because the JHG allocation satisfies large pooling. □

B Existence: missing proofs

B.1 Proof of Theorem 2

We distinguish between two possible classes of equilibrium. In one set of equilibrium candidates, sellers in both groups actively trade pooling contracts. Due to its inefficiency—there exist additional incentive compatible separating contracts that are profitable—Footnote 23 and our subsequent Proposition 3 argue that this case is pathological. In another set of equilibrium candidates, actively traded contracts in group $\mathcal{K}_1$ are pooling, and actively traded contracts in group $\mathcal{K}_2$ are separating. We will consider this case first.

"Pooling + Separating" Equilibria

We here construct an equilibrium where pooling contracts are only traded with sellers in group $\mathcal{K}_1$. Any active aggregate trades (Q_P, T_P) and $(q_L, t_L), (q_H, t_H)$ that satisfy conditions 1, 2, 3 and 4 with the restriction that the separating quantity q_H satisfies $q_H \in \arg\max_{q\geq 0} U_H(Q_P + q, T_P + qc_H)$ are admissible. This includes the possibility that $q_H = 0$. Sellers in group $\mathcal{K}_1$ offer contract (Q_P, T_P) or latent contracts, sellers in group $\mathcal{K}_2$ offer contracts (q_L, t_L), (q_H, t_H) or latent contracts. Latent contracts are as follows: The principal latent contract (q^ℓ, t^ℓ) is offered by inactive sellers in both groups $\mathcal{K}_1$ and $\mathcal{K}_2$, with the principal contract as defined in Lemma

3. For reasons that shall become apparent shortly, we also introduce a larger class of derivative latent contracts. Denote $(q^{n\ell}, t^{n\ell}) = (nq^{\ell}, nt^{\ell})$. And let $N \in \mathbb{N}$ be the smallest N such that $q^{N\ell} > Q_H$. The set of latent contracts offered in group $\mathcal{K}_1$ is

$$\left\{(q^{n\ell}, t^{n\ell}), (Q_P + q^{n\ell}, T_P + t^{n\ell}) : 1 \leq n \leq N\right\}.$$

The set of latent contracts offered in group $\mathcal{K}_2$ is

$$\left\{(q^{n\ell}, t^{n\ell}), (q_L + q^{n\ell}, t_L + t^{n\ell}), (q_H + q^{n\ell}, t_H + t^{n\ell}) : 1 \leq n \leq N\right\}.$$

Figure 7 summarizes the entire set of contracts offered.

	Subgroup $\mathcal{K}_1$	Subgroup $\mathcal{K}_2$
On-path sellers (actively trade)	(Q_P, T_P)	(q_L, t_L), (q_H, t_H)
Off-path sellers (latent contracts)	$(q^{n\ell}, t^{n\ell})$ $(Q_P + q^{n\ell}, T_P + t^{n\ell})$	$(q^{n\ell}, t^{n\ell})$ $(q_L + q^{n\ell}, t_L + t^{n\ell})$ $(q_H + q^{n\ell}, t_H + t^{n\ell})$

Figure 7: The on-path and off-path contracts in equilibrium

The proof of Theorem 2 under partial pooling is the consequence of three lemmata, each of which maintains identical assumptions and conditions as the main theorem. We first ensure that the latent contracts are not destabilizing, i.e., no agent has an incentive to purchase latent contracts.

Lemma 6. *No buyer is better off when actively trading at least one latent contract.*

We then show that for every cream-skimming deviation there exists a latent contract that blocks it.

Lemma 7. *No seller can offer a profitable deviating contract that only attracts low type buyers.*

We finally show that pivoting on latent contracts to attract high type buyers cannot be profitable. This concludes the proof of Theorem 2, for the set of relevant one-shot seller deviations, comprises (i) undercutting, (ii) pivoting on the on-path contracts to attract high types, (iii) efficiency-improving deviations, (iv) pivoting on the on-path contracts to attract low types, (v) pivoting on the off-path latent contracts to attract low types, (vi) pivoting on the off-path contracts to attract high types. (i) Undercutting was not profitable due to Condition 2, (ii) pivoting on the on-path contracts to attract high types was not profitable due to Condition 3, (iii) efficiency-improving deviations do not exist due to Condition 4, and so called cream-skimming deviations (iv) and (v) were not profitable due to the preceding lemma.

Lemma 8. *No seller can offer a profitable deviating contract that only attracts high type buyers.*

"Pooling + Pooling" Equilibria

We now construct an equilibrium where sellers from both groups actively trade pooling contracts. Any active aggregate trades (Q_P, T_P) and $(q_L, t_L), (q_H, t_H)$ that satisfy conditions 1, 2, 3 and 4 with the restriction that the separating quantities q_H, q_L are zero are admissible. In the aggregate, both types purchase the conditionally efficient quantity for the low type. For comparison, this is the low type's JHG allocation. To keep the analysis disjoint from the preceding, further assume that $0 \notin \arg\max_{q\geq 0} U_H(Q_P + q, T_P + qc_H)$, i.e., that the active aggregate trades are payoff-dominated by a more efficient equilibrium that occurs when there is "Pooling + Separation".

To construct the equilibrium "pooling+pooling" allocation, sellers in each group propose contracts $(\frac{1}{2}Q_P, \frac{1}{2}T_P) = \frac{1}{2}(Q_L^{JHG}, T_L^{JHG})$ or latent contracts. The principal latent contract (q^ℓ, t^ℓ) is still as given by Lemma 3. As before, denote $(q^{n\ell}, t^{n\ell}) = (nq^\ell, nt^\ell)$. And let $N \in \mathbb{N}$ be the smallest N such that $q^{N\ell} > Q_H$. The set of latent contracts offered in groups $\mathcal{K}_1$ and $\mathcal{K}_2$ is

$$\big\{(q^{n\ell}, t^{n\ell}), (\frac{1}{2}Q_P + q^{n\ell}, \frac{1}{2}T_P + t^{n\ell}) : 1 \leq n \leq N\big\}.$$

The set of contracts that sustain the equilibrium is depicted in Figure 8.

	Subgroup $\mathcal{K}_1$	Subgroup $\mathcal{K}_2$
On-path sellers (actively trade)	$(\frac{1}{2}Q_P, \frac{1}{2}T_P)$	$(\frac{1}{2}Q_P, \frac{1}{2}T_P)$
Off-path sellers (latent contracts)	$(q^{n\ell}, t^{n\ell})$ $(\frac{1}{2}Q_P + q^{n\ell}, \frac{1}{2}T_P + t^{n\ell})$	$(q^{n\ell}, t^{n\ell})$ $(\frac{1}{2}Q_P + q^{n\ell}, \frac{1}{2}T_P + t^{n\ell})$

Figure 8: The equilibrium contracts of "pooling +pooling "

To see that this is an equilibrium in which both types actively trade pooling contracts in both groups, one must consider the same set of deviations as in the analysis under "Pooling + Separation". Inspection of the proofs of Lemma 6 and Lemma 7 reveals that both results hold for identical reasons as before. It remains to verify that no profitable deviation only attracts high type buyers, i.e., Lemma, 8. In addition to the deviations considered in the preceding, buyers in either group can now pivot on $(\frac{1}{2}Q_P, \frac{1}{2}T_P)$ to attract type H buyers. Doing so is not feasible due to Condition 4: $\max_{q\geq 0} U_H(1/2Q_P + q, 1/2Q_Pc + qc_H) \leq U_H(Q_P, Q_Pc)$ ensures that no profitable deviation can attract high type buyers by pivoting on $(\frac{1}{2}Q_P, \frac{1}{2}T_P)$. In terms of pivoting on the latent contract, we use the same argument as in Lemma 8: $U_H(Q_H, T_H) =$

$\max_q U_H(q_L + q^\ell + q, t_L + t^\ell + q\,c) \geq \max_q U_H(q^\ell + q, t^\ell + q\,c)$, so that any pivoting contract that attracts the high type must have a unit price lower than c.

Proof of Lemma 6

Proof. First, it follows from Condition 4 and $t > q^\ell c$ that

$$U_L(Q_L, T_L) = \max_q U_L(Q_L + q, T_L + q\,c) \geq \max_q U_L(q_L + q, t_L + q\,c).$$

This is weakly greater than the utility when purchasing at least one latent contract: Condition 2 asserts that buyers can at most once purchase contract (q_L, t_L) (possibly part of a latent contract $(q_L + q^{n\ell}, t_L + t^{n\ell})$; and due to Condition 2 and $t > q^\ell c$ any other contract offered has unit price weakly greater than c.

Second, observe that by construction and Lemma 3

$$\begin{aligned}
&U_H(Q_H, T_H) = U_H(Q_L + q^\ell, T_L + t^\ell) \\
&\geq \max_q U_H(Q_L + q^\ell + q, T_L + t^\ell + q\,c) = \max_q U_H(q_L + q, t_L + q\,c).
\end{aligned}$$

Thereby derived utility is weakly greater than purchasing any latent contract for identical reasons. □

Proof of Lemma 7

Proof. First, consider a large cream-skimming deviation (q', t') in either group $\mathcal{K}_1$ or $\mathcal{K}_2$ so that $U_L(q', t') > U_L(Q_L, T_L)$. Then the flatter-curvature assumption implies that $U_H(q' + q^\ell, t' + t^\ell) > U_H(Q_L + q^\ell, T_L + t^\ell) = U_H(Q_H, T_H)$. So, the latent contract (q^ℓ, t^ℓ) blocks this deviation.

Analogously, consider a small cream-skimming deviation (q', t') in either group $\mathcal{K}_1$ or $\mathcal{K}_2$ that pivots on the latent contract $(q^{(n-1)\ell}, t^{(n-1)\ell})$ where $n > 1$. Or, $U_L(q' + q^{(n-1)\ell}, t' + t^{(n-1)\ell}) > U_L(Q_L, T_L)$. Then the flatter-curvature assumption implies that $U_H(q' + q^{(n-1)\ell} + q^\ell, t' + t^{(n-1)\ell} + t^\ell) > U_H(Q_L + q^\ell, T_L + t^\ell) = U_H(Q_H, T_H)$. So, the latent contract $(q^{n\ell}, t^{n\ell})$ blocks this deviation.

Finally, we show that there does not exist a small cream-skimming deviation (q', t') in either group $\mathcal{K}_1$ or $\mathcal{K}_2$ that pivots on the latent contract $(q^{N\ell}, t^{N\ell})$ and thereby exclusively attracts low type buyers. If it did, due to incentive compatibility, $U_L(q' + q^{N\ell}, t' + t^{N\ell}) > U_L(Q_L, T_L) \geq U_L(Q_H, T_H)$. Yet since $q^{N\ell} > Q_H$, also $U_H(q' + q^{N\ell}, t' + t^{N\ell}) > U_H(Q_H, T_H)$ due to single-crossing. Analogously, one cannot exclusively attract low type buyers by pivoting on the latent contracts $(q_L + q^{N\ell}, t_L + t^{N\ell})$, $(q_H + q^{N\ell}, t_H + t^{N\ell})$ and $(Q_P + q^{N\ell}, T_P + t^{N\ell})$.

Second, consider a cream-skimming deviation (q', t') in group $\mathcal{K}_1$ that pivots on a contract $(q_L + q^{(n-1)\ell}, t_L + t^{(n-1\ell})$ in group $\mathcal{K}_2$ where $1 \leq n \leq N$. Or, $U_L(q' + q_L + q^{(n-1)\ell}, t' + t_L + t^{(n-1)\ell}) > U_L(Q_L, T_L)$. Then the flatter-curvature assumption implies that $U_H(q' + q_L + q^{(n-1)\ell} + q^\ell, t' + t_L + t^{(n-1)\ell} + t^\ell) > U_H(Q_L + q^\ell, T_L + t^\ell) = U_H(Q_H, T_H)$. So, the contract $(q_L + q^{n\ell}, t_L + t^{n\ell})$

in group $\mathcal{K}_2$ blocks this deviation. (The same argument applies for cream-skimming deviations that pivot on a contract $(q_H + q^{(n-1)\ell}, t_H + t^{(n-1\ell)})$ in the group $\mathcal{K}_2$ where $1 \leq n \leq N$.).

Third, consider a cream-skimming deviation (q', t') in group $\mathcal{K}_2$ that pivots on a contract $(Q_P + q^{(n-1)\ell}, T_P + t^{(n-1\ell)})$ in group $\mathcal{K}_1$ where $1 \leq n \leq N$. Or, $U_L(q' + Q_P + q^{(n-1)\ell}, t' + T_P + t^{(n-1)\ell}) > U_L(Q_L, T_L)$. Then the flatter-curvature assumption implies that $U_H(q' + Q_P + q^{(n-1)\ell} + q^\ell, t' + T_P + t^{(n-1)\ell} + t^\ell) > U_H(Q_L + q^\ell, T_L + t^\ell) = U_H(Q_H, T_H)$. So, the contract $(Q_P + q^{n\ell}, T_P + t^{n\ell})$ in group $\mathcal{K}_1$ blocks this deviation.

$\square$

Proof of Lemma 8

Proof. First observe that a deviation (q', t') exclusively targeting high type buyers must satisfy $q'c_H < t'$ in order to be profitable. Then recall that $q^\ell c < t^\ell$. It follows that if a deviating contract (q', t') pivots on a latent contract that is different from $(q_L + q^{n\ell}, t_L + t^{n\ell})$, then the total unit cost of the quantity traded following the deviation must exceed c (due to competitive pricing, Condition 2). But then (q', t') traded in conjunction with the latent contract can impossibly be advantageous to high type buyers, for on-path utility satisfies $U_H(Q_H, T_H) = \max_q U_H(q_L + q^\ell + q, t_L + t^\ell + q\,c) \geq \max_q U_H(q^\ell + q, t^\ell + q\,c)$ (due to Lemma 3, moreover $t_L \leq q_L\, c$ due to competitive pricing in Condition 2).

Thus consider a deviation (q', t') in group $\mathcal{K}_1$ that pivots on a contract of the form $(q_L + q^{n\ell}, t_L + t^{n\ell})$. If $n = 0$ this deviation is not profitable and at the same time advantageous to high type buyers due to Condition 3. Thus suppose that $1 \leq n$. If high type buyers were better off by trading (q', t'), it follows (due to Lemma 3) that

$$\begin{aligned} U_H(q' + q_L + q^{n\ell}, t' + t_L + t^{n\ell}) > U_H(Q_H, T_H) &= \max_q U_H(q_L + q^\ell + q, t_L + t^\ell + q\,c) \\ &\geq U_H(q' + q^{n\ell} + q_L, t_L + t^\ell + (q' + q^{n\ell} + q_L - q_L - q^\ell)c). \end{aligned}$$

It must therefore hold that

$$t' + t^{n\ell} + t_L < t_L + t^\ell + (q' + q^{n\ell} - q^\ell)c \qquad \Leftrightarrow \qquad t' < q^{(n-1)\ell}c - t^{(n-1)\ell} + q'c.$$

Since $q^\ell c < t^\ell$, this implies that $t' < q'c$ and so (q', t') must be lossmaking. $\square$

C Positive Implications: missing proofs

C.1 Proof of Proposition 3

Proof of Proposition 3, claim 1. Denote (Q_P, T_P) the equilibrium pooling allocation. If there exists a non-zero quantity $Q_H - Q_P = \arg\max_{q \geq 0} U_H(Q_P + q, T_P + qc_H)$, an aftermarket seller can propose the contract $(Q_H - Q_P, (Q_H - Q_P)c_H + \epsilon)$. For $\epsilon > 0$ sufficiently low, this contract attracts type H and is profitable. Whence (Q_P, T_P) is not a serendipitous-aftermarket-proof.

If to the contrary pooling is conditionally efficient, no profitable pooling nor a separating contract targeting high type buyers will generate the desired demand. Cream-skimming deviations uniquely targeting low type buyers, meanwhile, will be infeasible due to single-crossing. $\square$

Proof of Proposition 3, claim 2. To prove the claim it suffices to consider aftermarket cream-skimming deviations that uniquely target the low type. For conditional efficiency ensures that $U_H(Q_H, T_H) = \max_{q\geq 0} U_H(Q_H + q, T_H + qc_H)$ and $U_L(Q_L, T_L) = \max_{q\geq 0} U_L(Q_L + q, T_L + qc)$, so that neither a pooling deviation nor a deviation uniquely targeting high type buyers can be both incentive compatible and profitable. The proof that there do not exist profitable cream-skimming deviations relies on the following arguments:

Step 1: For any three contracts $(Q_1, T_1), (Q_2, T_2)$ and (q', t') so that $Q_2 > Q_1$ satisfying $U_H(Q_1, T_1) = U_H(Q_2, T_2)$ and $U_H(Q_2, T_2) < U_H(Q_2 + q', T_2 + t')$: it must hold that $\frac{T_2 - T_1}{Q_2 - Q_1} > \frac{T_2 - T_1 + t'}{Q_2 - Q_1 + q'}$.

This is an immediate consequence of strict quasi-concavity. Otherwise, there exists a contract $(Q_3, T_3) : Q_3 > Q_2$ with $\frac{T_3 - T_2}{Q_3 - Q_2} = \frac{T_2 - T_1}{Q_2 - Q_1}$ and $U_H(Q_3, T_3) \geq U_H(Q_2, T_2)$. And, due to strict quasi-concavity, it must hold that $U_H(\lambda Q_1 + (1 - \lambda)Q_3, \lambda T_1 + (1 - \lambda)T_3) > \min\{U_H(Q_1, T_1), U_H(Q_3, T_3)\} = U_H(Q_1, T_1) = U_H(Q_2, T_2)$. Yet when setting $\hat{\lambda} = \frac{Q_3 - Q_2}{Q_3 - Q_1}$ (which is equal to $\frac{T_3 - T_2}{T_3 - T_1}$ by construction), it thereby follows that $U_H(Q_2, T_2) = U_H(\hat{\lambda} Q_1 + (1 - \hat{\lambda})Q_3, \hat{\lambda} T_1 + (1 - \hat{\lambda})T_3) > U_H(Q_2, T_2)$, which establishes the desired contradiction.

Step 2: For any three contracts $(Q_1, T_1), (Q_2, T_2)$ and (q', t') so that $Q_2 > Q_1$ satisfying $U_H(Q_1, T_1) = U_H(Q_2, T_2)$ and $U_H(Q_2, T_2) < U_H(Q_2 + q', T_2 + t')$: it must hold that $U_H(Q_1, T_1) < U_H(Q_1 + q', T_1 + t')$.

This is also a consequence of strict quasi-concavity: $U_H(\lambda Q_1 + (1 - \lambda)(Q_2 + q'), \lambda T_1 + (1 - \lambda)(T_2 + t')) > \min\{U_H(Q_1, T_1), U_H(Q_2 + q', T_2 + t')\} = U_H(Q_1, T_1)$ for all $\lambda \in (0, 1)$. Then set $\hat{\lambda} = \frac{Q_2 - Q_1}{Q_2 - Q_1 + q'}$ so that $\hat{\lambda} Q_1 + (1 - \hat{\lambda})(Q_2 + q') = Q_1 + q'$. In effect, $\hat{\lambda} T_1 + (1 - \hat{\lambda})(T_2 + t') = T_1 + t' + (T_2 - T_1) + \hat{\lambda}(t' - T_2 + T_1) = T_1 + t' + (T_2 - T_1) - (Q_2 - Q_1)\frac{T_2 - T_1 + t'}{Q_2 - Q_1 + q'} > T_1 + t' + (T_2 - T_1) - (Q_2 - Q_1)\frac{T_2 - T_1}{Q_2 - Q_1} = T_1 + t'$ where the inequality is due to the step 1. Since utility is decreasing in transfers, this proves that $U_H(Q_1 + q', T_1 + t') > U_H(\hat{\lambda}(Q_1, T_1) + (1 - \hat{\lambda})(Q_2 + q', T_2 + t')) > U_H(Q_1, T_1)$.

Step 3: For any "Pooling + Separating" allocation $(Q_L, T_L), (Q_H, T_H)$ satisfying conditions 1, 2, 3 and 4, the unique principal latent contract (q^ℓ, t^ℓ) (see Assumption 6 and Lemma 3) satisfies $Q_L + q^\ell > Q_H$.

To see this, recall that $U_H(Q_L + q^\ell, T_L + t^\ell) = U_H(Q_H, T_H)$. Then first, due to strict quasi-concavity for $\lambda = 1/2$, it holds that $U_H(Q_H + 1/2(Q_L + q^\ell - Q_H), T_H + 1/2(T_L + t^\ell - T_H)) > U_H(Q_H, T_H) = \max_q U_H(Q_H + q, T_H + q\ c_H) \geq U_H(Q_H + 1/2(Q_L + q^\ell - Q_H), T_H + 1/2(Q_L + q^\ell - Q_H)c_H)$ where the equality is due to conditional efficiency. It follows that $T_L + t^\ell - T_H < c_H(Q_L + q^\ell - Q_H)$. Second, due to strict quasi-concavity for $\lambda = 1/2$ it similarly holds that $U_H(Q_L+q^\ell+1/2(Q_H-Q_L-q^\ell),T_L+t^\ell+1/2(T_H-T_L-t^\ell)) > U_H(Q_L+q^\ell,T_L+t^\ell) = \max_q U_H(Q_L+q^\ell+q,T_L+t^\ell+q\,c) \geq U_H(Q_L+q^\ell+1/2(Q_H-Q_L-q^\ell),T_L+t^\ell+1/2(Q_H-Q_L-q^\ell)c)$ where the equality is due to Lemma 3. It follows that $T_H - T_L - t^\ell < c(Q_H - Q_L - q^\ell)$. To conclude, combining the conclusions of the first and the second argument establishes that $(Q_L + q^\ell - Q_H)c_H > (Q_L + q^\ell - Q_H)c$. This can impossibly hold if $Q_L + q^\ell \leq Q_H$.

Step 4: We now prove claim (ii): Consider an aftermarket cream-skimming deviation (q', t') so that $U_L(Q_L + q', T_L + t') > U_L(Q_L, T_L)$. Then Assumption 6 ensures that $U_H(Q_L + q^\ell + q', T_L+t^\ell+t', U_H(Q_L+q^\ell, T_L+t^\ell) = U_H(Q_H, T_H)$. Then set $(Q_1, T_1) = (Q_H, T_H)$ and $(Q_2, T_2) = (Q_L + q^\ell, T_L + t^\ell)$. Step 3 implies that $Q_2 > Q_1$. Then step 2 implies that $U_H(Q_1, T_1) < U_H(Q_1 + q', T_1 + t')$, as claimed. □

Proof of Proposition 3 claim 3 . To begin with, we show that there do not exist cream-skimming aftermarket deviations uniquely attracting low type buyers. Indeed, due to Assumption 6 there exists a (unique due to Lemma 3) contract (q^ℓ, t^ℓ) satisfying $U_H(Q_L+q^\ell, T_L+t^\ell) = U_H(Q_H, T_H)$ so that $U_L(Q_L+q, T_L+t) \geq U_L(Q_L, T_L)$ implies $U_H(Q_L+q^\ell+q, T_L+t^\ell+t) > U_H(Q_L+q^\ell, T_L+t^\ell)$ for all (q, t) (although strictly speaking we only require this for positive (q, t)). And by Steps 2 and 3 from the proof of Claim 2, $Q_L + q^\ell > Q_H$, and $U_H(Q_L + q^\ell + q, T_L + t^\ell + t) > U_H(Q_L + q^\ell, T_L + t^\ell)$ implies that $U_H(Q_H + q, T_H + t) > U_H(Q_H, T_H)$. It follows that no cream-skimming deviation that uniquely attracts low type buyers exists.

We then prove the "if" claim. If $0 = \arg\max_{q\geq 0} U_L(Q_L + q, T_L + qc)$, no profitable pooling contract can attract low type buyers. And due to the conditional efficiency of (Q_H^{RS}, T_H^{RS}), no separating contract targeting high type buyers can generate positive profit.

We then prove the "only if" claim: If to the contrary $Q_P \in \arg\max_{q\geq 0} U_L(Q_L + q, T_L + qc)$, there exists $\epsilon > 0$ so that $U_L(Q_L + Q_P, T_L + Q_P c + \epsilon) > U_L(Q_L, T_L)$. In effect, this contract surely attracts type L buyers and is already profitable and conditional on attracting both buyer types. □

D Welfare: missing proofs

We here prove our main welfare result within a more general environment where group 2 sellers can post menus:

- Stage 1: Each seller $k \in \mathcal{K}$ proposes a single contract (q_P^k, t_P^k) so that $q_P^k \geq \underline{Q}_P$ in group 1 or a menu of contracts (q_L^k, t_L^k) and (q_H^k, t_H^k) in group 2. Further denote $(q_\vartheta^0, t_\vartheta^0) = (0, 0)$ with $\vartheta \in \{L, H\}$ the group 2 null contract.

- Stage 2: Each buyer learns her type θ, and decides to purchase insurance or not. If so, she selects one group 1 seller k_1 in $\mathcal{K}$ and one group 2 seller k_2 in $\mathcal{K} \cup \{0\}$, makes a report $\vartheta \in \{L, H\}$ and derives utility $U_\theta(q_P^{k_1} + q_\vartheta^{k_2}, t_P^{k_1} + t_\vartheta^{k_2})$. Otherwise that buyer's utility is $U_\theta(0, 0)$.

We emphasize that any buyer that wishes to purchase some coverage must purchase a contract in group 1.

Proof of Theorem 3. We decentralize the equilibrium allocation without relying on latent contracts: all contracts offered in equilibrium are active. Only two sellers per group are required. At least two group 1 sellers offer $(Q_P, Q_P c)$ and at least two group 2 sellers offer $(q_L, q_L c_L)$ and $(q_H, q_H c_H)$. Since the aggregate allocation is incentive compatible and satisfies participation constraints, on path the low (high) risk type is better off purchasing than not purchasing even if this includes mandatory group 1 coverage $(Q_P, Q_P c)$, and upon purchasing weakly prefers $(q_L, q_L c_L)$ ($(q_H, q_H c_H)$ respectively) over purchasing $(q_H, q_H c_H)$ ($(q_L, q_L c_L)$ respectively) in group 2.

We then show that there do not exist any profitable group 1 seller deviations. Our proof does not consider group 2 deviations because those are encompassed by the familiar arguments pertaining to the exclusive Rothschild and Stiglitz (1976) model.

Suppose that some seller were to offer the deviating contract (Q'_P, T'_P) in group 1. If so, this deviation must necessarily satisfy the minimum quantity requirement, i.e., $Q'_P \geq Q_P$. To be profitable, such a contract can never uniquely attract high-risk buyers because all the gains from trading with high-risk buyers are already exhausted by the candidate allocation, i.e., it holds that $q_H \in \arg\max_{q \geq 0} U_H(Q_P + q, Q_P c + q c_H)$. In effect, any group 1 deviation must also attract low-risk buyers in which case low-risk buyers may also adjust their group 2 demand, denoted (q'_L, t'_L). We distinguish between three cases.

Case 1: high-risk buyers' incentive constraint is binding. Three possibilities arise. (i) If $Q'_P + q'_L > Q_P + q_L$, then the allocation $(Q'_P + q'_L, T'_P + t'_L)$ also attracts high-risk buyers and is therefore not profitable. This is an immediate consequence of single-crossing, i.e., Assumption 3 whereby $U_L(Q'_P + q'_L, T'_P + t'_L) > U_L(Q_P + q_L, T_P + t_L)$ implies $U_H(Q'_P + q'_L, T'_P + t'_L) > U_H(Q_P + q_L, T_P + t_L) = U_H(Q_H, T_H)$. (ii) If $Q'_P + q'_L = Q_P + q_L$, allocation $(Q'_P + q'_L, T'_P + t'_L)$ attracting low-risk buyers implies that $T'_P + t'_L < T_P + t_L$, so that once more $U_H(Q'_P + q'_L, T'_P + t'_L) > U_H(Q_P + q_L, T_P + t_L) = U_H(Q_H, T_H)$ due to the binding incentive constraint.

(iii) If $Q'_P + q'_L < Q_P + q_L$, proceed in two steps. We first show that H trading (Q'_P, T'_P) in group 1 and (q_L, t_L) in group 2 dominates H's initial allocation so that the group 1 deviation also attract high-risk buyers:

$$U_H(Q'_P + q_L, T'_P + t_L) > U_H(Q_H, T_H).$$

To see this, note that there exists $\alpha \in (0,1)$ so that $Q'_P + q_L = Q_P + \alpha q_L + (1-\alpha) q_H$. That is, a convex combination of L's and H's on-path coverage corresponds to the level of coverage available following the deviation. Since, by Assumption 1 and H's binding incentive constraint the convex combination of allocations is desirable to high-risk buyers, i.e., $U_H(Q_P + \alpha q_L + (1-\alpha) q_H) > U_H(Q_L, T_L) = U_H(Q_H, T_H)$, it suffices to show that the newly available allocation, while replicating the desirable coverage now becomes available at an even cheaper aggregate transfer: $T'_P + t_L < cQ_P + \alpha q_L c_L + (1-\alpha) q_H c_H$. This holds because

$$T'_L + t_L < cQ'_P + c_L q_L = c(Q_P + (1-\alpha)(q_H - q_L)) + c_L q_L$$
$$cQ_P + (1-\alpha)q_H c + q_L c_L - (1-\alpha) q_L c = cQ_P + \alpha q_L c_L + (1-\alpha) q_H \underbrace{c}_{<c_H} \underbrace{-(1-\alpha)q_L(c-c_L)}_{<0}.$$

We secondly show that the seller's loss on (Q'_P, T'_P) upon both types purchasing group 1 deviating contract (Q'_P, T'_P) cannot be offset by a possible profit on contract (q'_L, t'_L) (which exclusively attracts low-risk buyers).

To see this, employ an accounting argument: To begin with, note that there exists $\alpha \in (0,1)$ so that $Q_P + \alpha q_L = Q'_P + q'_L$. And, due to incentive efficiency, it holds that $U_L(Q_P + \alpha q_L, Q_P c + \alpha q_L c_L) < U_L(Q_P + q_L, Q_P c + q_L c_L)$. Yet, by assumption, low-risk types are attracted by the deviation so that $U_L(Q_P + q_L, Q_P c + q_L c_L) < U_L(Q'_P + q'_L, T'_P + t'_L)$. Then it must hold that $T'_P + t'_L < Q_P c + \alpha q_L c_L$. Furthermore, note that since $T'_P < Q'_P c$, $Q'_P > Q_P$ and $T_P = Q_P c$ it holds that $T'_P - Q'_P c < T_P - Q_P c$. Or, $T'_P - Q'c = Q_P(\frac{T'_P}{Q'_P} - c) + (Q'_P - Q_P)(\frac{T'_P}{Q'_P} - c) < Q_P(\frac{T_P}{Q_P} - c)$ as claimed. In effect, the seller's profit of the deviation of offering (Q'_P, T'_P) in group 1 and (q'_L, t'_L) in group 2 is

$$\begin{aligned}
T'_P - Q'_P c + m_L(t'_L - q'_L c_L) &= \big(T'_P - Q'_P c_L + t'_L - q'_L c_L\big) m_L + \big(T'_P - Q'_P c_H\big) m_H \\
&< \big(Q_P c + \alpha q_L c_L - Q'_P c_L - q'_L c_L\big) m_L + \big(T'_P - Q'_P c_H\big) m_H = Q_p(c - c_L) m_L + \big(T'_P - Q'_P c_H\big) m_H \\
&< Q_p(c - c_L) m_L + \big(Q_P c - Q_P c_H\big) m_H = 0.
\end{aligned}$$

Or, the deviation is not profitable.[32]

Case 2: low-risk buyers' incentive constraint is binding. If so $q_L = q_H = 0$, and the high-risk buyer's incentive constraint is binding as well. Then refer to Case 1.

Case 3: Both incentive constraints are slack. Note that since incentive constraints are slack, it holds that

$$U_L(Q_L, T_L) = \max_{(Q,T): Qc_L - T \le Q_P(c - c_L), Q \ge Q_P} U_L(Q, T)$$

(and analogously $U_H(Q_H, T_H) = \max_{(Q,T): Qc_H - T \le Q_P(c - c_H), Q \ge Q_P} U_H(Q, T)$). It follows that any deviation meant to attract L (H) risk buyers will earn a lower per L (H) type profit than the candidate allocation. Since on path each contract traded breaks even, for any deviation to be profitable it must now uniquely attract L risk buyers in group 1. Thus suppose that (Q'_P, T'_P) attracts low-risk buyers, but not high-risk buyers. If $Q'_P \ge Q_H$, then single-crossing, i.e., Assumption 3 readily implies that H risk buyers purchase the deviating allocation if L risk buyers do. Thus suppose that $Q'_P < Q_H$. If H risk buyers choose not to buy (Q'_P, T'_P), then they must prefer their allocation (Q_H, T_H) over purchasing (Q'_P, T'_P) in group 1 joint with some limit

[32] Observe that we did not require sellers to post limit orders to rule out the existence of profitable group 1 deviations.

order $(\alpha q_H, \alpha q_H c_H)$ in group 2:

$$U_H(Q_P + q_H, Q_P c + q_H c_H) \geq \max_{q\in[0,q_H]} U_H(Q'_P + q, T'_P + qc_L) \geq U_H(Q_H, T'_P + (Q_H - Q'_P)c_H).$$

In effect, $Q_P c + q_H c_H \leq T'_P + (Q_H - Q'_P)c_H$, or, equivalently, $T'_P \geq Q_P c + (Q'_P - Q_P)c_H$. We then show that L risk buyers do not find group 1 contract (Q'_P, T'_P) desirable either. Ensuing utility if they did is

$$\max_{\alpha\in[0,1]} U_L(Q'_P + \alpha q_L, T'_P + \alpha q_L c_L) \leq \max_{\alpha\in[0,1]} U_L(Q'_P + \alpha q_L, Q_P c + (Q'_P - Q_P)c_H + \alpha q_L c_L).$$

This, however, is dominated by the on-path allocation (recall that $Q'_P \geq Q_P$):

$$\begin{aligned} U_L(Q_L, T_L) &= \max_{q} U_L(Q_P + q, Q_P c + qc_L) \geq \max_{q\in[0,q_L]} U_L(Q'_P + q, Q_P c + (Q'_P - Q_P)c_H + qc_L) \\ &= \max_{\alpha\in[0,1]} U_L(Q'_P + \alpha q_L, Q_P c + (Q'_P - Q_P)c_H + \alpha q_L c_L) \end{aligned}$$

and strictly so if $Q'_P > Q_P$. □

Proof of Lemma 5. Part 1: We first show that the second-best decentralizing pooling quantity Q_P is non-decreasing in the ratio $\frac{\lambda_H}{\lambda_L}$. Quasi-linearity is not required for this to hold. Without loss normalize $\lambda_L = 1$ and consider $\overline{\lambda}_H > \underline{\lambda}_L$. Then denote $(\overline{Q}_P, \overline{Q}_P c), (\overline{q}_H, \overline{q}_H c_H), (\overline{q}_L, \overline{q}_L c_L)$ and $(\underline{Q}_P, \underline{Q}_P c), (\underline{q}_H, \underline{q}_H c_H), (\underline{q}_L, \underline{q}_L c_L)$ the aggregate trades that decentralize the second-best allocation for the respective welfare weights. Under the single-crossing assumption (cf. 3) indifference curves cannot bind for both type. And, noting that the second-best efficient allocation is continuous in welfare weights due to an application of Berge's theorem of the maximum, it suffices to focus on the case where the same incentive constraint binds for distinct welfare weights. Thus distinguish between two cases. First, suppose that H's IC constraint is slack. If so, by construction it must hold that

$$\begin{aligned} \overline{U}_L + \overline{\lambda}_H \max_{\tilde{q}} U_H(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_H) &\geq \underline{U}_L + \overline{\lambda}_H \max_{\tilde{q}} U_H(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_H) \\ \underline{U}_L + \underline{\lambda}_H \max_{\tilde{q}} U_H(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_H) &\geq \overline{U}_L + \underline{\lambda}_H \max_{\tilde{q}} U_H(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_H) \end{aligned}$$

where we denote $\overline{U}_L = U_L(\overline{Q}_P + \overline{q}_L, \overline{Q}_P c + \overline{q}_L c_L)$ and $\underline{U}_L = U_L(\underline{Q}_P + \underline{q}_L, \underline{Q}_P c + \underline{q}_L c_L)$ to facilitate the notation. Taking differences gives

$$\begin{aligned} \overline{U}_L - \underline{U}_L &\geq \overline{\lambda}_H \Big(\max_{\tilde{q}} U_H(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_H) - \max_{\tilde{q}} U_H(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_H) \Big) \\ &\underline{\lambda}_H \Big(\max_{\tilde{q}} U_H(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_H) - \max_{\tilde{q}} U_H(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_H) \Big) \geq \overline{U}_L - \underline{U}_L. \end{aligned}$$

Since $\overline{\lambda}_H > \underline{\lambda}_H$, this poses a contradiction provided that $\max_{\tilde{q}} U_H(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_H) > \max_{\tilde{q}} U_H(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_H)$, or, (in light of $c_H > c$) equivalently if $\underline{Q}_P > \overline{Q}_P$. We deduce that $\overline{Q}_P \geq \underline{Q}_P$ as desired.

The case where L's IC constraint is slack follows from largely symmetric arguments. If so, by construction it must hold that

$$\overline{U}_H + 1/\overline{\lambda}_H \max_{\tilde{q}} U_L(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_L) \geq \underline{U}_H + 1/\overline{\lambda}_H \max_{\tilde{q}} U_L(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_L)$$
$$\underline{U}_H + 1/\underline{\lambda}_H \max_{\tilde{q}} U_L(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_L) \geq \overline{U}_H + 1/\underline{\lambda}_H \max_{\tilde{q}} U_L(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_L)$$

where we denote $\overline{U}_H = U_H(\overline{Q}_P + \overline{q}_H, \overline{Q}_P c + \overline{q}_H c_H)$ and $\underline{U}_H = U_H(\underline{Q}_P + \underline{q}_H, \underline{Q}_P c + \underline{q}_H c_H)$ to facilitate the notation. Taking differences gives

$$\overline{U}_H - \underline{U}_H \geq 1/\overline{\lambda}_H \Big(\max_{\tilde{q}} U_L(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_L) - \max_{\tilde{q}} U_L(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_L) \Big)$$
$$1/\underline{\lambda}_H \Big(\max_{\tilde{q}} U_L(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_L) - \max_{\tilde{q}} U_L(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_L) \Big) \geq \overline{U}_H - \underline{U}_H.$$

Since $\overline{\lambda}_H > \underline{\lambda}_H$, this poses a contradiction provided that $\max_{\tilde{q}} U_L(\underline{Q}_P + \tilde{q}, \underline{Q}_P c + \tilde{q}c_L) < \max_{\tilde{q}} U_L(\overline{Q}_P + \tilde{q}, \overline{Q}_P c + \tilde{q}c_L)$, or, (in light of $c_L < c$) equivalently if $\underline{Q}_P > \overline{Q}_P$. We deduce that $\overline{Q}_P \geq \underline{Q}_P$ as desired.

Part 2: We show that as we move along the second-best frontier and increase the aggregate pooling quantity Q_P by one unit, the corresponding low-risk separating quantity q_L does not increase.[33] For convenience, we here impose quasi-linearity. First note that coupled with strict quasi-concavity, i.e., Assumption 1, this implies that $Q \mapsto U_\theta(Q)$ is strictly concave, whence differentiable almost everywhere. Then consider two cases. First, suppose that H's incentive constraint is slack so that the second-best efficient q_L is the solution to an unconstrained optimization problem. If so, it suffices to show that $\mathbf{q}_L(Q_P) = \arg\max_{\tilde{q}} U_L(Q_P + \tilde{q}) - Q_P c - \tilde{q}c_L$ is non-increasing. This follows readily from familiar arguments: Topkis (1998) shows that $\mathbf{q}_L(Q_P)$ is non-increasing if $(Q_P, \tilde{q}) \mapsto -U_L(Q_P + \tilde{q}) - Q_P c - \tilde{q}c_L$ satisfies nondecreasing differences, which, under quasi-linear utility is equivalent to $Q \mapsto U_L(Q)$ being concave. Second, suppose that H's incentive constraint is binding. Then, due to Assumption 3, L's incentive constraint is slack. In effect, $q_L \equiv \mathbf{q}_L(Q_P)$ is the solution to an implicit equation:

$$U_H(Q_P + \mathbf{q}_L(Q_P)) - Q_P c - \mathbf{q}_L(Q_P)c_L = \max_{\tilde{q}} U_H(Q_P + \tilde{q}) - Q_P c - \tilde{q}c_H \equiv U_H(Q_P + q_H) - Q_P c - q_H c_H.$$

Differentiating with respect to Q_P (and applying an envelope condition) gives

$$\dot{\mathbf{q}}_L(Q_P) = \frac{\dot{U}_H(Q_P + q_H) - \dot{U}_H(Q_P + \mathbf{q}_L(Q_P))}{\dot{U}_H(Q_P + \mathbf{q}_L(Q_P)) - c_L}.$$

[33] q_L non-increasing is likely not necessary. Large pooling, i.e., Condition 3, would hold as long as q_L increases by at most $\frac{c_H - c}{c_H - c_L}$.

This expression is negative because $\dot{U}_H(Q_P + \mathbf{q}_L(Q_P)) > \dot{U}_H(Q_P + q_H) = c_H$ under quasi-linear utility.

Part 3: It remains to show that partial efficiency holds for sufficiently large λ_H/λ_L. Or, since Q_P rises in λ_H/λ_L we prove that $Q_P \mapsto \dot{U}(Q_P + \mathbf{Q_L}(Q_P))$ is decreasing. Then, due to Assumption 2, it follows that $\dot{U}(Q_P + \mathbf{Q_L}(Q_P)) \leq c$ for sufficiently large λ_H/λ_L. Equivalently, we show that $Q_P \mapsto Q_P + \mathbf{Q_L}(Q_P)$ is increasing. Consider the case where H IC is tight. Then let $\mathbf{Q_L}(Q_P) = Q_P + \mathbf{q}_L(Q_P)$ with $\mathbf{q}_L(Q_P)$ as before. From the above we note that for $x > c_H$ it holds that $\dot{\mathbf{Q}}_L(Q_P) = 1 + \frac{c_H - x}{x - c_L} > 0$ which is equivalent to $c_H > c_L$.

To summarize, we have shown that as $\frac{\lambda_H}{\lambda_L}$ increases, the aggregate trades corresponding to a second-best efficient allocation are such that Q_P does not decrease and q_L does not increase. Hence there exists a maximal interval so that for all $\frac{\lambda_H}{\lambda_L} \in [\underline{L}, \overline{L}]$ it holds that $q_L \geq 0$, $(c_H - c_L)q_L \leq (c_H - c)Q_P$, i.e., Condition 3, and partial efficiency is satisfied. $\square$

References

Akerlof, G. A. (1970). The market for" lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, pages 488–500.

Ania, A. B., Tröger, T., and Wambach, A. (2002). An evolutionary analysis of insurance markets with adverse selection. *Games and Economic Behavior*, 40:153–184.

Asriyan, V. and Vanasco, V. (2023). Security design in non-exclusive markets with asymmetric information. *Review of Economic Studies*.

Attar, A., Casamatta, C., Chassagnon, A., and Décamps, J.-P. (2019). Multiple lenders, strategic default, and covenants. *American Economic Journal: Microeconomics*, 11(2):98–130.

Attar, A., Mariotti, T., and Salanié, F. (2011). Nonexclusive competition in the market for lemons. *Econometrica*, 79(6):1869–1918.

Attar, A., Mariotti, T., and Salanié, F. (2014). Nonexclusive competition under adverse selection. *Theoretical Economics*, 9(1):1–40.

Attar, A., Mariotti, T., and Salanié, F. (2021). Entry-proofness and discriminatory pricing under adverse selection. *American Economic Review*, 111(8):2623–59.

Attar, A., Mariotti, T., and Salanié, F. (2022). Regulating insurance markets: Multiple contracting and adverse selection. *International Economic Review*.

Auster, S., Gottardi, P., and Wolthoff, R. P. (2022). Simultaneous search and adverse selection.

Azevedo, E. M. and Gottlieb, D. (2017). Perfect competition in markets with adverse selection. *Econometrica*, 85(1):67–105.

Bauer, D., Russ, J., and Zhu, N. (2020). Asymmetric information in secondary insurance markets: Evidence from the life settlements market. *Quantitative Economics*, 11(3):1143–1175.

Bernheim, B. D. and Whinston, M. D. (1986). Common agency. *Econometrica*, pages 923–942.

Biais, B., Martimort, D., and Rochet, J.-C. (2000). Competing mechanisms in a common value environment. *Econometrica*, 68(4):799–837.

Bisin, A. and Gottardi, P. (1999). Competitive equilibria with asymmetric information. *Journal of Economic Theory*, 87(1):1–48.

Bisin, A. and Gottardi, P. (2003). Competitive markets for non-exclusive contracts with adverse selection: The role of entry fees. *Review of Economic Dynamics*, 6(2):313–338.

Bisin, A. and Gottardi, P. (2006). Efficient competitive equilibria with adverse selection. *Journal of Political Economy*, 114(3):485–516.

Cawley, J. and Philipson, T. (1999). An empirical examination of information barriers to trade in insurance. *American Economic Review*, 89(4):827–846.

Chade, H. and Schlee, E. (2012). Optimal insurance with adverse selection. *Theoretical Economics*, 7(3):571–607.

Chen, D. and Duffie, D. (2021). Market fragmentation. *American Economic Review*, 111(7):2247–74.

Chiappori, P.-A., Jullien, B., Salanié, B., and Salanie, F. (2006). Asymmetric information in insurance: General testable implications. *The RAND Journal of Economics*, 37(4):783–798.

Chiappori, P.-A. and Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of political Economy*, 108(1):56–78.

Crocker, K. J. and Snow, A. (1985). The efficiency of competitive equilibria in insurance markets with asymmetric information. *Journal of Public Economics*, 26(2):207–219.

Donaldson, J. R., Gromb, D., and Piacentino, G. (2020). The paradox of pledgeability. *Journal of Financial Economics*, 137(3):591–605.

Einav, L. and Finkelstein, A. (2023). *We've Got You Covered: Rebooting American Health Care*. Portfolio.

Einav, L., Finkelstein, A., and Levin, J. (2010). Beyond testing: Empirical models of insurance markets. *Annual Review of Economics*, 2(1):311–336.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, 103(1):178–219.

Encinosa, W. (2001). A comment on neudeck and podczeck's "adverse selection and regulation in health insurance markets". *Journal of Health Economics*, 20(4):667–673.

Engers, M. and Fernandez, L. (1987). Market equilibrium with hidden knowledge and self-selection. *Econometrica: Journal of the Econometric Society*, pages 425–439.

Farinha Luz, V. (2017). Characterization and uniqueness of equilibrium in competitive insurance. *Theoretical Economics*, 12(3):1349–1391.

Farinha Luz, V., Gottardi, P., and Moreira, H. (2023). Risk Classification in Insurance Markets with Risk and Preference Heterogeneity. *The Review of Economic Studies*, 90(6):3022–3082.

Finkelstein, A. and Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the uk annuity market. *Journal of Political Economy*, 112(1):183–208.

Glazer, J. and McGuire, T. G. (2006). Contending with risk selection in health insurance markets in germany. *Perspektiven der Wirtschaftspolitik*, 7:75–91.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *Journal of Finance*, 49(4):1127–61.

Gottlieb, D. and Moreira, H. (2023). Market power and adverse selection. Technical report.

Grossman, H. I. (1979). Adverse selection, dissembling, and competitive equilibrium. *The Bell Journal of Economics*, 10(1):336–343.

Guerrieri, V., Shimer, R., and Wright, R. (2010). Adverse selection in competitive search equilibrium. *Econometrica*, 78(6):1823–1862.

Harris, M. and Townsend, R. M. (1981). Resource allocation under asymmetric information. *Econometrica*, 47(1):33–64.

Hellwig, M. (1988). A note on the specification of interfirm communication in insurance markets with adverse selection. *Journal of Economic Theory*, 46(1):154–163.

Hellwig, M. F. (1983). *On moral hazard and non-price equilibria in competitive insurance markets.*

Hendren, N. (2013). Private information and insurance rejections. *Econometrica*, 81:1713–1762.

Huang, D. (2022). Adverse selection and equilibrium existence in almost nonexclusive markets.

Jaynes, G. D. (1978). Equilibria in monopolistically competitive insurance markets. *Journal of Economic Theory*, 19(2):394–422.

Malamud, S. and Rostek, M. (2017). Decentralized exchange. *American Economic Review*, 107(11):3320–62.

McFadden, D., Noton, C., and Olivella, P. (2015). Minimum coverage regulation in insurance markets. *SERIEs*, 6:247–278.

Mimra, W. and Waibel, C. (2021). Contracting environments and efficiency in markets with hidden information: An experiment.

Mimra, W. and Wambach, A. (2014). New developments in the theory of adverse selection in competitive insurance. *Geneva Risk and Insurance Review*, 39:136–152.

Miyazaki, H. (1977). The rat race and internal labor markets. *Bell Journal of Economics*, 8(2):394–418.

Netzer, N. and Scheuer, F. (2014). A game theoretic foundation of competitive equilibria with adverse selection. *International Economic Review*, 55(2):399–422.

Neudeck, W. and Podczeck, K. (1996). Adverse selection and regulation in health insurance markets. *Journal of Health Economics*, 15(4):387–408.

Pauly, M. V. (1974). Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection. *The Quarterly Journal of Economics*, 88(1):44–62.

Peters, M. (2001). Common agency and the revelation principle. *Econometrica*, 69(5):1349–1372.

Riley, J. G. (1979). Informational equilibrium. *Econometrica*, 47(2):331–359.

Rostek, M. and Yoon, J. H. (2021). Exchange design and efficiency. *Econometrica*, 89(6):2887–2928.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649.

Stiglitz, J. E., Yun, J., and Kosenko, A. (2020). Bilateral information disclosure in adverse selection markets with nonexclusive competition. Technical report, National Bureau of Economic Research.

Tirole, J. (2012). Overcoming adverse selection: How public intervention can restore market functioning. *American Economic Review*, 102(1):29–59.

Topkis, D. M. (1998). *Supermodularity and Complementarity.* Princeton University Press.

Veiga, A. and Levy, Y. (2023). Optimal contract regulation in selection markets. Technical report.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, 16(2):167–207.