Insurance as a lemons market: Coverage denials and pooling✩

Hector Chade, Edward E. Schlee *

Department of Economics, Arizona State University, United States of America

Received 16 April 2018; final version received 2 June 2020; accepted 17 June 2020
Available online 30 June 2020

Abstract

The standard monopoly insurance model with adverse selection implies that there are always gains to trade, that only the best (unobservable) risks can go uninsured, and that a profit-maximizing menu cannot pool all types. We show that insurance-provision costs can explain both coverage denials only to those likely to be the worst risks and complete pooling. Specifically, we prove a general comparative statics theorem formalizing coverage denials only to those deemed to be the worst risks; and two theorems showing that the insurer offers a single contract (complete pooling), with either zero or positive coverage. We point out some implications of these results for empirical work on insurance. Our results expand upon a point made by Hendren (2013), that the main effect of adverse selection on insurance might not be misallocation in active markets – the traditional emphasis after Rothschild and Stiglitz (1976) – but simply in shutting down markets, as in Akerlof (1970) classic lemons model.

✩ A previous version was entitled “Coverage Denied: Excluding Bad Risks, Inefficiency, and Pooling in Insurance.” We thank three referees and the Associate Editor for comments that improved our exposition. We are especially grateful to Dan Silverman for many comments and discussions. We also thank Camelia Bejan, Philip Bond, Yann Braouezec, Sandro Brusco, Ying Chen, Rohan Dutta, Liran Einav, Nathaniel Hendren, Fahad Khalil, Ilse Lindenlaub, George Mailath, Mihai Manea, Alejandro Manelli, Peter Norman, Asaf Plan, François Salanié, Andy Skrzypacz, conference participants at the 2013 SAET conference, Economic Theory Workshop at the University of Montreal, 2015 SWET conference, 2015 CEPET conference, 2015 World Congress of the Econometric Society, the 2018 NSF/CEME Decentralization Conference at Duke, and seminar participants at Arizona State, Arizona, Johns Hopkins, Michigan State, Kentucky, Ryerson, SUNY-Stony Brook, Washington–Bothell, Washington–Seattle, the St. Louis Fed, Universidad Catolica de Chile, and Universitat Pompeu Fabra for comments and discussions. Schlee’s research was supported in part by grants from the W.P. Carey School at Arizona State University, and by a grant from the Toulouse School of Economics. Part of this work was done during Schlee’s visit to Johns Hopkins University in the Spring of 2012, and Chade’s visit to Pompeu Fabra in the Summer of 2012.

* Corresponding author.
E-mail addresses: hector.chade@asu.edu (H. Chade), edward.schlee@asu.edu (E.E. Schlee).
1. Introduction

There are three striking predictions of the adverse selection model of monopoly insurance pioneered by Stiglitz (1977): there are always ex ante gains to trade between the insurer and a consumer with private information about the probability of a loss; only the best (unobservable) risks—those least likely to suffer a loss—are uninsured; and under plausible assumptions, there is complete sorting—each type gets a distinct contract.

The uninsured in the standard model go without coverage voluntarily, in the sense that the insurer offers each consumer a menu of contracts, but consumers with low enough probability of suffering a loss choose zero coverage. Casual evidence—as well as Gruber (2008), McFadden et al. (2012), Hendren (2013) and Braun et al. (2019)—makes clear that some consumers are involuntary uninsured in the sense that they are not offered any (nonzero) contracts, and that those who are denied coverage are the ones perceived to be the worst risks. For example, some insurers refuse to write health care policies for consumers with ‘pre-existing’ adverse health conditions.\(^1\) We know of no evidence, casual or otherwise, that shows that those perceived to be good risks are denied coverage.

To write the obvious, insurers deny coverage to a consumer because they expect to lose money from any policy the consumer would accept. Such a belief presumably comes from observing an attribute of a consumer, for example a medical history. So one might think that including in the model a noisy signal correlated with the consumer’s loss chance could account for these coverage denials. But such an inclusion is not enough: Chade and Schlee (2012) find that the monopolist always makes positive expected profit, i.e., there are ex ante gains to trade between the insurer and the consumer, no matter how pessimistic the insurer’s beliefs about the consumer’s loss chance.

Regarding sorting, a well-known fact about standard contracting models is that there will be a continuum of different contracts when there are a continuum of types. And standard conditions imply complete sorting of types. It is also a commonplace complaint that the theory predicts too many contracts, conditional on observables. The same complaint of course applies to the insurance model in Chade and Schlee (2012).

The standard model assumes that the insurer’s only cost is payment of claims: insurance provision itself is costless. Yet both the empirical insurance and practitioner literatures extensively discuss provision costs (sometimes referred to as administrative costs or simply loading). These costs are not small: in line with the number reported earlier by Diamond (1992), Gruber (2008) writes that administrative costs average 12% of the premium paid by consumers in the US health

---

\(^1\) Appendix F in Hendren (2013) contains an excerpt from Genworth Financial’s underwriting guidelines for long-term care insurance. There are two pages of ‘uninsurable conditions.’ Lists of uninsurable conditions from other insurers can be found at Hendren’s research page, http://scholar.harvard.edu/hendren/publications. See also the discussion on ‘lemon dropping’ (exclusion of bad risks) in the recent survey by Einav et al. (2010).
insurance industry. We show in this paper that adding the obvious to the standard insurance model—that insurance is costly—can neatly and simultaneously account for coverage denials for only bad risks and few contracts being offered by the insurer.

We begin with two results that explain when provision costs destroy trade between the insurer and a consumer. Then we turn to cases when there is trade, and show how these costs overturn well-known properties of insurance policies under adverse selection.

That trade costs reduce trade comes as no surprise. Our goal is different: to say which insurance trade costs explain the lack of trade precisely for those perceived to be the worst risks. As we explain in at the end of Section 4.1, not all forms of provision cost generate this prediction. But we do show that two commonly-discussed provision costs do: multiplicative loading, and a fixed cost of claims processing.

Specifically, we begin by proving a no-trade comparative static result (Theorem 1, Section 3.1): it shows that if there are no gains to trade at a belief about a consumer’s loss chance, then there are no gains to trade at any belief that is worse in the sense of likelihood ratio dominance (i.e., worse news about the consumer’s loss chance). Theorem 1 would be obvious if those with a higher loss chance were always less profitable. But such profit monotonicity fails here. An increase in the loss chance increases the demand for insurance as well as its cost: the most profitable consumers are generally those with intermediate, not extreme, loss chances. Theorem 1 does not assert that the insurer’s expected profit always falls with worse news, just that if it ever equals zero, it remains zero with worse news. As the proof suggests, the result extends beyond insurance to a large class of principal-agent models, a fact we explore in Section 7. Moreover, if we impose a smoothness assumption on the provision cost function, we can pin down precisely who trades and who does not.

These results are about the failure of trade. A natural question to ask is how provision costs affect profit-maximizing menus when there is trade. The second part of the paper takes up this question. Absent provision costs, Chade and Schlee (2012) confirm that three classic contracting properties hold for monopoly insurance under adverse selection: no consumer risk type pools with the ‘highest’ type, the one with the highest loss chance (no-pooling at the top); the highest type gets an interim efficient contract (efficiency at the top); and each of the other types gets a contract that is distorted downwards from an interim efficient one (downward distortions elsewhere). All three can fail with provision costs. More importantly, we show that all types can be pooled on a positive contract (Theorems 2 and 3 and Example 2 in Section 5). These results give another explanation for why firms offer relatively few contracts.

These results are relevant for applied work. First, empirical research on adverse selection has emphasized markets that are actively traded. In particular, it has emphasized the positive correlation property that riskier consumers buy more coverage. Our Theorems 1-3 together suggest that this emphasis might be misplaced. With pooling—either on a no-trade contract or positive insurance—adverse selection in this sense disappears: provision costs can make monopoly insurance markets resemble Akerlof’s classic lemons example, in which the main effect of adverse selection is simply that markets shut down. Hendren (2013) emphasizes the shutdown of trade in insurance because of adverse selection. We extend this insight to allow for provision costs, and go beyond the shutdown point to show that active markets can involve complete pooling, further weakening the positive correlation property. We formalize these points in Corollary 2

---

2 These costs are mentioned prominently in standard textbook and survey treatments of insurance under complete information (e.g. Rees and Wambach (2008)).
(Section 5.3), which unifies the two parts of our paper. Second, our Theorem 1 and Proposition 1 give new theoretical foundations for the “tail test” test for adverse selection in Hendren (2013) (namely, that the estimated loss distributions for those denied coverage have thicker right tails). Third, an even more direct connection is Braun et al. (2019), who build on results of an earlier draft of this paper. They calibrate a special case of our monopoly model, modified to allow for a public option, to data on the U.S. long-term care insurance market. They use it to explain some important facts about this market, in particular, rejections of bad risks, low take-up rates across all income groups, contracts with small indemnities, and a lack of correlation between nursing home risk and insurance coverage. Although the U.S. long-term care insurance market is not a literal monopoly, they argue that the concentration in this market makes a monopoly model a good approximation.

Although our monopoly assumption seems strong, it is precisely what is needed for our results on coverage denials: since profit is highest for a monopoly, it follows that if there are no gains to trade with a monopoly, then there are no gains to trade with any other market structure. Our pooling results do explicitly use the monopoly assumption. As just noted, Braun et al. (2019) show that an empirically-calibrated monopoly model can account for important facts about the U.S. long-term care insurance. And Cohen and Einav (2007, p. 749) argue that a monopoly model might be apt for the Israeli automobile insurance company that they analyze. So a monopoly model seems relevant for applied work even for markets that are not literal monopolies. Still, it is natural to ask whether our pooling results hold beyond monopoly. We show that they do not for the Rothschild-Stiglitz competitive model of insurance with adverse selection (Section 6.1). But we sketch a duopoly model illustrating that our pooling results are robust to allowing some forms of competition (Section 6.2).

Our paper is obviously related to the large literature on insurance under adverse selection (see Chade and Schlee (2012) and the papers they cite). Besides Stiglitz (1977), the two most-closely related papers are those by Hendren (2013) and Braun et al. (2019) that we already mentioned. In addition to giving a new theoretical foundation for Hendren’s tail test for coverage denials, our Proposition 1 generalizes his Theorem 1 to positive provision costs.

Dionne et al. (1999), Ramsay and Oguledo (2012), Liu and Browne (2007), Ramsay et al. (2013), and Allard et al. (2011) have results on insurance with loading or a fixed cost per contract for models with two consumer types. Allard et al. (2011) show that pooling can occur in a competitive market when there is a fixed cost of writing a contract; Ramsay et al. (2013) show graphically that pooling can occur with loading for a monopoly; and de Meza and Webb (2016) add claims processing costs to a competitive market. None though have our no-trade comparative static theorem or our general theorems on pooling for a continuum of types.

A large literature has explored pooling in screening models. We single out two recent contributions. Toikka (2011) has general results on pooling in a class of common-values models with quasilinear utility, results we exploit in Section 5 for the constant absolute risk aversion (CARA) case. And Gottlieb and Moreira (2017) show that the interaction between adverse selection and moral hazard can lead to pooling in a Principal-Agent model with limited liability and risk neutrality. Their pooling explanation is complementary to ours (and for a different model); our contribution is to show that provision costs can simultaneously explain coverage denials to only the worst risks and pooling.
2. A model of costly insurance provision

Except for provision costs, the model is the standard monopoly insurance model pioneered by Stiglitz (1977) and which we generalize in Chade and Schlee (2012). A consumer has initial wealth of \( w > 0 \), faces a potential loss of \( \ell \in (0, w) \) with chance \( \theta \in [0, 1] \), and has preferences represented by a \( C^2 \), strictly concave von Neumann-Morgenstern utility function \( u \) on \( \mathbb{R}^+ \), with \( u' > 0 \) and \( u'' < 0 \). The consumer knows the loss chance \( \theta \), but the insurer does not. From now on we call \( \theta \) the consumer’s type.

The monopoly insurer is risk neutral. It has a belief \( F \) (a cumulative distribution function) about the consumer’s type with support \( \Theta_F \subset [0, 1] \). To allow for both discrete and continuous type distributions, we assume throughout that any possible insurer belief has a continuous density with respect to a fixed measure \( \mu \) on \( [0, 1] \): for any belief \( F \), there is a continuous function \( f \) on the support of \( F \) with \( F(\theta) = \int_0^\theta f(s) d\mu(s) \) for every \( \theta \in [0, 1] \). We assume that \( \mu \) is either Lebesgue measure or a finite-support measure. The firm chooses a (measurable) menu of contracts, that is, for each \( \theta \in \Theta_F \), a point \( (x, t) \in \mathbb{R} \) where \( t \) is the premium and \( x \) the payment in the event of a loss. The expected utility of a type-\( \theta \) consumer for a contract \( (x, t) \) is

\[
U(x, t, \theta) = \theta u(w - \ell + x - t) + (1 - \theta) u(w - t).
\]

Except for allowing \( \theta = 1 \) in the type support, the model is so far standard. We introduce insurance provision costs, by which we mean that the expected cost of coverage \( x > 0 \) to a type-\( \theta \) consumer is greater than expected payment of claims, \( \theta x \). We write the expected cost of a contract \( x \) given to type \( \theta \) as \( C(x, \theta) = \theta c(x) \). We assume throughout that (i) \( c(x) \geq x \), (ii) \( c \) is strictly increasing, and (iii) \( c(0) = 0 \). Two commonly-mentioned provision costs in the literature are multiplicative loading (implying that the expected marginal cost of coverage exceeds the loss chance); and a fixed claims processing cost (which happens only in the event of a loss).\footnote{See Arrow (1965), Boland (1965), Lees and Rice (1965), Shavell (1977), Diamond (1977), and Gollier (2000). Besides Braun et al. (2019), Lockwood (2014), Brown and Finkelstein (2011), and Pauly and Percy (2000) estimate—formally or informally—the size of provision costs.}

We use as a leading example a cost function that includes these two costs, which clearly satisfies assumptions (i)–(iii).

Example 1. (Multiplicative Loading, Fixed Claims Cost) For a contract with coverage \( x \), the expected cost is

\[
C(x, \theta) = \theta c(x) = \begin{cases} 
\theta(\lambda x + k) & \text{if } x > 0 \\
0 & \text{if } x = 0 
\end{cases}
\]

where \( \lambda \geq 1 \) is a multiplicative loading factor and \( k \geq 0 \) is a fixed cost when a claim is made. We call the \( k = 0, \lambda > 1 \) case pure multiplicative loading.

In the example \( c \) is discontinuous at \( x = 0 \) when \( k > 0 \). We do not impose continuity as a maintained assumption partly to cover the case of a fixed claims cost. It is this version of the cost side of our model that Braun et al. (2019) calibrate; they set \( k = 0.019 \) and \( \lambda = 1.195 \) to match industry data for fixed and variable costs as a percentage of premiums in long-term care insurance.

Chade and Schlee (2012) prove that the insurer offers nonnegative contracts with coverage not greater than the loss and premium not greater than the coverage. Here we simply impose...
these conditions as constraints. We denote by $C$ the set of contracts satisfying $0 \leq x \leq \ell$ and $0 \leq t \leq x$.\footnote{The constraint $x \leq \ell$ does not bind, and neither does $(x, t) \geq 0$ for natural extensions of $c$ to $x < 0$.} By the revelation principle, the insurer’s problem is to choose a measurable menu of contracts \( \{x(\cdot), t(\cdot)\} \) with range in $C$ to solve

$$
\max_{\Theta_F} \int [t(\theta) - C(x(\theta), \theta)] dF(\theta)
$$

subject to

$$
U(x(\theta), t(\theta), \theta) \geq U(x(\theta'), t(\theta'), \theta) \quad \text{for} \ \theta, \theta' \in \Theta_F, \quad (\text{IC})
$$

$$
U(x(\theta), t(\theta), \theta) \geq U(0, 0, \theta) \quad \text{for} \ \theta \in \Theta_F. \quad (\text{P})
$$

Here IC stands for incentive compatibility and P for participation constraints. Since $U$ satisfies the strict single crossing property in $(x, t)$ and $\theta$, any menu that satisfies (IC) is increasing: if $\theta' > \theta$, then $x(\theta') \geq x(\theta)$ and $t(\theta') \geq t(\theta)$. This follows from standard comparative statics results, for example, Theorem 4’ in Milgrom and Shannon (1994).

3. Two-type example

In this section we present a simple example with two types ($0 < \theta_L < \theta_H < 1$), CARA preferences ($u(w) = -e^{-rw}$, $r > 0$), and pure multiplicative loading ($C(x, \theta) = \theta_1 x$, $\lambda > 1$) to illustrate all of our results using elementary arguments.

The CARA assumption eliminates wealth effects and allows us to represent the consumer’s preferences for $(x, t)$ when the type is $\theta$ in the quasilinear form $v(x, \theta) - t$, where $v(x, \theta) = -(1/r) \log(1 - \theta + \theta e^{\ell(x)}), v_1 > 0, v_{xx} < 0$, and $v_{xt} > 0$ (Example 3 in Chade and Schlee (2012)). From standard arguments, the only binding (IC) constraint in the insurer’s problem is for the high type and the only binding (P) constraint is for the low type. Moreover, incentive compatibility is equivalent to the incentive constraint for the high type, plus monotonicity $x_H \geq x_L$ where the subscript denotes the type receiving coverage $x$. These facts imply we can write the insurer’s problem as

$$
\max_{x_L, t_L, x_H, t_H} (1 - p)(t_L - \theta_L x_L) + p(t_H - \theta_H x_H)
$$

subject to

$$
v(x_H, \theta_H) - t_H = v(x_L, \theta_H) - t_L, \quad v(x_L, \theta_L) - t_L = v(0, \theta_L),
$$

and $x_H \geq x_L \geq 0$, where $p$ is the probability that the consumer is of high type.

If $\lambda = 1$ then, since $\theta_H < 1$, $x_H = \ell$ (full insurance), and $0 \leq x_L < \ell$. Without provision costs, there is always a positive probability of trade and no pooling, as in Stiglitz (1977). If $\lambda > 1$, however, there can be no trade. This is obvious from (2) if $\lambda \theta_L \geq 1$, but there can be no trade even if $\lambda \theta_H < 1$. Indeed, we show in Appendix A.1 that there is no trade if and only if

$$
\lambda \geq \max \left\{ \frac{v_x(0, \theta_H)}{\theta_H}, \frac{v_x(0, \theta_L)}{p \theta_H + (1 - p) \theta_L} \right\}.
$$

The first term on the right side of (3) asserts that there are no gains to trade between the $\theta_H$ and
the insurer, so no menu with $x_H > 0$ and $x_L = 0$ is profitable; the second asserts that there are no gains to trade with any pooling contract $(x, t)$ with $t = v(x, \theta_L) - v(0, \theta_L)$, the premium for which participation binds for $\theta_L$ at coverage $x$. Proposition 1 generalizes this no-trade example to an arbitrary number of types, general expected utility preferences, and a general cost function that is differentiable at $x = 0$.

The left side of (3) is increasing in the fraction $p$ of high types. Since (3) is necessary and sufficient for coverage to be denied, it follows here that if coverage is denied at some insurer belief $p$, then coverage is denied at any belief $p' > p$. A two-type distribution with a higher $p$ likelihood ratio dominates a distribution with a lower $p$. Our Theorem 1 extends this likelihood ratio comparative static to an arbitrary number of types, general expected utility preferences, and a general cost function.

Besides pooling on the no-trade contract, the profit-maximizing menu can entail pooling at a positive contract. In Appendix A.1 we show that the insurer pools both types at a positive contract if $p$ is low enough. Theorems 2–3, Corollary 2, and Example 2 substantially generalize this example.

For an intuition of pooling, the proof reveals that the full-information coverage, denoted by $x^*(\theta)$, is decreasing in $\theta$ under CARA. But any incomplete-information menu must be increasing by incentive compatibility. This conflicting monotonicity is well-known in screening problems with quasi-linear preferences (see for example Laffont and Martimort (2002) pp. 53–56). It does not arise in our problem if $\lambda = 1$, but if $\lambda > 1$ and $p$ is low enough, it is severe enough to preclude sorting the two types. Fig. 1 helps illustrate the possibility of pooling on a positive contract.

The rest of the paper generalizes this two-type example to an arbitrary number of types, general provision-cost functions, and general expected utility preferences.
4. No trade: coverage denials for only bad risks

If the value of problem (1) is zero, then there are no \textit{ex ante} gains to trade and we say that coverage is denied. Formally, if the value is zero, then a menu of no-trade contracts is interim weakly Pareto Optimal: there is no other menu of contracts that the firm and almost-every type strictly prefer to the no-trade menu.\footnote{There are knife-edge cases in which the maximum expected profit is zero, a positive measure of consumers can obtain gains to trade, but a small cost decrease will imply gains to trade. In the smooth version of our model in Section 4.2, even these knife-edge cases disappear. Since we use the revelation principle and impose participation (P) for every type, the worst contract the insurer can offer a type is the no-trade contract. If instead we allowed the insurer to offer an arbitrary menu of contracts, it is possible that the insurer would offer a menu of contracts that no type would accept. On our definition coverage would still be denied since there are no gains to trade with such a menu.} Chade and Schlee (2012) show that there are always ex ante gains to trade between the insurer and the consumer when insurance provision is costless (they assume that the type set does not include $\theta = 1$). Obviously, there are no gains to trade if provision costs are large enough. What is less obvious is what restrictions on these costs imply that the insurer denies coverage only to those perceived to be bad risks.

We prove two results on coverage denials to the worst risks. The first is a general comparative statics result that imposes minimal assumptions on the expected cost function $C$, and in particular covers the discontinuous cost in Example 1. It does not allow us, however, to pin down who trades and who does not. By restricting the cost function further, our second result gives a necessary and sufficient condition for no trade.

4.1. A general no-trade comparative statics theorem

Our first theorem asserts that if coverage is denied at some belief, then it is denied at any belief that is more pessimistic about the consumer’s type. To interpret the result, let the insurer observe a signal correlated with the consumer’s type before writing a menu of contracts. The signal is public—both parties observe its realization—and could be costly or costless, its informativeness exogenous or endogenous. An obvious example of an endogenous, costly signal in insurance is underwriting. After observing the signal, the insurer updates beliefs about the consumer’s type and decides what contracts to offer. We interpret $F$ in (1) as the insurer’s posterior belief after observing the signal.\footnote{It is wlog to work directly with posterior beliefs rather than signals since ‘observing the posterior’ is equivalent to ‘observing the signal realization’ (see Kihlstrom (1984)). We do not model the cost of a signal since we envision it as already sunk when the insurer offers a menu (as in underwriting). Also, we assume that the signal is observed before writing a menu and is public. If it were privately observed by the principal, then the model would become a complex informed principal problem with two-sided private information. And if it were revealed after the mechanism is played, then the principal could benefit from the correlation between the signal and the agent’s type by conditioning the contracts on both. Although these variations are interesting extensions, ours seems more natural for insurance.}

To model the idea of becoming more pessimistic we use the Likelihood Ratio (LR) order. The cdf $G$ LR-dominates $F$ if the corresponding ratio of densities $f/g$ is decreasing; with two types, an increase in the fraction of high types results in an LR-dominant change in beliefs. As a foundation for this choice, recall that Milgrom (1981) defined one signal realization $y$ to be \textit{better news} about a parameter $\theta$ than realization $x$ if the posterior belief after observing $y$ first-order stochastically dominates the posterior belief after observing $x$ for every prior belief.\footnote{The cdf $G$ first-order stochastically dominates $F$ ($G \text{ FOSD } F$) if $\int h(\theta)dG(\theta) \geq \int h(\theta)dF(\theta)$ for every increasing function $h$.} One reason for the qualifier “for every prior” is to make “better news” about the signal distribution, rather
than the Bayesian decision maker. He proved that \( y \) is better news than \( x \) if and only if the signal distribution has the monotone likelihood ratio property (MLRP, Propositions 1 and 2). Milgrom (1981) does not explicitly mention this fact, but the MLRP holds for the signal distributions if and only if the posterior after observing the higher signal LR dominates the posterior from the lower signal. This follows immediately from his equation (1).\(^8\)

**Theorem 1 (No-Trade Comparative Statics).** If there are no gains to trade at some belief, then there are no gains to trade at any belief that likelihood ratio dominates it.

If the insurer’s equilibrium profit were a decreasing function of the loss chance \( \theta \), then Theorem 1 would be obvious: increasing the probability of less profitable types would decrease profit. But profit is generally highest for intermediate, not extreme, types. This fact is evident even in the complete-information case. With complete information and no provision cost, the insurer’s profit is just the risk premium, which is a concave function of \( \theta \), and equals zero at the extreme types \( \theta = 0 \) and \( \theta = 1 \), positive otherwise. With incomplete information, the insurer’s maximum profit is not decreasing in the LR order, as the two-type case makes clear: if complete-information profit is positive for the high type, then the insurer’s incomplete-information profit is \( U \)-shaped as a function of the fraction of high types (Fig. 2). The proof of Theorem 1 does not show that the insurer’s incomplete-information profit is decreasing in the LR order, just that, if it equals zero at some belief, it remains zero at any belief that LR dominates it. Example 3 in Appendix A.2 shows that we cannot replace LR dominance by FOSD in Theorem 1.

The proof of Theorem 1 is easy for the two-type case. Suppose that the insurer cannot earn positive profit at some belief about the consumer’s type. Since the no-trade contract is costless and the consumer’s preferences satisfy the single-crossing property, the insurer cannot earn positive profit from the high type at any positive contract: otherwise, the insurer could simply substitute the no-trade contract for the low-type’s unprofitable positive contract. But then by the single-crossing property the resulting menu which leaves the coverage the same for the high type earns positive profit. So if the insurer cannot earn positive profit at some belief, then any low-type profit cannot make up for losses from the high type. If the high type now becomes more likely (a rightward-LR change), then it is all the more the case that low-type profit cannot make up for high-type losses. The argument for an arbitrary number of types is more subtle, and the details can be found in Appendix A.2.

An important application of Theorem 1 is to Hendren (2013), who determines whether the estimated loss distribution of potential consumers who are denied coverage have fatter right tails

---

\(^8\) If \( f \) is the conditional density of the signal given a parameter, then the signal distribution satisfies the MLRP if for every \( \theta’ > \theta \), \( f(y|\theta’)/f(y|\theta) > f(x|\theta’)/f(x|\theta) \). Letting \( g \) be the prior density (with respect to a fixed measure \( \mu \) in our setting) and \( g(\theta|x) \) the posterior density of \( \theta \) after observing \( x \), it follows from Bayes’s rule that (this is Milgrom’s (1981) equation (1))

\[
\frac{g(\theta'|z)}{g(\theta|z)} = \frac{g(\theta')}{g(\theta)} \frac{f(z|\theta')}{f(z|\theta)}
\]

for \( z = x, y \), so that \( y \) is better news than \( x \) if and only if \( g(\theta'|y)/g(\theta|y) > g(\theta'|x)/g(\theta|x) \) for every \( \theta' > \theta \), that is, if and only if \( g(\cdot|y) \) likelihood ratio dominates \( g(\cdot|x) \). This last observation gives us two equivalent foundations for the LR order: better news in the sense of Milgrom (1981); and that the distribution of signals that the insurer observes satisfies the MLRP. We thank the Associate Editor for suggesting this footnote and for pointing out that Milgrom (1981) does not explicitly mention the equivalence between the MLRP for the signal distributions and the LR order for the posteriors.
than consumers who get coverage. Theorem 1 gives a new foundation for comparing these right tails of the distributions.

The only properties of $C$ that we use in the proof of Theorem 1 are that $C(0, \theta) = 0$ and that $C(x, \theta)/\theta$ is increasing in $\theta$. These weak properties suggest that Theorem 1 holds quite generally beyond insurance. We develop this suggestion in Section 7.

4.2. Smooth cost

Theorem 1 does not pin down exactly who trades and who does not. We can do this by adding smoothness and curvature conditions to cost.

Let $MRS(\theta) = -U_x(0, 0, \theta)/U_t(0, 0, \theta) = \theta u'(w - \ell)/(\theta u'(w - \ell) + (1 - \theta)u'(w))$ be type-$\theta$’s marginal rate of substitution of $x$ for $t$ at the no-trade contract (graphically, it is the slope of type-$p$’s indifference curve in the $(x, t)$ space at the origin). Also, suppose that $c$ is differentiable at $x = 0$ and let $\mathbb{E}[\theta c'(0)|\theta \geq \hat{\theta}]$ be the expectation (with respect to $\theta$) of the marginal cost of $x$ for each $\theta$, conditional on the event $\{\theta \geq \hat{\theta}\}$. 

**Proposition 1** (No Trade with Smooth Cost). Fix a belief $F$. Suppose that $c$ is $C^1$ and convex on $[0, \ell]$ for each $\theta \in [0, 1]$.

(i) There are no gains to trade if and only if

$$MRS(\hat{\theta}) \leq c'(0)\mathbb{E}[\theta|\theta \geq \hat{\theta}] \quad \text{for all} \quad \hat{\theta} \in \Theta_F;$$

(ii) If $c(x) > x$ for all $x$, a sufficient condition for (4) is $MRS(\mathbb{E}[\theta]) \leq c'(0)\mathbb{E}[\theta]$.

The proposition implies that there are no gains to trade—coverage is denied—if and only if there are no gains to trade with a two-contract menu, with one the zero contract and the other a

---

Footnote: That $C(0, \theta) = 0$ rules out a fixed cost of writing a menu of contracts. In Fig. 2, a fixed menu cost between $V(1)$ and $V_{min}$ would overturn the conclusion of Theorem 1.
small, positive contract, since the right side of (4) is the slope of the insurer’s isoprofit line from a contract that pools all types above $\hat{\theta}$. If (4) holds at some belief $F$, then it holds at any other belief $G$ that LR dominates $F$, since the expectation on the right side increases. The smooth cost assumption allows us to sharpen the conclusion of Theorem 1: since (4) only depends on $F$ through $\mathbb{E}[\theta | \theta \geq \hat{\theta}]$, it holds for the weaker hazard rate order (HR). That is, if (4) holds at $F$, then it holds at any other belief $G$ that HR dominates it. (A cdf $G$ hazard rate dominates $F$ if $(1 - G)/(1 - F)$ increasing; it implies that $\mathbb{E}[\theta | \theta \geq \hat{\theta}]$ is larger under $G$ than under $F$ for all $\hat{\theta}$; see Shaked and Shanthikumar (2007) Section 1.B, equation (1.B.7)). Note that (4) collapses to (3) for the two-type example in Section 3; and that Example 3 in Section A.2 applies here to show that we cannot replace HR dominance by FOSD.

The inequalities in (4) reveal that if $c'(0)$ is not too large and $\theta_H < 1$, then there are gains to trade. Let $\theta_H < 1$ be the highest type. Since $MRS(\theta_H) > \theta_H$, $MRS(\theta_H) > c'(0)\theta_H$ for $c'(0)$ close to 1. And $\mathbb{E}[\theta | \theta \geq \theta_H] = \theta_H$ implies that (4) fails at $\hat{\theta} = \theta_H$.

When $C(x, \theta) = x\theta$, Proposition 1 (i) specializes to Theorem 1 in Hendren (2013); our Proposition 1 extends his result to the case of costly insurance provision. 10 By Theorem 1 (vi) in Chade and Schlee (2012), a necessary condition for coverage denials in this case of costless insurance is that the highest type suffers a loss with probability one: type $\theta = 1$ is in the support of the insurer’s belief. 11

Proposition 1 (ii) says that all the inequalities in (4) hold if a single inequality holds. This can be useful when $\Theta_F$ contains a large number of types. For an intuition, suppose that uncertainty is symmetric, in the sense that neither the insurer nor the consumer know the consumer’s loss chance. Then the insurer offers a contract tailored to the mean loss chance. One can easily verify that there is trade in the symmetric case if and only if $MRS(\mathbb{E}[\theta]) > c'(0)\mathbb{E}[\theta]$. So one can interpret inequality $MRS(\mathbb{E}[\theta]) \leq c'(0)\mathbb{E}[\theta]$ as follows: If there is no trade under symmetric uncertainty, then there is no trade under adverse selection. This condition bites only when $c'(0) \geq 1$, since it fails with $c'(0) = 1$. 12

Proposition 1 helps distinguish our explanation for coverage denials from another: regulation of premiums. A conjecture is that an insurer might deny coverage to the worst risks if premiums are capped (see Tennyson (2007) for an overview of insurance regulation). But Proposition 1 shows that there are gains to trade if and only if there are gains to trade with a single, “smal-
l” contract. Even without explicit premium regulation, insurers might voluntarily cap premiums out of fear of a regulatory response, or for reputational reasons. But again, Proposition 1 implies that a high-premium contract is profitable only if a “small” contract is profitable. Hence, explicit regulation of premiums or voluntary premium caps alone cannot explain coverage denials. If a minimum coverage requirement is added to a premium cap, then insurers would sometimes deny coverage even without provision costs. But the cancellation in recent years of private health insurance policies because they did not meet the minimum coverage requirements of the Affordable Care Act points to the existence of low-premium, low-coverage policies in this market.\textsuperscript{13} Whatever the fraction of coverage denials that regulation might account for, our modest and realistic departure from the standard insurance model explains why coverage denials would occur in the complete absence of price or quantity regulation.

5. Trade and pooling

So far we have shown how provision costs can account for coverage denials only for the worst risks. We now turn to how these costs affect menus of contracts when coverage is provided. Three classic properties of contracting menus with private information are that the highest type gets an interim efficient contract (\textit{efficiency at the top}); no type pools with the highest type (\textit{no pooling at the top}); and all other types get coverage smaller than an interim efficient level with the same expected utility (\textit{downward distortions elsewhere}). Chade and Schlee (2012) Theorem 1 shows that they hold under weak assumptions when insurance provision is costless (Hellwig (2010) shows these properties hold in a general private-values principal-agent model). We show that all three properties fail with costly insurance provision: all types can be pooled on the same positive contract and all high-enough types have contracts that are \textit{distorted upwards} from an interim efficient contract.\textsuperscript{14} We illustrated pooling in the two-type CARA case in Section 3. Here we show that complete pooling can happen even with a \textit{continuum} of types.

5.1. Complete information

Under costless provision, the profit-maximizing menu under complete information is simple: each type gets full insurance and is charged a premium that extracts all the surplus. This follows since the marginal cost of coverage is just the loss chance. When the marginal cost of coverage exceeds the loss chance ($C_x(x, \theta) = \theta c'(x) > \theta$), then the complete information menu can be \textit{decreasing} in the loss chance. As we will see, this difference with the standard model without provision costs gives some intuition for the pooling results of the next section.

**Proposition 2** (Complete Information). Let $c$ be $C^2$ and convex with $c'(x) > 1$ for all $x$. Then the complete information menu is decreasing, strictly so on the set of types with positive coverage, under either one of these conditions:

\textsuperscript{13} The National Association of Insurance Commissioners (2011) reports that 31 states in the U.S. merely required that premiums be actuarially justified in the private health insurance market: high premia were allowed for bad risks provided that firms can show that they are indeed bad risks.

\textsuperscript{14} A contract $(x', t') \in \mathcal{C}$ is interim efficient for a type-$\theta$ consumer if it maximizes the insurer’s expected profit $t - C(x, \theta)$ on $\{(x, t) \in \mathcal{C} | U(x, t, \theta) = U(x', t', \theta)\}$, the set of contracts that are indifferent to $(x', t')$ for a type-$\theta$ consumer. A contract $(x'', t'')$ is \textit{distorted downwards} from an efficient contract $(x', t')$ for type $\theta$ if $(x'', t'') < (x', t')$ and $U(x'', t'', \theta) = U(x', t', \theta)$. If the inequality is reversed, then $(x'', t'')$ is \textit{distorted upwards} from an efficient contract.
(i) The consumer’s vN-M utility satisfies increasing absolute risk aversion.

(ii) The number $\eta = \min_{x \in [0,\xi]} c'(x)$ is large enough.

Conditions (i) and (ii) are strong, but each one is merely sufficient.\footnote{One can construct examples with pure multiplicative loading and square root or logarithmic utility such that the complete information menu is not decreasing everywhere, so it is not true that complete-information monopoly coverage is always decreasing in type. And one can also construct examples with these utility functions where complete-information coverage is strictly decreasing in type, so increasing absolute risk aversion in part (i) is not necessary.} We intend the proposition only to demonstrate that complete-information menus can be decreasing. As it turns out, each condition dampens wealth effects as the loss chance increases. When $\theta$ increases, the participation constraint becomes slack if the contract is unchanged. At the new solution, the participation constraint must hold as an equality. This decrease in real wealth raises the demand for insurance if preferences satisfy decreasing absolute risk aversion, potentially overturning the conclusion of Proposition 2.

The importance of Proposition 2 for pooling should be clear, especially under CARA (subsumed by part (i)), where the interim efficient coverage is independent of the level of utility the consumer receives. Indeed, efficiency calls for coverage to be decreasing in type, while incentive compatibility requires coverage to be increasing in the type. Under some conditions spelled out in the next section, this conflicting monotonicity can lead to pooling. As we will see, a similar logic applies beyond CARA.

5.2. Pooling with a continuum of types

Insurance is more complicated than the classic monopoly pricing problem in, for example, Maskin and Riley (1984), for two reasons: since the insurer’s cost depends directly on the consumer’s type, it is a common-values model (they assume private values); and except for CARA, there are wealth effects (they assume quasilinear preferences).

Our next result imposes CARA, implying we can write the consumer’s utility for $(x, t)$ in the quasilinear form $v(x, \theta) - t$, as in Section 3. We can then solve the problem as in standard in monopoly pricing models by maximizing the expectation of the virtual surplus, $v - (v_0/\rho) - \theta c$, with respect to the single variable $x$, subject to $x$ being increasing, where $\rho = f/(1 - F)$ is the hazard rate function. If $\rho$ is increasing in $\theta$, then we say that the monotone hazard rate condition (MHRC) holds.

The next result shows that pooling at the top and even complete pooling can hold in a profit-maximizing menu with a continuum of types. Since everyone could be pooled at zero coverage, we also give conditions for pooling at positive coverage.

**Theorem 2** (Pooling: CARA preferences). Let the consumer’s preferences be CARA, let $c$ be $C^2$ and convex, and let beliefs be a $C^1$ density $f$ on $[\theta_L, \theta_H]$, $\theta_L > 0$.

(i) If $c'(0) > (1 + \theta_H)/2\theta_H$ and $e^{\xi}(1 - c'(0)\theta_H)/(c'(0)(1 - \theta_H)) > 1$, then the profit-maximizing menu exhibits pooling at the top;

(ii) If $(f'/f)/\rho$ and $1/\rho$ are bounded above in $\theta$, and $\eta = \min_{x \in [0,\xi]} c'(x)$ is large enough, then the profit-maximizing menu is pooling.

(iii) If the MHRC holds, $c'(x) > 1$ for every $x$, complete-information profit is positive for $\theta = \theta_L$, and $f(\theta_L)$ is large enough, then every type is pooled on a positive contract.
The proof of part (i) shows that the solution to the relaxed problem—choose coverage type-by-type to maximize virtual surplus, ignoring the constraint that coverage be increasing—is strictly decreasing at $\theta_H$. Hence, the omitted monotonicity constraint on $x$ binds near the top, and this implies pooling at the top by a standard “ironing” argument that accounts for the monotonicity constraint.

Regarding part (ii), the conditions on $f$ and $\rho$ are easy to check. If $f$ is positive and satisfies the MHRC, then $1/\rho$ is uniformly bounded above by 0; and if $f$ is either positive everywhere or decreasing, then $(f'/f)/\rho$ is uniformly bounded above. The proof of (ii) and also of part (i) reveals that the supermodularity of $c$ creates a force in the direction of making the solution of the maximization of virtual surplus to be decreasing in type. Indeed, if $C(x, \theta) = \theta c(x)$ is sufficiently supermodular in $(x, \theta)$ ($\eta$ is large enough), then virtual surplus will be submodular in $(x, \theta)$ on the relevant domain. It follows that the solution to the relaxed problem is globally decreasing. The resulting “ironed” menu which restores the omitted monotonicity constraint implies pooling.

Part (ii) requires that the cost function be sufficiently supermodular, implying that provision costs be sufficiently high. Part (iii) is different: it relaxes the cost condition merely to $c'(x) > 1$, but requires that the insurer’s belief be sufficiently optimistic. The proof does not show that the virtual surplus is supermodular, but just that it satisfies the strict single crossing property in $(x, -\theta)$. Part (iii) demonstrates that a monopolist can pool a continuum of types on a positive contract for arbitrarily small provision costs.

Part of the intuition underlying Theorem 2 is the one hinted at in the previous section. Proposition 2 implies that, under CARA, the complete information coverage is decreasing in type, and strictly so wherever coverage is interior. So, ignoring (IC), the insurer wants to provide less coverage to higher types, but (IC) requires that higher types receive more coverage. This conflicting monotonicity in the insurer’s optimization problem can be so severe to preclude screening of types (either partially or completely). Theorem 2 identifies conditions under which this conflict leads to pooling at the top and complete pooling, respectively.\(^{16}\)

Note that the insurance problem with CARA is a special case of the principal-agent problem in Guesnerie and Laffont (1984, Section 5.1), and in particular their Theorem 4 applies. It implies that the profit-maximizing coverage coincides with the one that solves the relaxed problem except on a finite number of disjoint intervals where there is pooling (their assumption that the solution to the relaxed problem has a finite number of “peaks” can be ensured with a suitable class of densities $f$ and thus of hazard rates $\rho$). What is surprising about Theorem 2 is that one of the pooling regions can occur at the top, and that one can easily identify conditions on costs and beliefs that lead to complete pooling.

The CARA preference assumption greatly simplifies the problem since it rules out wealth effects. It is clear that complete pooling does not depend on CARA, but the general case is more difficult. Our next result relaxes CARA, but imposes stronger conditions on costs and beliefs. We restrict cost functions to pure multiplicative loading (Example 1 with $k = 0$).

**Theorem 3 (Pooling: Pure Multiplicative Loading).** Let beliefs be a $C^1$ density $f$ on $[\theta_L, \theta_H]$, $\theta_L > 0$, and $C(x, \theta) = \lambda \theta x$ for all $(x, \theta)$, with $\lambda < 1/\theta_H$.\(^{16}\)

\(^{16}\) The effects of this conflict on the solution to the principal’s problem was first studied by Guesnerie and Laffont (1984), and it is known as nonresponsiveness. Some references define this term as the case where full-information and incentive compatible allocations have opposite monotonicity (e.g., Laffont and Martimort (2002), Morand and Thomas (2003)), while others refer to nonresponsiveness in the weaker sense that the full-information allocation is not implementable, that is, fails to be monotone in the direction implied by incentive compatibility (e.g., Caillaud et al. (1988)).
(i) If $\lambda > (1 + \theta_H)/2\theta_H$ and complete information coverage is positive at $\theta_H$, then the profit-maximizing menu exhibits pooling at the top;

(ii) If $\lambda > (1 + \theta_L)/2\theta_L$ and $f'(\theta)/f(\theta) \leq (3\theta - 1)/(\theta(1 - \theta))$ for every $\theta \in [\theta_L, \theta_H]$, then the profit-maximizing menu exhibits complete pooling.

Part (i) of the theorem is the analogue of part (i) of Theorem 2. It imposes conditions only on the loading factor $\lambda$ and the highest element in the belief support, plus a condition on the complete information coverage for the highest type that is easily justified from primitives. Moreover, since $1/\theta_H > \lambda > (1 + \theta_H)/2\theta_H$, the result holds with $\lambda \theta < 1$ for all $\theta$. The logic is similar to that of Theorem 2 (i): the solution to the relaxed problem is strictly decreasing near $\theta_H$, and thus “ironing” yields pooling at the top. Part (ii) is the analogue of the complete pooling results in Theorem 2. It imposes a stronger condition on $\lambda$ (since $(1 + \theta)/\theta$ is decreasing in $\theta$), and also restricts the density of types. But it delivers complete pooling under loading for any twice continuously differentiable and strictly concave utility function. The logic is similar to that of Theorem 2 (ii): the solution to the relaxed problem is strictly decreasing, and complete pooling ensues at the profit maximization menu. Both conditions require that $\lambda$ be large enough, which implies that the cost function is sufficiently supermodular, as in Theorem 2.

The intuition is similar to that for the CARA case. It is, however, complicated by the presence of wealth effects, which makes the profit-maximizing coverage both under complete and incomplete information depend on the amount of surplus/utility the consumer is given. But we show in Appendix A.7 that, suitably modified, the insurer’s problem exhibits the conflicting monotonicity alluded to above. Under the conditions of Theorem 3 this conflict is severe enough to render sorting of types suboptimal (either at the top or on the entire type support).

Theorem 3 gives conditions for complete pooling with a continuum of types, but is silent about whether it is on a zero contract or a positive one. Just as an illustration, we next present an example of pooling at the top and complete pooling on a positive contract for the special case of pure multiplicative loading and a uniform type distribution.

Example 2 (Pooling at the Top; Complete Pooling). Let $f$ be uniform on $[\theta_L, \theta_H]$, with $1/3 \leq \theta_L < \theta_H$. Since $(1 + \theta_H)/2\theta_H < (1 + (1/3))/2(1/3) = 2$, it follows that if $2 < \lambda < 1/\theta_H$, then the profit-maximizing menu pools at the top by Theorem 3 (i). And since $f' = 0$ and $\theta_L \geq 1/3$, the condition on $f'/f$ in Theorem 3 (ii) holds. Hence, the profit-maximizing menu exhibits complete pooling. It remains to show that the single contract can be positive.

As risk aversion increases uniformly without bound, the no-trade condition in Proposition 1, $MRS(\hat{\theta}) \leq \lambda \mathbb{E}[\theta|\theta \geq \hat{\theta}] = \lambda(\theta_H + \hat{\theta})/2$, fails (and hence there is trade) for the uniform distribution if $1 > \lambda(\theta_H + \theta_L)/2$. This is because $MRS(\hat{\theta}) \rightarrow 1$ for all $\hat{\theta}$ as risk aversion increases uniformly without a bound. With $\theta_L = 1/3$ the condition $1 > \lambda(\theta_H + \theta_L)/2$ is compatible with $\lambda > 2$ if $\theta_H < 2/3$.

Turning to the classic contracting properties, Example 2 and Theorem 2 (i) already illustrate the failure of no-pooling-at-the-top. Consider the other two properties: efficiency at the top, and

---

17 The proof of the theorem is more involved because wealth effects force us to use optimal control arguments. In the CARA case without wealth effects, we can write the agent’s utility as linear in the transfer. We then use the usual trick of integrating by parts to eliminate the transfer and simply maximize the resulting virtual surplus pointwise with respect to $x$ for each type. With wealth effects we cannot use this trick to bypass optimal control arguments.
downward distortions elsewhere. For definiteness, assume CARA and the conditions in Theorem 2 (ii), and assume also that all types are pooled at a positive coverage $\hat{x} \in (0, \ell)$. If the full-information coverage of $\theta_H$ were zero, then clearly both properties would fail near $\theta_H$. So assume that the full information coverage of $\theta_H$ is positive, and denote it by $x^*(\theta_H)$, which solves $v_x(x^*(\theta_H), \theta_H) - \theta_H c'(x^*(\theta_H)) = 0$. We know from Theorem 4 in Guesnerie and Laffont (1984) that $\hat{x} \geq x^*(\theta_H)$, and also that one of the optimality conditions that $\hat{x}$ must satisfy is

$$\int_{\theta_L}^{\theta_H} \left( v_x(\hat{x}, \theta) - \theta c'(\hat{x}) - \frac{v_{\theta x}(\hat{x}, \theta)}{\rho(\theta)} \right) f(\theta) d\theta = 0. \quad (5)$$

If $\hat{x} > x^*(\theta_H)$ then it is clear that efficiency at the top and downward distortions elsewhere both fail, so it suffices to rule out $\hat{x} = x^*(\theta_H)$. Assume to the contrary that $\hat{x} = x^*(\theta_H)$, and thus $v_x(\hat{x}, \theta_H) - \theta_H c'(\hat{x}) = 0$. Since under the conditions of Theorem 2 (ii) the term in parentheses in (5) is strictly decreasing in $\theta$ and is zero at $\theta = \theta_H$, it follows that it is strictly positive on $[\theta_L, \theta_H]$ and hence (5) is violated, a contradiction.

Theorems 2 and 3 impose smooth cost. We now extend them to the case of a fixed claims cost. The cost discontinuity however changes the conclusion: profit is maximized with a menu with at most two contracts, the no-trade contract and a positive contract, but now each might be chosen by a positive measure of consumers. These menus are what Braun et al. (2019) call “choice” menus. Their empirical model has only two risk types. Corollary 1 shows that we can generate “choice” menus with a continuum of types.

**Corollary 1 (Fixed Claims Cost).** Suppose that, for $k > 0$, the cost function is $c(x) = \theta (c_0(x) + k)$ for $x > 0$ with $c(0) = 0$ and $c_0$ is differentiable and strictly increasing; and let the insurer belief be given by a $C^1$ density $f$ on $[\theta_L, \theta_H]$, $\theta_L > 0$. Assume that either

(i) The belief conditions of Theorem 2 (ii) hold and $\eta_0 = \min_{\ell \in [0, \ell]} c'_0(x)$ is large enough; or

(ii) The complete-information contract is positive for type $\theta_L$, $c'_0(x) > 1$ for every $x \in [0, \ell]$, and $f(\theta_L)$ is large enough; or

(iii) $c_0(x) = \lambda x$, and the belief/cost conditions of Theorem 3 (ii) hold when the cost is $C(x, \theta) = \theta c_0(x)$.

Then profit is maximized with at most two contracts, one positive, and one zero.

If the virtual surplus satisfies the strict single crossing property in $(x, -\theta)$ when the cost function is $c_0$, then it continues to do so if the cost function is $c$, including the fixed claims cost. We do not get complete pooling because of the cost discontinuity: the ironing arguments we use in Theorems 2 and 3 break down because the virtual surplus is not even quasiconcave in coverage. But since the discontinuity is just at $x = 0$, we can show that the profit maximizing menu consists of at most two distinct pooling intervals.

### 5.3. Insurance as a lemons market

We have answered two questions. How do insurance provision costs affect who trades and who doesn’t? And how do these costs affect profit-maximizing menus when there is trade? We now bring together our answers to these two questions and set out some implications of our answers for applied work.

We motivated Theorem 1 by supposing that the two beliefs were two possible posterior beliefs of the insurer after observing a signal about the consumer’s type. To formalize this motivation,
Corollary 2 (Insurance as a Lemons Market). Suppose that the conditions of either Theorem 2 (a) or Theorem 3 (ii) hold for every belief in \( \{F_1, ..., F_n\} \), ordered by LR. Then there is an \( 1 \leq i^* \leq n \) such that, i) for \( i < i^* \), coverage is denied; and ii) for \( i \geq i^* \), every consumer type is pooled on a positive contract.

Part i) and the fact that all other types get positive coverage follows from Theorem 1, and Part ii) follows from Theorem 2 (a) and Theorem 3 (ii). If the set of beliefs has common support, then the conclusion holds if simply the mean type increases in \( i \); if all types are pooled at \((\hat{x}, \hat{\theta})\) then expected profit is \( \hat{\theta} - \mathbb{E}_x[\theta]c(\hat{x}) \), which decreases with an increase in \( E_1[\theta] \). A special case is that \( i \) orders the beliefs by FOSD.

Empirical research into the presence and welfare consequences of adverse selection typically looks at insurance markets that are actively traded. The literature emphasizes the positive correlation property that people who buy more coverage have on average higher claims, and the inefficiency that results from some types getting interim inefficient contracts. Indeed, this correlation is the basis of a standard statistical test for adverse selection. Our results, however, suggest that adverse selection in this sense might be invisible: under the conditions of Theorem 2 (a) and Theorem 3 (ii), either the market shuts down altogether, or everyone is pooled on the same contract. These results also suggest that the main welfare cost of adverse selection is not misallocation in active markets, but that some markets simply do not exist. This suggestion is consistent with Hendren’s (2013) empirical findings, and with the notion that provision costs make insurance markets with adverse selection look more like Akerlof’s lemons model. In this sense, provision costs “lemonize” monopoly insurance markets.\(^{18}\)

6. Beyond monopoly

We now examine the robustness of the insights beyond monopoly. We first extend the perfectly competitive model of Rothschild and Stiglitz (1976) to provision costs, and show that our monopoly pooling results do not hold in this model. We then outline a duopoly model that preserves our monopoly pooling results.

\(^{18}\) Consider the famous example in Section B of Akerlof (1970). As one varies the multiplicative taste parameter of the group-two traders (those with a zero endowment of cars), either there is no trade, or there is a unique equilibrium with positive trade. If the type-two traders are on the “long end” of the market (their aggregate income is high enough), then all cars are traded whenever there is trade.
6.1. Perfect competition

Since provision costs dramatically affect monopoly insurance, it is natural to ask how they affect competitive insurance contracts. A competitive equilibrium in Rothschild and Stiglitz (1976) is a set of contracts such that, when consumers choose contracts to maximize expected utility, no contract in that set makes negative profit; and, given this set, no contracts outside the set would earn positive profit if offered. Its more modern rendition consists of at least two identical firms competing in Bertrand fashion, where each firm posts a finite number of contracts and then consumers choose one contract or none at all. Here there is a unique subgame perfect equilibrium allocation of the two-stage game that equals the Rothschild-Stiglitz allocation (see Mas-Colell et al. (1995), chapter 13).

In either version, the textbook two-type case shows that when an equilibrium exists (i) there is no pooling; (ii) the high type gets an interim efficient contract (full insurance with no provision costs); (iii) the low risk consumer’s contract is distorted downwards from efficiency. These properties straightforwardly extend to any finite number of types—if in (iii), the word “low” is replaced by “each lower” (see Appendix A.9).

Properties (i)–(ii) follow from these facts: consumer preferences satisfy the strict single-crossing property; at any given contract expected profit is lower for higher types; and profit is continuous in type and coverage. These properties continue to hold for our cost specification if cost is continuous as, for example, in the pure multiplicative loading case. These properties can be extended to any equilibrium with trade for the case of a fixed claims cost in our Example 1, since the only discontinuity is at zero coverage. So (i)–(ii) continue to hold for a competitive equilibrium with these provision costs (see Appendix A.9). Property (iii) continues to hold if complete information coverage is decreasing in type (including the nonresponsive case).

As we have seen, each of these properties can fail in our monopoly insurance model. The stark difference between how provision costs affect on monopoly and competitive insurance is potentially important for empirical work on the subject. For example, one interesting implication of these facts is a way to distinguish empirically between competition and monopoly under adverse selection when insurance provision is costly: observing a menu with just one nonzero contract (pooling) is consistent with monopoly, but not with competition under adverse selection.

There are also implications for coverage denials. Suppose that our cost function $C(x, \theta) = \theta c(x)$ is both convex and continuous in $x$, and that the number of types is finite. Then the monopolist’s solution is unique: in particular, profit is zero if and only if the no-trade menu is the unique solution. Suppose that there are no gains to trade in the sense that the maximum monopoly profit is zero. There are then two possibilities for a competitive equilibrium: one does not exist; or only a no-trade equilibrium exists.

Putting aside the equilibrium existence issue, the important question is this: if there is an equilibrium, what determines whether it involves trade or not? Property (ii) implies that, if a competitive equilibrium exists, there are nonzero contracts if and only if the complete information (zero-profit) contract for the highest risk type is nonzero. Put another way: if the support of

---

19 The elimination of pooling proceeds as in Rothschild and Stiglitz (1976). Similarly with the derivation of the only separating equilibrium candidate contracts: for example, under pure loading, they solve $t_i = \lambda \theta_i x_i$, $i = L, H$ (zero profit of each contract), $u(\omega - t_H) = U(x_L, t_L, \theta_H)$ (incentive constraint of high type binds), and $(1 - \lambda \theta_H)u'(\omega - \ell + x_H - t_H) = \lambda (1 - \theta_H)u'(\omega - t_H)$ (efficient high type contract).

20 For a two-type model with loading, Dionne et al. (1999), Proposition 2, confirms property (i) and Ramsay et al. (2013), Result 4.1, confirm properties (i) and (ii).
insurers’ beliefs remains the same, changes in insurer beliefs are irrelevant for whether or not there are gains to trade under competition.

This fact has an important implication for testing for adverse selection in the sense of shutting down markets, as in Hendren (2013). Under competition, if type distributions (beliefs) all have the same support, then there is no relationship between type distributions and coverage denials. Correlation between the type distribution and coverage denials only emerge in market structures other than (Rothschild-Stiglitz) competition.21

6.2. Oligopoly

A stumbling block in extending our pooling results extend beyond monopoly is that there is no consensus on an appropriate model of oligopoly in industrial organization in general or insurance in particular. Indeed, the literature on oligopolistic screening under incomplete information is slim (see Chade and Swinkels (2018) for a recent contribution under private values and the literature they cite).

Here we sketch a plausible duopoly model that builds on the screening model with random participation of Rochet and Stole (2002). The market consists of two firms, each with a loading factor, and a continuum of consumers of different types. Consumers have idiosyncratic preferences for each of the two firms, modeled as an additive (independent and identically distributed) shock (one for each of the firms) to the utility they obtain from a contract offered by a firm. These shocks are unobservable and not contractible to the firms. Note that the set up includes both demand and cost differentiation. Regarding the timing of the interaction, first the two firms simultaneously offer menus of contracts, consisting on a coverage and a premium function dependent on the reported type. Then consumers observe the offers and their idiosyncratic shocks, and choose the best contract in the menus or the autarky option.

We have confirmed that in any pure strategy Nash equilibrium of the game between the firms in which at least one firm offers a non-null menu, there is pooling of an interval of types, and that if the loading factor is large enough, then there is pooling at the top. We leave the proof of this result, and a more complete analysis of pooling everywhere, for future work.22

7. Beyond insurance

Our pooling results and examples directly exploit the specifics of the insurance model. Yet we conjecture that versions of these results extend to other economic applications of common-values principal-agent models. This is clearest in the CARA set-up of Theorem 2, which reduces to a standard quasilinear screening problem with a particular function \( v \). If one starts with \( v \) as a primitive, as is standard in applications, and assumes it is \( C^3 \) and satisfies the SCP in \( (x, \theta) \), then a similar pooling result should follow.

Regarding coverage denials, the weak cost function conditions of Theorem 1 suggest that it is immediately applicable beyond insurance. Indeed, the main tool we use to prove the theo-

21 Hendren (2013) only considers continuous type distributions. Here we view the continuous distribution as an approximation to a finite set of types, since a Rothschild-Stiglitz competitive equilibrium does not exist with a continuum of types.
22 Another promising approach for extending our pooling results beyond pure monopoly is a regulated monopoly, in which a regulator with the same information as the firm chooses a menu of contracts to maximize a weighted average of consumers’ surplus and expected profit. Our preliminary results suggest it is easier to show pooling on a positive contract for the regulator’s problem than for pure monopoly.
rem, Lemma 1 (Section A.2), is formulated for a general optimization problem (see (8) of that section) and imposes only a weak condition that holds in many principal-agent models, such as the common-values modification of Maskin-Riley in Toikka (2011); a monopsony procurement problem in which the single buyer is unsure about the supplier’s product quality;23 a monopoly version of the rat-race employment model of Miyazaki (1977); and a monopoly version of the Stiglitz and Weiss (1981) model of credit markets.24 The last two are of obvious interest, since they suggest asymmetric-information theories of unemployment and loan denials or redlining.

Although Theorem 1 and Lemma 1 are formulated for a monopoly, we repeat that these results have implications for other market structures: since a monopoly is the most profitable market structure, if there are no gains to trade for a monopoly, then there are no gains to trade from any other market structure. This implication holds for all the applications just mentioned.

Appendix A

A.1. Omitted algebra from two-type example

Let \( v \) be the multiplier on the monotonicity constraint, \( x_H \geq x_L \). Use (2) to solve for \( t_L \) and \( t_H \), substitute them into the objective function, then differentiate with respect to \( x_L \) and \( x_H \) to find the first order conditions:

\[
(1 - p)(v(x_L, \theta_L) - \theta_L \lambda) - v \leq p(v(x_L, \theta_H) - v(x_L, \theta_L)),
\]

(6)

\[
pv(x_H, \theta_H) + v \leq p\theta_H \lambda,
\]

(7)

together with (2), \( x_H \geq x_L \geq 0 \), and complementary slackness for these nonnegativity and monotonicity constraints. Since \( v(\cdot, \theta) \) is concave, these first-order conditions are both necessary and sufficient for a point to solve the problem (this is easiest to see by changing variables from the space of contracts to the space of state-contingent utilities, which makes constraints linear and the insurer’s objective function concave).

Let \( \mathbb{E}[\theta] = p\theta_H + (1 - p)\theta_L \). To show that (3) is necessary for no trade, suppose that \( x_L = x_H = 0 \) solves the first-order conditions, and so maximizes insurer profit. Then (7) implies that \( \lambda \theta_H \geq v(x_H, \theta_H) \) since \( v \geq 0 \); substitute the inequality \( v \leq p\theta_H \lambda v(x_L, \theta_H) \) into (6) and rearrange to find \( \lambda \mathbb{E}[\theta] \geq v(x_L, \theta_L) \), so (3) holds. To show that (3) is sufficient for no trade, suppose some \( (x_H, x_L) > 0 \) solves the first-order conditions, so \( x_H > 0 \) and (7) holds as an equality. We argue that (3) fails. If \( v > 0 \), then \( \lambda = v(x_H, \theta_H)/\theta_H < v(x_L, \theta_H)/\theta_L \), since \( x_H > 0 \) and \( v(\cdot, \theta_H) \) is strictly concave, so (3) fails. If \( v < 0 \), then \( x_H = x_L \Rightarrow \beta > 0 \). Solve (7) for \( v \), insert into (6) to find that \( \lambda = v(x_H, \theta_H)/\mathbb{E}[\theta] < v(x_L, \theta_L)/\mathbb{E}[\theta] \) and again (3) fails. Since \( v(x_L, \theta) \rightarrow \theta \) as risk aversion \( r \rightarrow 0 \), (3) can hold for any \( \lambda > 1 \), provided risk aversion is low enough; in particular it can hold for \( \lambda > 1 \) satisfying \( \lambda \theta_H < 1 \).

---

23 Ilya Segal suggested this application to us in the context of Chade and Schlee (2012).

24 A project either succeeds and pays \( t > 0 \) or fails and pays 0. A risk averse borrower can either self-finance the project or take out a loan of size \( B \) from a monopoly bank. A loan contract specifies the repayment \( t \) if the project succeeds and collateral \( t - x \) collected if it fails (default). The borrower, but not the bank, knows the failure chance \( \theta \). If the borrower’s initial wealth is \( w - t \), then the borrower’s random wealth with contract \( (x, t) \) is exactly as in our insurance model. Since any borrower rejects any contract with \( x = 0 \) and \( t > 0 \), we identify \( x = 0 \) with a no-trade contract. Suppose the bank incurs a fixed cost \( k \) if and only if \( x > 0 \) and the borrower defaults on an accepted contract. Then the bank’s expected profit is the same as in our insurance model with cost function from Example 1 for \( \lambda = 1 \) and \( k > 0 \). More general versions of this model fit the weaker conditions of Lemma 1 in the Appendix, which do not require that the payoffs be the same as our insurance model.
To show that the profit maximizing menu can entail pooling at a positive contract, suppose that \( 1 < \lambda < v_\ell(0, \theta_L) / \theta_L \), so the complete-information contract for the low type is positive; since \( v_\ell(0, \theta_H) / \theta_H > v_\ell(0, \theta_L) / \theta_L \), it follows that (3) fails at \( p = 0 \), so it fails for all \( p \in (0, \hat{p}) \) for some \( \hat{p} < 1 \). For any such \( p \), \( x_H > 0 \) at any profit-maximizing menu (since \( x_H \geq x_L \geq 0 \), so (7) holds as an equality. If \( v = 0 \) for some \( p \in (0, \hat{p}) \), then by (7), we have \( x_H = \ell - \frac{1}{2} \ln \left( \frac{(1-\theta_H)\lambda}{1-\theta_H} \right) =: x^*(\theta_H) \). Direct substitution reveals that \( v_\ell(x^*(\theta_H), \theta_L) - \lambda \theta_H > 0 \). Using \( x_L \leq x_H = x^*(\theta_H) \) and the concavity of \( v(\cdot, \theta_L) \), rearrange inequality (6) to find \( 0 < v_\ell(x^*(\theta_H), \theta_L) - \lambda \theta_H \leq v_\ell(x_L, \theta_L) - \lambda \theta_L \leq [v_\ell(x_L, \theta_H) - v_\ell(x_L, \theta_L)](1 - p)/p \). Since \( v_\ell(x_L, \theta_H) - v_\ell(x_L, \theta_L) > 0 \) and the number \( x^*(\theta_H) \) does not depend on \( p \), these inequalities imply that there is a positive lower bound for \( p \), call it \( p^* \), such that \( v = 0 \) implies \( p \geq p^* \). Equivalently, for any \( 0 < p < \max(\hat{p}, p^*) \), \( v > 0 \) and the insurer pools both types on a positive contract.

A.2. Proof of Theorem 1

The result is easier to prove after a change of variables: rather than choosing a menu of feasible contracts, the firm chooses a menu of feasible expected profits. For any belief \( F \) and menu \( \{ x(\cdot, t(\cdot)) \} \) from \( \Theta_F \) to \( C \) that satisfies (IC) and (P), there is a real-valued function \( \pi \) on \( \Theta_F \) that gives the expected profit for each type \( \theta \), namely, \( \pi(\theta) = t(\theta) - C(x(\theta), \theta) \). Let \( \Phi_F \) be the set of such functions.

The next assumption says that if a menu of expected profits is feasible, then so is a menu that sets expected profit of all types below some threshold type equal to zero, and does not lower aggregate expected profit from the other types.

Assumption 1. Let \( F \) be the distribution of types. If \( \pi(\cdot) \in \Phi_F \), then for any \( \theta' \in \Theta_F \), there is a \( \pi'(\cdot) \in \Phi_F \) with \( \pi'(\theta) = 0 \) for \( \theta < \theta' \) and \( \int_{[\theta', 1]} \pi'(r)dF(r) \geq \int_{[\theta', 1]} \pi(r)dF(r) \).

We confirm later that this Assumption holds in our insurance model. For now we simply impose it. Now consider the problem of choosing a measurable \( \pi(\cdot) \) to solve

\[
V(F) = \max_{\pi(\cdot) \in \Phi_F} \int_{\Theta_F} \pi(\theta)dF(\theta)
\] (8)

Lemma 1. Let \( G \) LR dominate \( F \) and have the same support and suppose that Assumption 1 holds. If \( V(F) = 0 \), then \( V(G) = 0 \).

Proof of Lemma 1. We give the proof for the case of continuous densities with respect to the Lebesgue measure. The proof for the finite-type case just replaces integration-by-parts with summation-by-parts. Let \( \theta_L \) be the smallest and \( \theta_H \) the largest element in the common support \( \Theta_F \). Let \( g \) be the density of \( G \) and \( f \) the density of \( F \). From the definition of LR dominance, \( f/g \) is decreasing. Set \( h = f/g \).

We want to show that \( V(F) = 0 \) implies that \( V(G) = 0 \). We prove the contrapositive. Suppose that \( V(G) > 0 \), and let \( \pi^G \) maximize expected profit on \( \Phi \) at belief \( G \). It suffices to show that \( \int_{\Theta_F} \pi^G(q)g(q)dq \) is absolutely continuous, we can integrate by parts to find
\[
\int_{\theta_L}^{\theta_H} \pi^G(\theta) f(\theta) \, d\theta = \int_{\theta_L}^{\theta_H} \pi^G(\theta) h(\theta) g(\theta) \, d\theta \\
= h(\theta_H) \int_{\theta_L}^{\theta_H} \pi^G(\theta) dG(\theta) + \int_{\theta_L}^{\theta_H} \left[ \int_{\theta_L}^{\theta} \pi^G(q) dG(q) \right] d(-h(\theta)). \quad (9)
\]

Consider the two terms in (9). The first term is nonnegative by hypothesis. Assumption 1 implies that cumulative profit is always nonnegative\(^{25}\):

\[
\int_{\theta_L}^{\theta} \pi^G(q) dG(q) \geq 0 \text{ for every } \theta \in [\theta_L, \theta_H]. \quad (10)
\]

Since \(h\) is decreasing, it follows that the second term in (9) is nonnegative. We are done if at least one of the terms in (9) is positive. There are two possibilities. First, \(h(\theta_H) > 0\), in which case the first term in (9) is positive. Second, \(h(\theta_H) = 0\). Since \(\theta_H\) is the largest element of the common support, \(\int_{\theta_H-\varepsilon}^{\theta_H} d(-h(\theta)) > 0\) for all \(\varepsilon > 0\). Since \(\int_{\theta_H}^{\theta_H} \pi^G(q) dG(q) \geq 0\) with a strict inequality at \(\theta = \theta_H\), and the integral is continuous at \(\theta_H\), it follows that the second term in (9) is positive. \(\square\)

**Proof of Theorem 1.** We first confirm that Assumption 1 holds in our insurance model. Consider a belief \(F\), let \(\{x(\cdot), t(\cdot)\}_{\theta \in \Theta_F}\) be a feasible menu of contracts and suppose that \(\theta' \in \Theta_F\). If \(\theta'\) is the smallest element of the support, the conclusion holds trivially, so suppose that \(\theta'\) is not the smallest element. We now define a new menu of contracts by giving every type less than or equal to \(\theta'\) the no-trade contract, and leaving the other contracts the same. Since \(C(0, \theta) = 0\), expected profit from each of these no-trade contracts is zero. There are two possibilities. First, type \(\theta'\) weakly prefers \((0, 0)\) to any contract \((x(\theta), t(\theta))\) for \(\theta > \theta'\). In this case the new menu is feasible: since the consumer’s preferences satisfy the single-crossing property, no type lower than \(\theta'\) strictly prefers \((0, 0)\) to \((x(\theta), t(\theta))\) for \(\theta > \theta'\). Clearly expected profit to the higher types is unchanged, and the conclusion follows.

The second case is that type \(\theta'\) strictly prefers some higher type’s contract to \((0, 0)\). In this case, raise the premium and change the coverage for each original contract for each type \(\theta \geq \theta'\) (including type \(\theta'\)) so that utility falls by the same amount in each state until type \(\theta'\) is indifferent between \((0, 0)\) and this less-attractive version of its original contract. Each type \(\theta \leq \theta'\) now weakly prefers \((0, 0)\) to any contract given to the higher types and the new menu is feasible. Since wealth in each state fell for each of the higher types, expected profit conditional on \(\theta > \theta'\) is higher. So Assumption 1 holds, and the conclusion of Theorem 1 follows whenever \(F\) and \(G\) have the same support.

Now suppose that \(F\) and \(G\) do not have the same support. By the definition of LR dominance, the smallest elements and largest elements of the two supports are ordered, with the extreme elements of \(G\)’s support being higher. There are two possibilities. First, \(\int_{\Theta_F \cap \Theta_G} \pi^G(\theta) dG(\theta) > 0\), that is, the profit-maximizing menu for \(G\) earns positive profit when restricted to the support.

\(^{25}\) If for any \(\pi \in \Phi\) there is a \(\theta' \in [\theta_L, \theta_H]\) with \(\int_{\theta_L}^{\theta'} \pi(q) dG(q) < 0\), then by Assumption 1 there is a \(\pi' \in \Phi\) with \(\pi'(\theta) = 0\) for \(\theta < \theta'\) and \(\int_{\theta}^{\theta_H} \pi' dG \geq \int_{\theta}^{\theta_H} \pi dG\), so \(\pi\) does not solve the problem at \(G\).
of $F$. Since $V(F) \geq \int_{\Theta_F \cap \Theta_G} \pi^G(\theta)dF(\theta)$, it is enough to show that the last integral is positive, which follows from the argument in the proof of Lemma 1 (simply extend $\pi^G(\cdot)$ to all of $\Theta_F$ by setting $\pi^G(\theta) = 0$ for $\theta \in \Theta_F - \Theta_G$, and proceed as in Lemma 1).

The second possibility is that $\int_{\Theta_F \cap \Theta_G} \pi^G(\theta)dG(\theta) \leq 0$, in which case complete information profit must be positive for some type in the support of $G$ but not in the support of $F$. We will show that this fact implies $V(F) > 0$.

The complete-information, profit-maximizing objective is $T(x, \theta) - C(x, \theta)$ for $x > 0$ (and 0 otherwise), where $T(\cdot)$ is defined implicitly by

$$U(x, T(x, \theta), \theta) = U(0, 0, 0).$$

Suppose now complete-information profit is positive for some $\theta' \in \Theta_G - \Theta_F$, so $T(x', \theta') - C(x', \theta') > 0$ for some $x' > 0$ or, equivalently, $[T(x', \theta') - C(x', \theta')]/\theta' > 0$. We begin by showing that complete-information profit is positive for every lower type. Since $C(x', \theta)/\theta$ is increasing in $\theta$, we are done if $T(x', \theta)/\theta$ is decreasing in $\theta$. We will show the stronger property that $T(x, \cdot)$ is concave for every $x$ (it is stronger since $T(x, 0) = 0$). Differentiate (11) with respect to $\theta$ at $x = x'$ to find that $U_1T_\theta + U_\theta = U_\theta^0$, where $U_i$ and $U_\theta$ are partial derivatives evaluated at $(x', T(x', \theta), \theta)$ and $U_\theta^0$ is $U_\theta(0, 0, \theta)$. Differentiate again and use $U_\theta U_\theta = 0$ to find that $U_1T_\theta + U_\theta^0 T_\theta^2 + 2U_\theta T_\theta = 0$. Recalling that $U(x, t, \theta) = \theta u(w - \ell - t + x) + (1 - \theta)u(w - t)$ and that $u$ is strictly increasing and strictly concave, one can easily verify that $U_1 < 0$, $U_\theta < 0$, $U_\theta^0 < 0$, and $T_\theta > 0$, so that $T_\theta^0 \leq 0$. It follows that complete-information profit is positive at any $\theta < \theta'$.

In particular, complete-information profit is positive at the largest element in the support of $F$, call it $\theta_H$. Let $(t_H, x_H)$ a complete-information profit maximizing contract for type $\theta_H$. Now use the argument for Theorem 1-(vii) in Chade and Schlee (2012) to construct a two-contract menu, one the no-trade contract $(0, 0)$ and the other equal to $(t_H - \varepsilon, x_H)$ for $\varepsilon > 0$. If $\varepsilon$ is small enough a positive measure of types will buy the positive contract; and since $C(x, \cdot)$ is increasing, expected profit from that two contract menu is positive at $F$, so $V(F) > 0$. The conclusion of Theorem 1 follows.

Theorem 1 assumes the LR order. We now present an example that shows that LR cannot be replaced by the weaker FOSD order.

**Example 3 (FOSD cannot replace LR in Theorem 1).** Let $G$ have support $[\theta_L, \theta_H]$, $0 < \theta_L < \theta_H < 1$, and $F$ have support $[\theta'_L, \theta_H]$, $0 \leq \theta'_L < \theta_L$. Assume that $g(\theta_H) = f(\theta'_L) = 1 - p$ and $g(\theta_H) = f(\theta'_L) = p$, $p \in (0, 1)$. It is clear that $G$ dominates $F$ in the FOSD order. But since $g(\theta'_L)f(\theta_H) = 0 < p < (1 - p) = p$ and $g(\theta_H)g(\theta_H)$, it follows that $G$ does not LR-dominate $F$.

Now consider the two-type CARA example from Section 3. Set $\lambda$ equal to the supremum of loading factors for which complete-information profit for $\theta_H$ is positive. The complete-information profit for type $\theta_H$ is then zero. By the argument given after equation (11) in the proof of Theorem 1, complete-information profit from every type in $(0, \theta_H)$ is positive. Consider any $\theta_L \in (0, \theta_H)$, and let $G$ be the resulting distribution. For a large-enough fraction of low types, the profit-maximizing menu for this type distribution is pooling on a positive contract and expected profit is positive, as in Section 3. Fix such a fraction of low types, and consider the distribution $F$ with $\theta'_L = 0$. As mentioned, $G$ FOSD dominates $F$, but does not LR dominate $F$. The profit-maximizing menu for the type distribution $F$ is pooling on the no-trade contract.
A.3. Proof of Proposition 1

(i) To prove that (4) is sufficient for no trade, assume that it holds for all types. Suppose first that $\Theta_F = \{\theta_1, \ldots, \theta_n\}$ with $\theta_n > \ldots > \theta_1$. Recall that with finite types, we can reduce the insurer’s problem to one of maximizing expected profit subject to $x_1 \leq x_2 \leq \ldots \leq x_n$, the binding participation constraint of the lowest type, and the binding local downward incentive compatibility constraints. We split the problem into a series of programs that can be solved recursively starting from $\theta_n$.

We first show by induction that any solution to the monopolist problem involves pooling, namely, $x_1 = x_2 = \ldots = x_n$ and $t_1 = t_2 = \ldots = t_n$. Fix $\{(x_1, t_1), \ldots, (x_{n-1}, t_{n-1})\}$ with $x_i \geq t_i \geq 0$ for $i = 1, \ldots, n - 1$. Consider the problem of choosing $(x_n, t_n)$ to maximize $t_n - \theta_n c(x_n)$ subject to the constraints that $x_n \geq x_{n-1}$ and $U(x_n, t_n, \theta_n) = U(x_{n-1}, t_{n-1}, \theta_n)$. By (4), the convexity of $c$, the strict concavity of $u$, and $x_{n-1} \geq t_{n-1} \geq 0$ it follows that (recall that $MRS(x, t, \theta) = -U_x/U_t$ evaluated at $(x, t, \theta)$.)

$$\theta_n c'(x_{n-1}) \geq \theta_n c'(0) \geq MRS(0, 0, \theta_n) \geq MRS(x_{n-1}, t_{n-1}, \theta_n).$$

Let $\delta = MRS(x_{n-1}, t_{n-1}, \theta_n)$. Now consider any $(x, t)$ satisfying the constraint for the insurer’s problem for type $n$. Since $U(\cdot, \cdot, \cdot)$ is strictly concave for every $\theta$, and $(x, t) \geq (x_{n-1}, t_{n-1})$ it follows that $t - t_{n-1} \leq \delta (x - x_{n-1})$. Use the inequality $\theta_n c'(x) \geq \delta$ and convexity of $c(\cdot)$ to obtain $t - t_{n-1} \leq \delta (x - x_{n-1}) \leq \theta_n c(x_{n-1}) - \theta_n c(x)$. Rearranging yields $t_{n-1} - \theta_n c(x_{n-1}) \geq t - \theta_n c(x)$ so $x_n = x_{n-1}$ and $t_n = t_{n-1}$ solves the problem.

Now fix $\{(x_1, t_1), \ldots, (x_{n-k}, t_{n-k})\}$ nonnegative with $x_i \geq t_i$ for $i = 1, \ldots, n - k$, and set $x_n = x_{n-1} = \ldots = x_{n-k+1}$ and $t_n = t_{n-1} = \ldots = t_{n-k+1}$. Consider the problem

$$\max_{(x_{n-k+1}, t_{n-k+1}) \geq 0} t_{n-k+1} - \mathbb{E}[\theta c(x_{n-k+1})|\theta \geq \theta_{n-k+1}]$$

subject to that $x_{n-k+1} \geq x_{n-k}$ and $U(x_{n-k+1}, t_{n-k+1}, \theta_{n-k+1}) = U(x_{n-k}, t_{n-k}, \theta_{n-k+1})$. By an analogous argument it follows that $x_{n-k} = x_{n-k+1}$ and $t_{n-k} = t_{n-k+1}$. So the only solution to the insurers’s problem is a pooling menu. By (4) applied to $\hat{\theta} = \theta_1$, that pooling menu must be a null menu.

Now consider an arbitrary type distribution $F$. Suppose that (4) holds. Consider a sequence of finite support distribution functions $F_n$ which converge weakly to $F$ and such that (4) holds for all $n$ (Hendren (2013) confirms that such a sequence exists). By the preceding argument the profit at each $F_n$ is 0 and the unique optimal menu is null. Since the monopolist’s objective is continuous in the weak convergence topology, the constraint set does not depend on the type distribution, and wlog, the constraint set is compact (in either the relaxed or unrelaxed problem), by Berge’s Theorem (e.g. Aliprantis and Border (2006), Theorem 16.31) the maximum profit at $F$ for the relaxed problem is 0 and the unique optimal menu is the null contract $(0, 0)$ given to all types.

To prove that (4) is necessary for a null menu to maximize expected profit, follow Hendren (2013) Lemma A.2 and suppose that (4) does not hold for some $\theta' \in \Theta_F$. Construct a two-contract menu that gives $(0, 0)$ to every type below $\theta'$ and a contract $(x, t)$ $\gg 0$ to every type $\theta \geq \theta'$ which leaves type $\theta'$ indifferent between $(x, t)$ and $(0, 0)$. If $(x, t)$ is close enough to $(0, 0)$, then this menu yields positive profit to the insurer.

(ii) We will show that $MRS(\mathbb{E}[\theta]/\mathbb{E}[\theta])$ is an upper bound for $MRS(\hat{\theta})/\mathbb{E}[\theta | \theta \geq \hat{\theta}]$ for all $\hat{\theta} \in \Theta_F$. Note that the premise plus convexity of $c$ implies that $c'(0) > 1$. 
Consider any $\hat{\theta} \geq \mathbb{E}[\theta]$ and assume that $MRS(\mathbb{E}[\theta]) \leq c'(0)\mathbb{E}[\theta]$. Then

$$c'(0) \geq \frac{MRS(\mathbb{E}[\theta])}{\mathbb{E}[\theta]} \geq \frac{MRS(\hat{\theta})}{\mathbb{E}[\theta \geq \hat{\theta}]} \geq \frac{MRS(\hat{\theta})}{\mathbb{E}[\theta \geq \hat{\theta}]}.$$

where the second inequality follows from $MRS(z)/z$ decreasing in $z$ and $\hat{\theta} \geq \mathbb{E}[\theta]$, and the third one from $\hat{\theta} \leq \mathbb{E}[\theta \geq \hat{\theta}]$. Thus,

$$MRS(\mathbb{E}[\theta]) \leq c'(0)\mathbb{E}[\theta] \Rightarrow MRS(\hat{\theta}) \leq c'(0)\mathbb{E}[\theta \geq \hat{\theta}], \ \forall \hat{\theta} \geq \mathbb{E}[\theta].$$

Consider any $\hat{\theta} \in \mathbb{E}[\theta]$ and assume that $MRS(\mathbb{E}[\theta]) \leq c'(0)\mathbb{E}[\theta]$. Then

$$c'(0) \geq \frac{MRS(\mathbb{E}[\theta])}{\mathbb{E}[\theta]} > \frac{MRS(\hat{\theta})}{\mathbb{E}[\theta \geq \hat{\theta}]} \geq \frac{MRS(\hat{\theta})}{\mathbb{E}[\theta \geq \hat{\theta}]}.$$

where the second inequality follows from $MRS(z)$ increasing in $z$ and $\hat{\theta} < \mathbb{E}[\theta]$, and the third one from $\mathbb{E}[\theta] \leq \mathbb{E}[\theta \geq \hat{\theta}]$. Thus,

$$MRS(\mathbb{E}[\theta]) \leq c'(0)\mathbb{E}[\theta] \Rightarrow MRS(\hat{\theta}) \leq c'(0)\mathbb{E}[\theta \geq \hat{\theta}], \ \forall \hat{\theta} < \mathbb{E}[\theta].$$

Combine the two cases considered to complete the proof. □

A.4. Proof of Proposition 2

(i) Let $x^*(\theta)$ be the complete information coverage when the consumer’s type is $\theta$, i.e., the solution to $\max_{x,t}(t - \theta c(x))$ subject to $U(x,t,\theta) = U(0,0,\theta)$. Since $U(x,t,\theta)$ is strictly decreasing in $t$, we can invert the constraint and write it as $t = T(x,\theta)$; it is easy to check that $T$ is increasing in $\theta$. Then the problem becomes $\max_{x}(T(x,\theta) - \theta c(x))$, and it is easy to see that its solution strictly decreases in $\theta$ whenever it is positive, we show that $T(x,\theta) - \theta c(x)$ satisfies the strict single crossing property in $(x, -\theta)$. Now, $T_x(\theta, x) = -(U_x/U_t)(T(x,\theta),x,\theta)$, and the first-order necessary condition for an interior maximum is $T_x - \theta c'(x) = 0$. Let

$$m(x,t,\theta) = -\frac{U_x}{U_t}(x,t,\theta) \frac{1}{\theta}$$

and rewrite the first-order condition as $m(x,T(x,\theta),\theta) - c'(x) = 0$. Some calculation (using $m = c' > 1$ and $c$ convex in $x$) reveals that $m_x + \theta mm_t - c'' < 0$, as the Implicit Function Theorem requires. At any $\theta$ with $x^*(\theta) > 0$, apply the Implicit Function Theorem to find that

$$\frac{dx^*(\theta)}{d\theta} = -\frac{m_t}{(m_x + \theta mm_t - c'')} T_\theta + \frac{m_\theta}{(m_x + \theta mm_t - c'')}.$$

where the right side is evaluated at $(x^*(\theta), T(x^*(\theta), \theta))$. Specifically

$$m_t = -(1 - \theta) \frac{u^\ell u'_u (R_n - R_\ell)}{\mathbb{E}[u]^2},$$

where $u_\ell = u(w - \ell + x - t), u_n = u(w - t), \mathbb{E}[u^\ell] = \theta u'_\ell + (1 - \theta)u'_n$, and $R_\ell$ and $R_n$ are the coefficients of absolute risk aversion evaluated at the loss and no loss state wealths; and

$$T_\theta = \frac{U_\theta(x^*, T(x^*, \theta), \theta) - U_\theta(0,0,\theta)}{\mathbb{E}[u')} > 0,$$

$$m_x = -\frac{(1 - \theta)u^\ell u'_n R_\ell}{\mathbb{E}[u']^2} < 0, \ \text{and}$$

(14)
\[ m_\theta = \frac{-u'_x(u'_x - u'_\mu)}{\mathbb{E}[u']^2} < 0. \]  

(16)

It follows from (12)-(15) that \( x \) strictly decreases in \( \theta \) (whenever it is positive) if \( u \) exhibits increasing absolute risk aversion (so that \( R_u \geq R_\ell \)), which includes CARA.

(ii) Note that the functions \( m, m_t, T_\theta, m_x, m_\theta, \) and \( c' \) are uniformly continuous (since they are continuous on the compact set \( \Theta_F \times x^*(\Theta_F) \times T(x^*(\Theta_F), \Theta_F) \)).

Let \( \kappa = \theta_L c'(0) > \theta_L \) and define \( \kappa^* \) to be the smallest value of \( \kappa \) such that \( x(\theta, \kappa) = 0 \) for every \( \theta \in \Theta_F \). The first-order condition reveals that \( x(\theta, \cdot) \) is decreasing in \( \kappa \) for every \( \theta \in \Theta_F \).

It follows that \( x^* \) converges uniformly to the zero function as \( \kappa \to \kappa^*+ \). From (14), \( T_\theta \) converges uniformly to the zero function and, since \( m_x \) and \( m_t \) are uniformly continuous, \( |m_t T_\theta/(m_x + \theta m_m - c'')| \) converges uniformly to the zero function as \( \kappa \to \kappa^*+ \). Finally it is easy to verify that \( m_\theta \) converges uniformly to a function which is negative for every \( \theta \in \Theta_F \). These facts and (12) give us the conclusion. \( \square \)

A.5. Proof of Theorem 2

To begin, since preferences are CARA, we can represent preferences over \( (x, t) \) with a function \( v(x, \theta) - t \), where

\[ v(x, \theta) = -\frac{1}{r} \log[\theta e^{-r(x-\ell)} + 1 - \theta]. \]

As is standard (e.g., Toikka (2011) or Maskin and Riley (1984)), the virtual surplus is

\[ V(x, \theta) = v(\theta, x) - v_\theta(x, \theta) \frac{1 - F(\theta)}{f(\theta)} - \theta c(x), \]

for all \( (x, \theta) \), where

\[ v_\theta(x, \theta) = \frac{\theta e^{-r(x-\ell)}}{\theta e^{-r(x-\ell)} + 1 - \theta}. \]

We can then write the insurer’s problem as choosing \( \tilde{x}(\theta) \) to maximize \( V(x, \theta) \) for each \( \theta \), subject to the constraint that \( \tilde{x} \) is increasing.

Differentiate the virtual surplus to find

\[ V_x(x, \theta) = v_x(x, \theta) - \frac{v_x \theta}{\rho} - \theta c'(x), \]

where

\[ v_x(x, \theta) = \frac{\theta e^{-r(x-\ell)}}{\theta e^{-r(x-\ell)} + 1 - \theta} \]

and

\[ v_{x\theta}(x, \theta) = \frac{e^{-r(x-\ell)}}{(\theta e^{-r(x-\ell)} + 1 - \theta)^2}. \]

So

\[ V_{x\theta}(x, \theta) = v_{x\theta}(x, \theta) \left( 1 + \frac{\rho'(\theta)}{\rho^2(\theta)} \right) - \frac{v_x \theta}{\rho(\theta)} - c'(x), \]

or, using the CARA form,
\[ \mathcal{V}_{x, \theta}(x, \theta) = \frac{e^\xi}{\theta e^\xi + 1 - \theta} \left( 1 + \frac{\rho'(\theta)}{\rho^2(\theta)} \right) + \frac{1}{\rho(\theta)} \frac{2e^\xi(e^\xi - 1)}{\theta(e^\xi + 1 - \theta)^2} - c'(x) \]

where \( \xi = -r(x - \ell) > 0 \). Gather terms to find

\[ \mathcal{V}_{x, \theta}(x, \theta) = \frac{e^\xi}{\theta e^\xi + 1 - \theta} \left( \left( 1 + \frac{\rho'(\theta)}{\rho^2(\theta)} \right) + \frac{1}{\rho(\theta)} \frac{2(e^\xi - 1)}{\theta(e^\xi + 1 - \theta)} \right) - c'(x). \]  \hspace{1cm} (17)

It is easy to verify that

\[ \frac{\rho'(\theta)}{\rho^2(\theta)} = \frac{f'(\theta)}{f(\theta)} + 1, \]  \hspace{1cm} (18)

and that \( \lim_{\theta \to \theta_H} \frac{\rho'(\theta)}{\rho^2(\theta)} = 1 \).

To prove part (i), we will show that the solution to the relaxed problem is strictly decreasing in \( \theta \) near \( \theta_H \), which implies pooling at the top. Note that (17) yields

\[ \lim_{\theta \to \theta_H} \mathcal{V}_{x, \theta}(x, \theta) = \frac{2e^\xi}{\theta(e^\xi + 1 - \theta)^2} - c'(x). \]

Since \( e^\xi (1 - c'(0)(\theta_H)/c'(0)(1 - \theta_H) > 1, x(\theta_H) is interior and solves \( v_x(\theta_H, x - \theta_H c'(x) = 0 \) which, using the expression for \( v_x \) and for \( \xi \), can be written as \( e^\xi/(\theta_H e^\xi + 1 - \theta_H) = c' \) or as \( e^\xi = c'(1 - \theta_H)/(1 - \theta_H c') \). Since by the assumptions made the solution is also interior near \( \theta_H \), we can differentiate the first-order condition with respect to \( \theta \) and verify that \( x'(\theta_H) < 0 \) if and only if \( \mathcal{V}_{x, \theta}(\theta_H, x(\theta_H)) < 0 \). Using these expressions we obtain

\[ \mathcal{V}_{x, \theta}(\theta_H, x(\theta_H)) = \frac{2e^\xi}{\theta e^\xi + 1 - \theta_H} - c'(x(\theta_H)) \]  

\[ = c'(x(\theta_H)) \left( \frac{2}{\theta_H e^\xi (\theta_H) + 1 - \theta_H} - 1 \right), \]

which is negative if the expression in parentheses is negative, and this simplifies to \( c'(x(\theta_H)) > (1 + \theta_H)/2\theta_H \). Since \( x(\theta_H) \in [0, \ell) \), it suffices that \( c'(0) > (1 + \theta_H)/2\theta_H \), completing the proof of part (i).

To prove part (ii), we will show that under the assumed conditions \( \mathcal{V}_{x, \theta} < 0 \) on \( [0, \ell] \times [\theta_L, \theta_H] \) if \( \eta \) is sufficiently large, and so is strictly submodular. This implies that the profit-maximizing menu is pooling: strict submodularity implies that the maximizers of \( \mathcal{V} \) are decreasing in \( \theta \) if we ignore the monotonicity constraint; since \( \mathcal{V}(\cdot, \theta) \) is concave in \( x \), any solution that satisfies the monotonicity constraint is constant. \( ^{26} \)

Since \( \theta_L > 0 \) and the right side of (18) is uniformly bounded above, so is the product on the right side of (17). So as long as \( \eta = \min_{x \in [0, \ell]} c'(x) \) is sufficiently large, \( \mathcal{V}_{x, \theta} < 0 \) for all \( \theta \in [\theta_L, \theta_H] \) and \( x \in [0, \ell] \), and the profit-maximizing menu is pooling.

To prove (iii), first write \( R(x, \theta) = v(x, \theta) - v(\theta)(x)/\rho(\theta) \). Fix any cost function \( c \) with \( 1 < c'(x) < 1/\theta_H \) for every \( x \in [0, \ell] \) (in particular, \( c' \) can be uniformly within any \( \epsilon > 0 \) of 1). By Proposition 2, the complete-information, profit-maximizing menu is decreasing in type.

Let \( x_L \) be the complete-information coverage for the lowest type at the chosen \( c \). Since \( v_{\theta_L} > 0 \)

\( ^{26} \) To relate our argument to Toikka (2011), if the virtual surplus is strictly submodular then \( \int_{[\theta_L, \theta]} \mathcal{V}(x, r) dr \) (this equation (4.2)) is strictly concave in \( \theta \). This implies that the convexification of this function never equals the function itself. By his footnote 12, the menu must be pooling.
(equation (17)), it follows that for any \( \theta \geq \theta_L \), any maximizer of virtual surplus \( V(x, \theta) \) is at most \( x_L \). So we can without loss restrict the domain of \( V \) to be \( D = [0, x_L] \times [\theta_L, \theta_H] \). Crucially, \( x_L < \ell \) since \( c' > 1 \).

We use the following lemma.

**Lemma 2.** If \( R_{x\theta} - R_x/\theta < 0 \) on \( D \), then \( V \) satisfies the strict single crossing property in \( x \) and \(-\theta \) on \( D \).

**Proof.** The inequality implies that \( R_x/c'(x)\theta \) is strictly decreasing in \( \theta \): for \( 1 > \theta'' > \theta' > 0 \), we have

\[
\frac{R_x(x, \theta'')}{\theta''} < \frac{R_x(x, \theta')}{\theta'}.
\]

Rearrange to find

\[
R_x(x, \theta'') < \frac{\theta''}{\theta'} R_x(x, \theta'),
\]

which is equivalent to

\[
V_x(x, \theta'') = R_x(x, \theta'') - c'(x)\theta'' < \frac{\theta''}{\theta'} R_x(x, \theta') - c'(x)\theta'' \frac{\theta'}{\theta'} = \frac{\theta''}{\theta'} V_x(x, \theta').
\]

So

\[
V_x(x, \theta'') < \frac{\theta''}{\theta'} V_x(x, \theta').
\]

For any \( 0 \leq x' < x'' \leq x_L \), integrate both sides with respect to \( x \) between \( x' \) and \( x'' \) to conclude that if \( V(x'', \theta'') - V(x', \theta') \leq 0 \), then \( V(x'', \theta'') - V(x', \theta'') < 0 \). \( \square \)

The strict single crossing property implies that the maximizer of \( V \) is decreasing in \( \theta \) in the sense that if \( \theta_1 > \theta_0 \), and for \( i = 0, 1, x_i \) maximizes \( V(x, \theta_i) \) on \( [0, x_L] \), then \( x_1 \leq x_0 \). We now prove that, if \( f(\theta_L) \) large enough, then \( R_{x\theta} - R_x/\theta < 0 \) holds for \( V \). In what follows let \( \xi = e^{-\rho(x-L)} \). Some calculation reveals that

\[
R_{x\theta} - \frac{R_x}{\theta} = v_{x\theta} - \frac{v_{x\theta}}{\rho} - \frac{\rho'}{\rho} v_{x\theta} - \frac{v_x}{\theta} + \frac{v_{x\theta}}{\theta\rho},
\]

where

\[
v_{x\theta} = \frac{2\xi(\xi - 1)}{[\theta \xi + 1 - \theta] \xi}.
\]

Since \( \rho' \geq 0 \) and \( v_{x\theta} > 0 \), the third term is nonpositive. Both \( v_{x\theta} \) and \( v_{x\theta} \) are uniformly bounded on \( D \). Since \( f(\theta_L) \leq \rho(\theta) \) for any \( \theta \in [\theta_L, \theta_H] \), it follows that for any positive real number \( K \), there is a density with \( f(\theta_L) \) large enough and satisfying the MHRC such that \( \rho(\theta) \geq K \) for all \( \theta \in [\theta_L, \theta_H] \). As a result, for any \( \varepsilon > 0 \), there is a \( K \) large enough so that the second and final terms are uniformly less than \( \varepsilon \) on \( D \). Finally

\[
v_{x\theta} - \frac{v_x}{\theta} = \frac{\xi \theta (1 - \xi)}{[\theta \xi + 1 - \theta]^2}
\]

which is uniformly bounded below 0 on \( D \) (recall \( \theta_L > 0 \), and \( x_L < \ell \), since \( c' > 1 \), so \( \xi \) is bounded uniformly above 1). So for \( f(\theta_L) \) large enough, \( R_{x\theta} - R_x/\theta < 0 \) holds.
The last fact implies that the profit-maximizing menu is pooling. It remains to show the pooling contract is positive. The expected profit from pooling at $x \geq 0$ is $v(x, \theta_L) - c(x) \int \theta f(\theta) d\theta$, since the premium must make the lowest type indifferent between buying coverage of $x$ or zero coverage. Since the complete-information profit is positive for type $\theta_L$, we have $v(x_L, \theta_L) - \theta_L c(x_L) > 0 = v(0, \theta_L) - \theta_L c(0)$. By choosing $f(\theta_L)$ large enough, $\int \theta f(\theta) d\theta$ can be made arbitrarily close to $\theta_L$. It follows that, if $f(\theta_L)$ is large enough, the expected profit from pooling at $(x, t) = (x_L, v(x_L, \theta_L))$ is positive, and thus so is the profit-maximizing pooling contract. \hfill \Box

A.6. Proof of Theorem 3

We follow the same methodology as in Section 4 in Chade and Schlee (2012). We first reformulate the problem as choosing $(u_n(\cdot), \Delta(\cdot))$, where $u_n(\theta) = u(w - t(\theta))$ and $\Delta(\theta) = u(w - t(\theta)) - u(w - \ell + x(\theta) - t(\theta))$ for all $\theta$. After obtaining the optimal $u_n(\cdot)$ and $\Delta(\cdot)$, one can recover the original menu by

$$
t(\theta) = w - h(u_n(\theta))
$$

$$
x(\theta) = \ell - (h(u_n(\theta)) - h(u_n(\theta) - \Delta(\theta)))
$$

for all $\theta$, where $h$ is the inverse function of $u$. As a result, the expected utility of a contract $(x, t)$ for type $\theta$ is now $u_n - \theta \Delta$, and we define the indirect utility of type $\theta$ when facing menu $(u_n(\cdot), \Delta(\cdot))$ and reporting the truth is $U(\theta) = u_n(\theta) - \theta \Delta(\theta)$.

Using $U$ to eliminate $u_n$ from the problem, the insurer’s problem can now be formulated as the following optimal control one, with control $\Delta$ and state $U$.

$$
\max_{U(\cdot), \Delta(\cdot)} \int \theta \left[ w - \lambda \theta \ell - (1 - \lambda \theta) h(U(\theta) + \theta \Delta(\theta)) - \lambda \theta h(U(\theta) - (1 - \theta) \Delta(\theta)) \right] f(\theta) d\theta
$$

subject to

$$
\Delta(\cdot) \quad \text{decreasing}
$$

$$
\Delta(\theta) \geq 0 \quad \forall \theta
$$

$$
\Delta(\theta) \leq \Delta_0 \quad \forall \theta
$$

$$
U'(\theta) = -\Delta(\theta) \quad \text{for almost all } \theta
$$

$$
U(0) = U(0, 0, \theta)
$$

$$
U(\theta) \quad \text{free.}
$$

As in Chade and Schlee (2012), we start with the relaxed problem in which we omit constraints (21)–(23). But unlike that paper, we now provide conditions for $\Delta$ to be strictly increasing in $\theta$. If that property happens at the top, then the optimal one will be flat at the top. And if it is strictly increasing for all $\theta$, then the $\Delta$ that solves the relaxed problem has only “one peak” and hence the optimal one will have only one “flat segment” (Guesnerie and Laffont (1984)). Outside flat segments the optimal $\Delta$ is strictly decreasing and coincides with the solution to the relaxed problem. But if the relaxed problem’s solution is strictly increasing, then the optimal solution $\Delta$

---

27 If for a sequence of densities $f_n$, $\lim_{n \to \infty} f_n(\theta_L) = \infty$, and the MHRC holds, then the sequence of cdf’s $F_n$ converges weakly to the degenerate distribution at $\theta_L$. 
is constant in \( \theta \) since the objective in the relaxed problem is concave. That the same applies to the menu \((x(\cdot), t(\cdot))\) follows from this lemma:

**Lemma 3.** If the function \( \Delta(\cdot) \) that solves the relaxed problem is strictly increasing in \( \theta \) in an open neighborhood of, say, \( \theta = \theta_0 \), then the corresponding \( x(\cdot) \) and \( t(\cdot) \) are strictly decreasing in \( \theta \) in that neighborhood.

**Proof.** Assume that \( \Delta'(\theta) > 0 \) in a neighborhood \( B \) of \( \theta_0 \). From \( u_n(\theta) = U(\theta) + \theta \Delta(\theta) \) and (24) it follows that \( u'_n(\theta) = \theta \Delta'(\theta) > 0 \) in \( B \). Since \( u_n(\theta) = u(w - t(\theta)) \), \( u'_n(\theta) > 0 \) implies that \( t'(\theta) < 0 \) in \( B \). Finally, differentiating the expression for \( x(\theta) \) above and using \( u'_n(\theta) = \theta \Delta'(\theta) \) reveals that \( x'(\theta) < 0 \) in \( B \). \( \square \)

We are now ready to prove Theorem 3:

**Proof of Theorem 3.** By the Maximum Principle and the concavity of the objective for the relaxed program, the following condition is necessary and sufficient to solve the optimal control problem:

\[
\begin{align*}
\int_\theta^\theta_H a(s)f(s)ds, \quad (27)
\end{align*}
\]

where \( a(\theta) = (1 - \lambda \theta) h'(U(\theta) + \theta \Delta(\theta)) + \lambda \theta h'(U(\theta) - (1 - \theta) \Delta(\theta)) > 0 \) since \( \lambda \theta < 1 \).

Differentiating (27) with respect to \( \theta \) reveals that \( \Delta'(\theta) > 0 \) if and only if

\[
\left(2\lambda \theta - \theta - 1\right) \frac{h'_n(\theta)}{\lambda(1 - \theta)} - \left(\frac{f'(\theta)}{f(\theta)} - \lambda (1 - \theta) + 1 - 3\theta\right) \frac{(1 - \lambda \theta)h'_n(\theta) - h'_t(\theta)}{\lambda(1 - \theta)} > 0, \quad (28)
\]

where to simplify the notation we have set \( h'_n(\theta) \equiv h'(U(\theta) + \theta \Delta(\theta)) \) and \( h'_t(\theta) \equiv h'(U(\theta) - (1 - \theta) \Delta(\theta)) \).

Part (i) follows easily from (28) since at \( \theta = \theta_H \) the term in square brackets is zero since complete information coverage is interior for \( \theta_H \). Hence, \( \Delta'(\theta_H) > 0 \) if and only if \( \lambda > (1 + \theta_H)/2\theta_H \). By continuity, \( \Delta'(\theta) > 0 \) for \( \theta \) in a left-neighborhood of \( \theta_H \). Lemma 3 and the discussion before it deliver part (i).

Regarding part (ii), \( \Delta'(\theta) > 0 \) for all \( \theta \) if each of the two terms in (28) are positive for all \( \theta \), and this holds if \( \lambda > (1 + \theta_H)/2\theta_H \) and \( f'(\theta)/f(\theta) < (3\theta - 1)/(\theta(1 - \theta)) \). Once again, appealing to Lemma 3 and the discussion before it yields part (ii). \( \square \)

### A.7. Interim efficiency and incentive compatibility conflict

We showed that under CARA any interim efficient coverage is strictly decreasing in type while incentive compatibility always demands that it be increasing. We used this as an intuition for the conditions for pooling at the top and complete pooling. We will see that a similar conflict exists in the general case with pure loading. The complication is that, by the presence of wealth effects, an interim efficient coverage depends on the level of utility the consumer must obtain.

Consider the problem of maximizing expected profit from type \( \theta \) subject to giving that type a utility level \( \tilde{U}(\theta) \) (in the full information case \( \tilde{U}(\theta) = U(\theta) \)). Formally,
\[
\begin{align*}
\max_{x,t} & \quad t - \theta \lambda x \\
\text{s.t.} & \quad \theta u(\omega - \ell + x - t) + (1 - \theta)u(\omega - t) = \bar{U}(\theta).
\end{align*}
\]

Using the reparameterization of the previous section, it can be rewritten as follows:
\[
\max_{0 \leq \Delta \leq \Delta_0} \omega - \lambda \theta \ell - (1 - \lambda \theta)h(\bar{U}(\theta) + \theta \Delta) - \lambda \theta h(\bar{U}(\theta) - (1 - \theta)\Delta),
\]

where to distinguish its maximizer from the one in the adverse selection case we denote it by \( \hat{\Delta} \).

Assume for simplicity an interior solution, which is characterized by
\[
\frac{1 - \lambda \theta}{\lambda(1 - \theta)} h'(\bar{U}(\theta) + \theta \Delta) - h'(\bar{U}(\theta) - (1 - \theta)\Delta) = 0.
\]

Differentiating this expression with respect to \( \theta \) yields
\[
\hat{\Delta}' = \frac{\lambda(h'_n - h'_\ell)}{\theta(1 - \lambda \theta)h''_n + (1 - \theta)^2 \lambda h''_\ell}.
\]

Now, in the adverse selection problem, one of the conditions for incentive compatibility is that \( \bar{U}' = -\hat{\Delta} \). So if \( \bar{U} \) satisfies this condition, which is true in the relaxed problem under adverse selection, then
\[
\hat{\Delta}' = \frac{\lambda(h'_n - h'_\ell)}{\theta(1 - \lambda \theta)h''_n + (1 - \theta)^2 \lambda h''_\ell} > 0.
\]

That is, interim efficient contracts when \( \bar{U} \) satisfies \( \bar{U}' = -\hat{\Delta} \) have the opposite monotonicity property as the one required by incentive compatibility. And since in the relaxed problem \( U' = -\Delta \) is assumed, a conflict akin to the one pointed out for CARA emerges. The conditions of Theorem 3 provides sufficient conditions for this conflict to be severe enough to engender pooling at the top or complete pooling.

A.8. Proof of Corollary 1

We first show that, at a profit-maximizing menu, there is a minimum coverage \( \bar{x} \) such that every type either gets the no-trade contract or one with coverage at least \( \bar{x} \). First, in the proof of Theorem 1, we proved that Assumption 1 holds for a general cost function, including a discontinuous one with a fixed claims cost given in the corollary statement. As explained in the proof of Theorem 1, Assumption 1 implies equation (10), which asserts that cumulative profit is always nonnegative at any type \( \bar{\theta} \) in any profit-maximizing menu. Let \( \{\bar{x}(\theta), \bar{\ell}(\theta)\}_{\theta \in [\theta_L, \theta_H]} \) be any feasible menu of contracts. Let \( \bar{x} = \inf_{\theta \in [\theta_L, \theta_H]} \bar{x}(\theta) \geq 0 \) and suppose that \( \bar{x} = 0 \). We will show that the menu fails (10), so cannot maximize profit. Let \( \bar{\theta} = \inf_{\theta \in [\theta_L, \theta_H]} \bar{x}(\theta) > 0 \), and \( \bar{\ell} = \inf_{\theta \in [\theta_L, \theta_H]} \bar{x}(\theta) > 0 \). By participation (P), \( \bar{x} = 0 \), and continuity of \( \ell \), it follows that \( \bar{\ell} = 0 \). So there is a \( \bar{\theta} > \bar{\theta} \) such that, for all \( \theta \leq \bar{\theta} \), \( \bar{\ell}(\theta) > k \theta \). For all \( \bar{\theta} < \theta \leq \bar{\theta} \), \( \bar{\ell}(\theta) = c_0(\bar{x}(\theta)) - k \theta < \bar{\ell}(\theta) - k \theta < 0 \), violating (10), so the menu cannot maximize profit.

Now suppose that condition (i) or (ii) or (iii) in the statement of the corollary holds when the cost function is \( c_0 \), and add the fixed claims cost to \( c_0 \).\(^{28}\) By the argument in the last paragraph,

\(^{28}\) If the virtual surplus \( V \) with cost function \( c_0 \) satisfies the strict single crossing property (resp. strictly increasing differences) in \( x \) and \( \theta \) on \( X \times [\theta_L, \theta_H] \) for some interval \( X \subset [0, \ell] \), then \( V \) with cost function \( \epsilon \) (including the fixed
there is some $x > 0$ such that each type’s coverage is either 0 or at least $x$. Let $\theta$ be the infimum of the set of types with positive coverage. The menu maximizes profit only if the submenu for types $\theta \geq \theta$ subject to $x(\theta) \geq x$ maximizes profit on that subset of the type space subject to $x \geq x$. On this constraint set, the properties of the condition assumed (either (1) or (2) or (3)) are satisfied and the virtual surplus is strictly quasiconcave in $x$ for $x \geq x$. By the corresponding part of Theorem 2 or 3, every type $\theta \geq \theta$ is pooled in the profit-maximizing submenu. So the profit-maximizing full menu consists of at most two contracts, the no-trade contract and a positive contract. □

A.9. Proof of Section 6

Perfect Competition. Let $c$ be continuous and strictly increasing with $c(x) \geq x$ for all $x$. In a competitive equilibrium with a finite number of risk types, we first argue that (i) there is no pooling; (ii) the highest risk type gets an efficient contract; (iii) if complete information zero-profit contracts are decreasing in types, then all types other than the highest buy contracts distorted downwards from these efficient contracts. The arguments follow essentially from those in Rothschild and Stiglitz (1976) or in Mas-Colell et al. (1995) chapter 13, so we present them in outline.

First, if there is an equilibrium, each contract makes zero expected profit: if any contract $(x, t)$ earns positive expected profit, then a new contract $(x', t')$ could be introduced that all types buying $(x, t)$ prefer, earns positive profit from them, but that higher types do not prefer to the contracts they buy in the original set. (If lower types buy it, then these contracts continue to earn positive profit.) Second, there is no pooling: if more than one risk type buys a contract $(x, t)$ that makes zero expected profit, then, by the single-crossing property and continuity of profit in type and coverage, another contract could be offered that only the lowest risk that bought $(x, t)$ would prefer to it; and is sufficiently close to $(x, t)$, so expected profit from the new contract is positive (see Figure II, p. 635 in Rothschild and Stiglitz (1976)). So property (i) holds. Third, the highest risk type gets an efficient contract: if the only contracts bought by the highest risk are inefficient and earn zero profit, then another contract could be offered that makes positive expected profit if bought by the highest risk types (and so if bought by anyone; see Figure III on p. 636 in Rothschild and Stiglitz (1976) and the discussion surrounding it). So (ii) holds.

If complete-information competitive contracts are decreasing in risk type (including the non-responsive case), then (iii) all types below the top get contracts that are distorted downward. This follows from complete sorting, the strict single crossing property, and expected utility maximization. If however complete-information competitive contracts are strictly increasing then each type might strictly prefer its complete information contract to the complete information contract of any other type.

References


