Theoretical Economics 7 (2012), 571–607 1555-7561/20120571

Optimal insurance with adverse selection

Hector Chade
Department of Economics, Arizona State University

Edward Schlee
Department of Economics, Arizona State University

We solve the principal–agent problem of a monopolist insurer selling to an agent whose riskiness (loss chance) is private information, a problem introduced in Stiglitz's (1977) seminal paper.

For an *arbitrary* type distribution, we prove several properties of optimal menus, such as efficiency at the top and downward distortions elsewhere. We show that these results extend beyond the insurance problem we emphasize. We also prove that the principal always prefers an agent facing a larger loss and prefers a poorer one if the agent's risk aversion decreases with wealth.

For the standard case of a continuum of types and a smooth density, we show that, under the mild assumptions of a log-concave density and decreasing absolute risk aversion, the optimal premium is *backward-S-shaped* in the amount of coverage—first concave, then convex. This curvature result implies that quantity discounts are consistent with adverse selection in insurance, contrary to the conventional wisdom from competitive models.

Keywords. Principal–agent model, monopoly insurance, common values, wealth effects, quantity discounts, empirical tests for adverse selection.

JEL classification. D82.

1. Introduction

Moral hazard and adverse selection are fundamental problems in insurance. A large literature has explored how each affects insurance contracts. The usual approach for moral hazard is the principal–agent model (which includes the case of monopoly).[1] The usual approach for adverse selection is competitive models, either the Rothschild and

Hector Chade: hector.chade@asu.edu
Edward Schlee: edward.schlee@asu.edu

We thank Stephan Lauermann, Wei Li, Larry Samuelson, Ilya Segal, Lones Smith, and audience members at the University of Montreal, CEA-Universidad de Chile, Miami, Southern California, Georgia, Santa Barbara, Ohio State, Arizona, Arizona State, Duke, Riverside, Johns Hopkins, Iowa, the 2007 Midwest Economic Theory Meetings at Michigan, the 2008 Winter Meetings of the Econometric Society, the 2008 NSF/NBER Decentralization Conference at Tulane, the 2008 SWET Conference at UCSB, the 2008 SED Conference at Michigan, 2009 Cowles Summer Theory Conference, and 2011 Barcelona Jocs for comments. We are also grateful to Jeff Ely (Associate Editor) and three anonymous referees for their detailed suggestions.

[1] Prescott and Townsend (1984) and Chiappori and Bernardo (2003) are exceptions.

DOI: 10.3982/TE671

Stiglitz (1976) model or one of its variants. An exception is the by now classic paper by Stiglitz (1977), who introduces a model of a monopolist insurer selling to an insuree who is privately informed about the chance of a loss. He solves the case of two types of insurees using an intuitive graphical argument and derives a few properties of optimal insurance with a continuum of types. Still, we know surprisingly little about monopoly insurance policies, even for the case of a finite number of types or the continuum-type case distributed according to a continuous density function.

We solve the problem of a monopolist insurer selling to a risk averse agent (the insuree) who is privately informed about the chance of suffering a loss. This problem is important for at least two reasons. First, from a technical standpoint, it does not fit the standard principal–agent model of a monopolist selling to a privately informed consumer (e.g., Maskin and Riley 1984): the agent's risk aversion implies that there are *wealth effects* (except for the constant absolute risk aversion case, henceforth CARA), the agent's type enters the principal's objective function directly (*common values*), and the agent's reservation utility is *type dependent*. Second, most of the received wisdom regarding adverse selection in insurance comes from competitive markets models following Rothschild and Stiglitz (1976), so it is important to explore the opposite benchmark of a monopoly insurer. Since several recent empirical papers find evidence of market power in insurance markets, the need for such analysis is even more pressing.

We divide the paper into two parts. In the first part, we allow for an *arbitrary* type distribution, imposing neither a finite support nor a continuous density function. Despite the generality, we extend *all* of the known results for the two-type case and add others (Theorem 1): the type with the highest chance of a loss gets full coverage (efficiency at the top); all other types get less than full coverage (downward distortions elsewhere); the premium and coverage are nonnegative for all types and co-monotone; and the principal makes positive expected profit (there are always gains to trade). We also argue that the elementary arguments we use in the proofs of these results can be adapted to other screening problems in addition to the canonical insurance problem we emphasize.

As mentioned, one difference with the standard monopoly model is that the agent's wealth matters. An important question is how the agent's initial wealth and the size of the loss affect the principal's profit. Using monotone methods, we prove a new comparative static result showing that the principal always prefers an agent facing a larger loss and prefers a poorer one if the agent's risk aversion decreases with wealth (Theorem 2).

In the second part, we specialize to the case of a continuum of types distributed according to a smooth density. The additional structure allow us to derive new conditions for complete sorting of types (Theorem 3), exclusion (or inclusion) of types, and curvature of the menu. Wealth effects prevent us from bypassing optimal control arguments, as is usually done in the quasilinear case.

Our most surprising result is on the curvature of the premium as a function of the coverage amount. We show that under two mild assumptions—the density is log-concave and the agent's risk aversion decreases with wealth—the premium is *backward-S-shaped* (Theorem 4)—first concave, then convex—a shape that is consistent with global quantity discounts. The curvature property sharply distinguishes a monopolist insurer from a "standard" (i.e., Maskin–Riley) monopolist and from competitive insurers. Maskin and Riley (1984) show that in their model, the price–quantity menu is globally concave under mild assumptions. By contrast, a monopoly insurer's menu has a convex segment for coverage close to full insurance.

An implication of many competitive insurance models (e.g., Rothschild and Stiglitz 1976) is that insurers offer global quantity *premia*, and this is often tested in the empirical literature on adverse selection. A monopolist insurer, however, can offer global quantity *discounts*. Our curvature result shows that we cannot simply infer the absence of adverse selection from the absence of quantity premia.

To illustrate the results and for counterexamples, we derive in closed form the solution for the CARA case with a continuum of types and for the square-root utility case with two types. In particular, the CARA example reveals that a monotone hazard rate does *not* suffice for complete sorting. We also use it to illustrate how the presence of common values affects known results on sorting and curvature in the standard private-values monopoly model.

Related literature

This paper is closely related to four literatures. First, it is related to the literature on insurance with adverse selection started by Rothschild and Stiglitz (1976) for competition, and by Stiglitz (1977) for monopoly; each focuses on the two-type case. And although Stiglitz (1977) contains some results on the continuum of types case, he does not provide general properties of the solution.[2] We solve a more general problem than does Stiglitz (1977), and we compare the predictions of monopoly and competition. In independent work, Szalay (2008) solves the smooth-density case with a different version of the optimal control problem than ours (compare our Section 4 with his Section 3). His solution sheds light on some properties of the optimal contract and allows him to rederive several of Stiglitz's results easily. But none of our main results—our four theorems—follows from his.[3]

Second, a large literature tests implications of the *joint* hypothesis of adverse selection and (some version of) competition (e.g., Chiappori et al. 2006, Cawley and Philipson 1999). Our results help separate implications of adverse selection from competition (e.g., that monopoly insurers can offer quantity discounts).

Third, the paper is related to the literature on principal–agent models with privately informed agents, illustrated by Spence (1977), Mussa and Rosen (1978), Maskin and Riley (1984), Guesnerie and Laffont (1984), Matthews and Moore (1987), Page (1992), Jullien (2000), Nöldeke and Samuelson (2007), and Hellwig (2010). The complications of

[2]The renegotiation stage in Fudenberg and Tirole (1990) resembles a monopoly insurance problem in which the (random) effort chosen in the first stage is the agent's type. With a continuum of effort levels, they derive an optimality condition similar to ours in Section 4. But they mainly use it to find the support of the equilibrium effort distribution and do not explore sorting, exclusion, or curvature.

[3]Schlesinger (1983) derives the first-order condition for yet another version of the monopoly insurer's optimal control problem for the special case of CARA and points out that the optimal menu need not be either concave or convex (but does not provide any results on curvature).

the insurance problem—wealth effects, common values, and type-dependent reservation utilities—are absent in Mussa and Rosen (1978) and in Maskin and Riley (1984), while only wealth effects appear in Matthews and Moore (1987) and only common values appear in Guesnerie and Laffont (1984).[4] Jullien (2000) allows type-dependent reservation utility, and Nöldeke and Samuelson (2007) allow for common values, but each imposes quasilinear preferences (and focuses on particular aspects of the solution). Hellwig (2010) derives the no-pooling and efficiency-at-the-top results in a general principal–agent problem with wealth effects, using a nontrivial extension of the maximum principle. We handle the insurance problem with a general type distribution using elementary arguments.

Finally, the paper is related to Thiele and Wambach (1999) and Chade and Vera de Serio (2011), who determine how an agent's wealth affects the principal's profit with moral hazard. We provide a complete analysis of the issue for our adverse selection problem.

2. The model

We model the monopolist's choice of insurance policies as a principal–agent problem with adverse selection. The agent (insuree) has initial wealth $w > 0$, faces a potential loss $\ell \in (0, w)$ with chance $\theta \in (0, 1)$, and has risk preferences represented by a strictly increasing and strictly concave von Neumann–Morgenstern utility function $u(\cdot)$ on $\mathbb{R}_+$. The loss chance θ, from now on the agent's *type*, is private information to the agent.

The principal (monopolist insurer) is risk neutral, with beliefs about the agent's type given by a cumulative distribution function $F(\cdot)$ with support $\Theta \subset (0, 1)$.[5] Let $\underline{\theta}$ and $\overline{\theta}$ be the smallest and largest elements of Θ; by assumption, $0 < \underline{\theta} < \overline{\theta} < 1$.

For each $\theta \in \Theta$, the principal chooses a contract $(x, t) \in \mathbb{R}^2$ consisting of a premium t and an indemnity payment x in the event of a loss. The expected profit from a contract (x, t) chosen by a type-θ agent is $\pi(x, t, \theta) = t - \theta x$, and the ex ante expected profit from a (measurable) menu of contracts $(x(\theta), t(\theta))_{\theta\in\Theta}$ is $\int_\Theta \pi(x(\theta), t(\theta), \theta)\, dF(\theta)$.

The expected utility of a type-θ agent for a contract (x, t) is $U(x, t, \theta) = \theta u(w - \ell + x - t) + (1 - \theta)u(w - t)$. The function U satisfies the following *strict single-crossing property* (henceforth SSCP): for any two *distinct* contracts (x', t') and (x, t) with $(x', t') \geq (x, t)$ and $\theta' > \theta$, if $U(x', t', \theta) \geq U(x, t, \theta)$, then $U(x', t', \theta') > U(x, t, \theta')$. If $u(\cdot)$ is differentiable with $u'(\cdot) > 0$, this is equivalent in our insurance setting to the Spence–Mirrlees SSCP that $-U_x(x, t, \theta)/U_t(x, t, \theta)$ is strictly increasing in θ; i.e., indifference curves cross once, with higher types being willing to pay more for a marginal increase in insurance.

[4] Biais et al. (2000, Section 4), consider a risk neutral monopoly market maker who trades a risky asset with a risk averse investor who has CARA utility and private information about the asset's mean return and his endowment. The risk neutral market maker can be viewed as an insurer and the investor can be viewed as an insuree, with the private information about the mean return as information about the mean loss in an insurance setting (rather than the probability of a loss). But the difference in the definition of a type leads to completely different conditions for separation and curvature.

[5] The support of a probability measure on the real line (endowed with the Borel σ-field) is the smallest closed set of probability 1. Formally, $\Theta = \{\theta \in (0, 1) \mid F(\theta + \varepsilon) - F(\theta - \varepsilon) > 0, \forall \varepsilon > 0\}$.

By the revelation principle, we consider (measurable) menus $(x(\theta), t(\theta))_{\theta\in\Theta}$ that the agent accepts and announces its true type θ. Formally, the principal solves

$$\max_{x(\cdot),t(\cdot)} \int_{\Theta} \pi(x(\theta), t(\theta), \theta)\, dF(\theta)$$

subject to

$$U(x(\theta), t(\theta), \theta) \geq U(0, 0, \theta) \quad \forall \theta \in \Theta \tag{P}$$

$$U(x(\theta), t(\theta), \theta) \geq U(x(\theta'), t(\theta'), \theta) \quad \forall \theta, \theta' \in \Theta. \tag{IC}$$

As mentioned, this problem is not a special case of the standard monopoly problem (e.g., Maskin and Riley 1984) for three reasons: In the standard problem, the agent has quasilinear utility, her type does not directly affect the principal's profit (private values), and her type does not enter her reservation utility. Here, the agent's risk aversion implies nontrivial wealth effects (except for CARA preferences), the agent's type directly affects the principal's profit (common values), and the agent's type enters her reservation utility.

We shall see that the first two differences are significant. Consider common values. In the standard, private-values monopoly model, profit is increasing in type whether information is complete (first-best) or incomplete. In our model, however, even first-best profit is not increasing in type (higher types demand more insurance, but the cost of selling to them is *higher*). Indeed, first-best profit from a type equals its risk premium, which is concave in the type—first increasing, then decreasing. Moreover, with incomplete information, profit from the highest type can easily be *negative*, implying that the principal does not offer quantity discounts for high coverage, as we explain in Section 5.[6] In addition, common values lead to different sufficient conditions for complete sorting and for the presence or absence of quantity discounts, and wealth effects rule out standard arguments from the quasilinear case, substantially complicating the analysis.

3. The general case: Arbitrary type distribution

We begin with some general properties of optimal menus and then explore some comparative statics, with emphasis on how changes in the agent's wealth affect the principal's profit. We stress that these results hold for an *arbitrary* type support $\Theta \subset (0, 1)$ and a *general* cumulative distribution function on it. Thus, they hold in the canonical special cases of a finite number of types and of a continuum of types with a smooth density. We think the extra generality is important for at least two reasons. First, if we interpret the type distribution as the principal's prior belief, rather than an empirical frequency, there seems to be little justification for putting restrictions on it. And although we specialize later to the smooth density case, it is worth understanding what follows if we do not restrict the type distribution. Second, as Hellwig (2010) points out, in some applications,

[6] We leave as an open question conditions under which profit is monotone in type in the second-best case. The question is potentially important since Chiappori et al. (2006) use a "profit monotonicity" condition—profit does not increase from contracts with higher coverage—to test for adverse selection and competition.

type distributions with both interval support and mass points arise naturally. Suppose, for example, that the (common) prior has no mass points and that there is a positive chance that the agent privately learns its type before the principal offers the menu, but also a positive chance the agent learns nothing about its type. Then the principal faces a type distribution with interval support but a mass point at its mean.

3.1 *Useful lemmata*

Consider a contract that does not give full coverage. Now change the indemnity in the direction of (but not beyond) full coverage and adjust the premium so that expected utility of a type falls. Then the principal's profit from that type increases. We repeatedly use this result to find improvements to a feasible menu.

Lemma 1 (Profitable changes). *Let $\theta \in \Theta$, and let $|x'' - \ell| < |x' - \ell|$ with $(x'' - \ell)(x' - \ell) \geq 0$. If $U(x'', t'', \theta) \leq U(x', t', \theta)$, then $\pi(x'', t'', \theta) > \pi(x', t', \theta)$.*

Proof. Fix $\theta \in \Theta$. Since the agent is strictly risk averse and $U(x', t', \theta) \geq U(x'', t'', \theta)$, it follows that $t'' - t' > \theta(x'' - x')$; otherwise the consumption plan generated by the (x'', t'') would second-order stochastically dominate the plan generated by (x', t') and the agent would strictly prefer (x'', t'') to (x', t'). Thus, $t'' - \theta x'' > t' - \theta x'$. □

Intuitively, if a change from a given contract offers *more* insurance and yet makes the agent *worse off*, then the additional insurance must be "actuarially unfair." But then the change *increases* expected profit.

As is well understood, the next result follows from the incentive compatibility constraints (IC) and the strict single-crossing property. We omit the proof.

Lemma 2 (Monotonicity). *Any feasible menu is monotone in the agent's type: $x(\theta)$, $t(\theta)$, and $x(\theta) - t(\theta)$ are increasing in θ.*

An immediate implication of this result is that higher risks (types) obtain more coverage at an optimal menu, a property that serves as a basis for testing for adverse selection in the empirical literature cited in the Introduction.

3.2 *Properties of optimal menus*

We now list several properties of optimal menus for an arbitrary type distribution.

Theorem 1 (Properties of an optimal menu). *Any solution to the principal's problem is F-almost everywhere (a.e.) equal to a solution $(x(\theta), t(\theta))_{\theta\in\Theta}$ satisfying*

- *(i)* (No Overinsurance) *$x(\theta) \leq \ell$ for all θ;*
- *(ii)* (Nonnegativity) *$x(\theta)$, $t(\theta)$, and $x(\theta) - t(\theta)$ are nonnegative for all θ;*
- *(iii)* (Participation) *(P) is binding for the lowest type $\underline{\theta}$;*

(iv) (Efficiency at the top) $x(\overline{\theta}) = \ell$;

(v) (No pooling at the top) *If* u *is* C^1 *with* $u' > 0$, *then* $x(\theta) < \ell$ *for all* $\theta < \overline{\theta}$;

(vi) (Gains to Trade) *The principal's expected profit is positive.*

Proofs that are not in the text are in the Appendix.

The proof has many steps, but except for a few measure-theoretic details, each is elementary, requiring only a little more work than for the finite-type case. We prove parts (i), (ii), and (iv) by contraposition: we use Lemma 1 to show that if a feasible menu does not satisfy one of these properties, then there is another feasible menu that increases profit for a positive measure of agents.[7] If (iii) fails for a feasible menu, the principal can reduce the utility of each type in each state by the same amount, so profit rises but IC still holds, and if the reduction is small enough, (P) still holds. For (v), if a positive measure of types gets full coverage, the principal can feasibly reduce the coverage of a fraction of those types and raise the premium charged to the remaining types with full coverage. The premium increase leads to a positive first-order increase in profit, while the coverage reduction has only a second-order effect on profit. Finally, we prove (vi) by showing that there is a pooling contract that is accepted by a positive mass of high enough types and yields positive expected profit.

Stiglitz (1977) derives (iii)–(v) for the two-type case and for the continuum case with a smooth positive density. In independent work, Hellwig (2010) proves (iv) and (v) for a principal–agent problem with wealth effects and an arbitrary type distribution, but with private values. Much of his contribution is to derive and exploit the first-order conditions for an optimal control problem with discontinuous densities. We prove these properties for an arbitrary type distribution using elementary arguments, invoking Lemma 1 to find profitable deviations from a feasible menu that fails one of the properties.

Although the proof of Theorem 1 is short, the only part where we used the assumptions that the type is the loss chance and the agent maximizes expected utility is in the proof of part (iii), that participation binds for the lowest type. In the Appendix, we prove that all of the conclusions of Theorem 1 hold simply if U satisfies the strict single-crossing property in (x, t) and θ, and *either* the first-best quantity does not depend on type or wealth, *or* utility is quasilinear in the transfer t. So the conclusions of Theorem 1 extend to the following environments:

- An insurance model in which the agent does not satisfy the expected utility hypothesis (but we impose the SSCP as a primitive assumption);[8]

[7] Since (i) and (ii) show that there is no loss of generality in bounding the set of menus, existence of a solution to the principal's problem follows from Page (1992).

[8] That Theorem 1 extends to this setting complements Machina (1995), who explores whether well known properties of insurance under complete information extend to (smooth) nonexpected utility models. Notice that the assumption of the SSCP is important, since the SSCP can fail without expected utility; e.g., Ormiston and Schlee (2001) show that this property fails for mean–variance preferences.

- An insurance model in which the type θ orders the agents by their degree of risk aversion (but we impose no overinsurance as a primitive assumption);[9]
- The standard monopoly model pricing model with quasilinear preferences and no wealth effects, but allowing for common values.[10]

Allowing common values in the standard monopoly model is especially useful if we think of the type as the quality of the good exchanged in a procurement problem: the agent is a supplier of the good who is privately informed about its quality θ with preferences over contracts represented by $t - c(x, \theta)$, where c satisfies strictly increasing differences; the principal is a monopsony buyer with preferences over contracts given by $\theta x - t$.

3.3 *Comparative statics*

In this section, we still refrain from any assumption on the type distribution. The main result is that the principal's maximum expected profit is increasing with the agent's loss size or with risk aversion, and decreasing with wealth if risk aversion decreases with wealth.

3.3.1 *The principal prefers a poorer agent* The agent's risk aversion introduces wealth effects that are absent in the standard screening model with quasilinear utility: here changing the wealth endowment changes the set of feasible menus. An important question is how the agent's wealth endowment affects the principal's profit: Does the principal prefer a richer or poorer agent? Does he prefer one facing a larger or smaller potential loss?

In the first-best case (observable types), the answers are immediate: the risk premium is higher if the loss amount or risk aversion is higher, or the agent's initial wealth is smaller and risk aversion decreases with wealth (DARA). The first-best argument, however, *fails* with adverse selection, since the incentive compatibility and participation constraints change with the agent's wealth endowment or risk aversion. And, unfortunately, the constraint sets cannot be ordered by inclusion as wealth or risk aversion changes. Despite this nontrivial complication, the principal still prefers a poorer agent (under DARA) and a larger loss size.

It is important to emphasize that we analyze the principal's preferences over the initial wealth of the agent for a *given* loss amount. Clearly, the principal might prefer a richer agent if the potential loss rises with wealth. Also, recall that we assume $\ell < w$; it should be understood that we consider only changes in w and ℓ that preserve this inequality. Since $t \le x \le \ell$ in an optimal menu, wealth is *positive* in both states whenever $\ell < w$, thereby avoiding problems of liquidity constraints.

Theorem 2 (Wealth effects). *The principal's maximum profit is*

[9] In this setting, indifference curves "double-cross" below and above full insurance. The assumption rules out the second crossing and thus the SCCP holds.

[10] For the standard monopoly model, the loss amount ℓ is replaced by the first-best output for the highest demand type. Unlike the insurance model, however, "no pooling at the top" and "no overprovision" do not imply "downward distortions below the top."

(i) increasing in the loss size ℓ;

(ii) increasing in the agent's risk aversion;

(iii) decreasing in wealth w if the agent's preferences satisfy DARA.

The proof of each part is similar. Fix an optimal menu. After an increase in risk aversion or the loss amount, the menu continues to satisfy the *downward* incentive and participation constraints; i.e., lower contracts (involving less insurance at a lower premium) become less attractive. We do, however, need to worry about possible violations of *upward* incentive constraints. Now, if we simply let each type choose its best contract from the *original* menu, each agent type will choose either the same contract or a *higher* contract. If the chosen contract is higher, then by Lemma 1, the principal's profit from that type does not fall. Since the resulting menu satisfies (IC) and (P), maximum expected profit cannot fall.[11]

Contrast Theorem 2 with how the agent's wealth affects the principal under moral hazard (Thiele and Wambach 1999). Under moral hazard, a fall in wealth makes the agent less lazy (loosens the incentive constraints), but under DARA, much stronger conditions are imposed to conclude that the principal prefers a poorer agent (namely, the coefficient of prudence is less than three times the coefficient of absolute risk aversion). With adverse selection, a decrease in agent's wealth loosens both the downward incentive and the participation constraints under *just* DARA, and we show that the potential tightening of the upward incentive constraints does not lower expected profit.

The proof outline of Theorem 2 suggests a conjecture about what happens to the menu as risk aversion or the loss amount increases: each type gets a higher contract. We were surprised that this conjecture is false, as the following example shows.

Example 1 (Menu nonmonotonicity in risk aversion). Suppose the agent can be one of two types, $\theta_1 < \theta_2$, and let $0 < f_i < 1$, $i = 1, 2$, be the chance that $\theta = \theta_i$. Assume also that the agent's utility function is $u(\cdot) = \sqrt{(\cdot)}$, which satisfies DARA. We show in the Appendix that for any $(\theta_1, \theta_2, \ell, w)$ with $\theta_1 < \theta_2$ and $\ell < w$, the premium of the high type $t(\theta_2)$ is locally strictly increasing in w if f_1 is sufficiently close to 1 and the loss amount is small enough. That is, the high type pays a lower premium when he becomes poorer (hence more risk averse). ◊

3.3.2 *Principal's profit and changes in the type distribution* As mentioned, in the standard monopoly pricing model with private values, the principal's profit is higher for higher types under both complete and incomplete information, so a first-order stochastic dominance (FOSD) shift in the type distribution increases the principal's expected profit. In our monopoly insurance model, complete information profit is the agent's risk premium, which is strictly concave in the loss chance and equals 0 at $\theta = 0$ and at

[11] Machina's 1995 extension of expected-utility comparative statics of increased risk aversion on insurance demand to smooth nonexpected utility representations suggests that at least part (ii) of Theorem 2 extends to risk averse nonexpected utility preferences after we impose the single-crossing property. But we conjecture that additional assumptions are needed to extend parts (i) and (iii).

$\theta = 1$. Since the risk premium is strictly decreasing over an interval of types, it is easy to construct examples in which an FOSD shift in the type distribution lowers the principal's profit. Let $0 < \theta_1 < \theta_2 < 1$ be two types in the interval on which the risk premium is strictly decreasing, and consider two different two-point distributions with support $\theta_i \pm \varepsilon$ for $i = 1, 2$. As ε tends to 0, profit from distribution i tends to the risk premium at θ_i, so for ε small enough, profit is lower for the higher distribution.

Since the risk premium is concave in type, a natural conjecture is that the principal's profit falls with a mean-preserving increase in risk in the type distribution. The next example shows that this conjecture is false.

Example 2 (An increase in risk can raise the principal's profit). Suppose the agent can be one of two types, $\theta_1 < \theta_2$, and let $0 < f_i < 1$, $i = 1, 2$, be their probability distribution. Assume that $u(\cdot)$ is twice-continuously differentiable, that θ_2 is in the interval of types over which the risk premium is strictly increasing, and that f_1 is small enough so that θ_1 is excluded from the optimal menu (see Stiglitz 1977 and Section 4.3). Thus, the optimal contract offers the high type full insurance at the first-best premium $t(\theta_2) = w - h(\theta_2 u(w - \ell) + (1 - \theta_2)u(w))$, which yields an expected profit equal to $f_2 \times \pi(\theta_2)$, where $\pi(\theta_2)$ is the risk premium at θ_2 (i.e., $\pi(\theta_2) = w - \theta_2\ell - h(\theta_2 u(w - \ell) + (1 - \theta_2)u(w)))$. Consider an increase in θ_2 and a decrease in θ_1 so that the mean type is unchanged, but the changes are sufficiently small so that the low type continues to be excluded and the high type continues to fall in the interval in which the risk premium strictly increases. Then this increase in risk raises the principal's expected profit. ♢

3.4 *A reformulation*

So far we have restricted contracts to be deterministic: each type is offered a single premium–indemnity pair. In this section, we show that this restriction is without loss of generality (w.l.o.g.). The simplest way to prove this result is to reformulate the problem as one in which the principal chooses a menu of state-contingent utilities, rather than a menu of premium–indemnity pairs. We also use this formulation in the rest of the paper to derive other properties of optimal contracts.

Given a menu $(x(\theta), t(\theta))_{\theta\in\Theta}$, define, for each $\theta \in \Theta$,

$$u_n(\theta) = u(w - t(\theta))$$
$$\Delta(\theta) = u(w - t(\theta)) - u(w - \ell + x(\theta) - t(\theta)).$$

A menu $(x(\theta), t(\theta))_{\theta\in\Theta}$ uniquely defines a menu $(u_n(\theta), \Delta(\theta))_{\theta\in\Theta}$. Conversely, given $(u_n(\theta), \Delta(\theta))_{\theta\in\Theta}$, we can recover $(x(\theta), t(\theta))_{\theta\in\Theta}$ by (henceforth, $h = u^{-1}$)

$$t(\theta) = w - h(u_n(\theta)) \tag{1}$$
$$x(\theta) = l - \big(h(u_n(\theta)) - h(u_n(\theta) - \Delta(\theta))\big). \tag{2}$$

An equivalent formulation of the principal's problem is

$$\max_{u(\cdot),\Delta(\cdot)} \int_\Theta \big[w - \theta\ell - (1 - \theta)h(u_n(\theta)) - \theta h(u_n(\theta) - \Delta(\theta))\big]\, dF(\theta)$$

subject to

$$u_n(\theta) - \theta\Delta(\theta) \geq u_n(\theta') - \theta\Delta(\theta') \quad \forall \theta, \theta' \in \Theta$$
$$u_n(\theta) - \theta\Delta(\theta) \geq U(0, 0, \theta) \quad \forall \theta \in \Theta.$$

In other words, we can think of a menu of contracts as specifying, for each type θ, a utility $u_n(\theta)$ in the no loss state, and a change in utility $\Delta(\theta)$ in case of a loss. In this formulation, the constraints are *linear* in the contracting variables, and the objective function is strictly concave in them. This formulation makes it clear that stochastic menus cannot improve upon deterministic ones.

Proposition 1 (Deterministic menus). *Any solution to the principal's problem involves a deterministic contract for almost all types.*

Proof. Suppose the principal offers each type θ in a set of positive probability a contract consisting of random variables $(\tilde{x}(\theta), \tilde{t}(\theta))$. In the reformulated problem, this implies that the principal offers each type in that same set a contract consisting of random variables $(\tilde{u}_n(\theta), \tilde{\Delta}(\theta))$. Since the constraints are linear in these variables, any type-θ agent's constraints are satisfied if $\tilde{u}_n(\theta)$ and $\tilde{\Delta}(\theta)$ are replaced with their expected values. But profit increases from this change, since the objective function is strictly concave. □

Arnott and Stiglitz (1988, Proposition 10), proved a similar result for the two-type case. Proposition 1 extends it to any number of types arbitrarily distributed. A more recent contribution is Strausz (2006), who proves that in a principal–agent model with quasilinear utilities and a finite number of types, random contracts are not optimal if there is no bunching in any optimal deterministic menu. Our Proposition 1 holds even if there is bunching in the optimal deterministic menu.

4. Continuum of types: The smooth case

A major objective of the paper is to analyze the curvature of optimal menus (for example, the existence of quantity discounts). The issue is of paramount importance in the analysis of nonlinear pricing models, and, as we explain in Section 4.4.2, it plays a prominent role in the empirical literature on adverse selection. We now specialize to a smooth model with a continuum of types. Specifically, we assume that $u(\cdot)$ is C^2, with positive first and negative second derivatives, $\Theta = [\underline{\theta}, \overline{\theta}]$, and $F(\cdot)$ is C^2 on Θ with density $F'(\cdot) = f(\cdot)$ that is positive on $(\underline{\theta}, \overline{\theta})$.[12] Except for the possibility of a zero density at the endpoints, this setup is the most common one used in contracting problems with adverse selection. Under these assumptions, we prove strong results for complete sorting of types and for the existence of quantity discounts.

[12]We leave as an open question whether the argument in Appendix A in Maskin and Riley (n.d.) could be adapted to show that the (w.l.o.g.) continuous solution of our smooth continuum model can be well approximated by a model with a large, but finite, type set.

4.1 *The optimal control problem*

Let $V(\theta)$ be the agent's indirect utility function for a bounded menu $(u_n(\theta), \Delta(\theta))_{\theta\in\Theta}$ satisfying $\Delta(\cdot) \geq 0$, (IC), and (P), that is, $V(\theta) = u_n(\theta) - \theta\Delta(\theta)$. Since the agent's objective is affine in θ, $V(\cdot)$ is convex, and since it is also continuous on $[\underline{\theta}, \overline{\theta}]$ (Lemma 4), V is absolutely continuous on $[\underline{\theta}, \overline{\theta}]$. By standard arguments, (IC) holds if and only if $V'(\theta) = -\Delta(\theta)$ almost everywhere and $\Delta(\cdot)$ is nonincreasing, so there is no loss of generality in replacing (IC) with these two conditions. Moreover, by Theorem 1(i) and (ii), we can assume that $0 \leq \Delta(\theta) \leq \Delta_0 = u(w) - u(w - \ell)$ for all $\theta \in \Theta$, while by (iii), $V(\underline{\theta}) = U(0, 0, \underline{\theta})$. We now write the principal's problem as an optimal control problem with a control variable $\Delta(\cdot)$ (measurable, by monotonicity), an absolutely continuous state variable $V(\cdot)$, and a free endpoint at $V(\overline{\theta})$:

$$\max_{V(\cdot),\Delta(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \big[w - \theta\ell - (1-\theta)h(V(\theta) + \theta\Delta(\theta)) - \theta h(V(\theta) - (1-\theta)\Delta(\theta))\big] f(\theta)\, d\theta$$

subject to

$$\Delta(\cdot) \quad \text{nonincreasing} \tag{3}$$

$$\Delta(\theta) \geq 0 \quad \forall\theta \tag{4}$$

$$\Delta(\theta) \leq \Delta_0 \quad \forall\theta \tag{5}$$

$$V'(\theta) = -\Delta(\theta) \quad \text{for almost all } \theta \tag{6}$$

$$V(\underline{\theta}) = U(0, 0, \underline{\theta}) \tag{7}$$

$$V(\overline{\theta}) \quad \text{free.}$$

Notice that the optimal control problem is not identical to the original problem, since the control problem includes constraints (4) and (5) that are absent in the original problem, and the original problem imposes participation on all types, something the optimal control problem imposes only on the lowest type (and with equality). The solution to the optimal control problem nonetheless solves the original problem. It does so since, in the presence of (4) and (5), constraints (7) and (6) imply the omitted participation constraints, and (4), (5), and (7) are all justified by Theorem 1.

In the standard model with quasilinear preferences, the objective function is *linear* in the indirect utility. The usual next step in that case is to use Fubini's theorem to eliminate the transfer and maximize pointwise with respect to the remaining variable, a great simplification (Guesnerie and Laffont 1984). Since our objective is not linear in the indirect utility $V(\cdot)$, we are forced to proceed with optimal control arguments.

Consider the "relaxed problem" that ignores (3)–(5), and let $\lambda(\cdot)$ be the costate variable of the problem. If a solution to the relaxed problem satisfies the omitted constraints, then of course it solves the original problem.

The Hamiltonian is

$$H(V, \Delta, \lambda, \theta) = \big[w - \theta\ell - (1-\theta)h(V(\theta) + \theta\Delta(\theta)) \\ - \theta h(V(\theta) - (1-\theta)\Delta(\theta))\big] f(\theta) - \lambda(\theta)\Delta(\theta),$$

and by the maximum principle (e.g., Clarke 1976 or Vinter 2000, Chapter 6), at any solution to the relaxed problem, there is an absolutely continuous function $\lambda(\cdot)$ such that for almost all θ,

$$-\lambda(\theta) = f(\theta)\theta(1-\theta)\big[h'(V(\theta)+\theta\Delta(\theta)) - h'(V(\theta)-(1-\theta)\Delta(\theta))\big] \tag{8}$$

$$\lambda'(\theta) = f(\theta)\big[(1-\theta)h'(V(\theta)+\theta\Delta(\theta)) + \theta h'(V(\theta)-(1-\theta)\Delta(\theta))\big] \tag{9}$$

$$\lambda(\overline{\theta}) = 0, \tag{10}$$

as well as (6) and (7).[13]

Note that $\lambda(\theta) \leq 0$, and $\lambda(\theta) < 0$ if $\Delta(\theta) > 0$; also $\lambda'(\theta) > 0$ for almost all θ. Integrate (9) with respect to θ, use (10), and replace the resulting expression in (8) to find that for almost all θ,

$$f(\theta)\theta(1-\theta)\big[h'(V(\theta)+\theta\Delta(\theta)) - h'(V(\theta)-(1-\theta)\Delta(\theta))\big] = \int_{\theta}^{\overline{\theta}} a(s)f(s)\,ds, \tag{11}$$

where $a(s) = (1-s)h'(V(s)+s\Delta(s)) + sh'(V(s)-(1-s)\Delta(s)) > 0$.

Equation (11) illustrates the standard *efficiency versus information rent* trade-off of screening problems: the left side is the *marginal benefit* (increase in profit) of providing type θ with additional insurance (lower $\Delta(\theta)$), i.e., more efficiency, while the right side is the *marginal cost* (decrease in profit) of doing so, as it leads to an increase in the information rent left to all higher types to ensure that incentive compatibility is satisfied. To see this last point, note that the cost of giving type θ one more unit of utility is $a(\theta)f(\theta)$, but giving θ an additional unit of utility also increases the utility of all higher types by one unit, and thus the cost to the principal is given by $\int_{\theta}^{\overline{\theta}} a(s)f(s)\,ds$.

The right side of (11) is continuous in θ. Since the left side $f(\theta)\theta(1-\theta)[h'(V(\theta)+\theta\delta) - h'(V(\theta)-(1-\theta)\delta)]$ is continuous in θ and strictly increasing in δ, it follows that any solution $\Delta(\cdot)$ to (11) is continuous. *Hence, any solution to the relaxed problem is almost everywhere equal to a continuous function* (and any discontinuities to the solution to the relaxed problem are removable). It follows that there is no loss of generality to restrict $\Delta(\cdot)$ to be continuous and hence $V(\cdot)$ to be C^1 (the last by Theorem 3 in Milgrom and Segal 2002). After imposing continuity, (11) implies that the omitted constraint (4) is satisfied for all types. Note also that (consistent with Theorem 1) $\Delta(\overline{\theta}) = 0$ and $\Delta(\theta) > 0$ for all $\theta < \overline{\theta}$: type $\overline{\theta}$ gets full coverage and all other types get partial coverage.[14]

Before presenting our results on sorting and curvature, we give the (closed form) solution for the special case of CARA preferences for which wealth affects are absent. We use it to isolate how the presence of common values affects our results and also for counterexamples. We define $\rho(\cdot)$ to be the *hazard rate* $f(\cdot)/(1-F(\cdot))$ of the type distribution.

[13] The Hamiltonian is strictly concave in (V, Δ), so these conditions are also sufficient for optimality.

[14] This result is immediate if $f(\overline{\theta}) > 0$. And if $f(\overline{\theta}) = 0$, then $\Delta(\theta_n)$ tends to zero for any sequence θ_n in Θ tending to $\overline{\theta}$. To see this second point, divide both sides of (11) by $f(\theta_n)$ and use the mean value theorem to write the right side as $\psi(\hat{\theta})(1-F(\theta_n))/f(\theta_n)$ for some $\hat{\theta}$. The conclusion now follows since $\lim_{\theta\to\overline{\theta}} f(\theta)/(1-F(\theta)) = \infty$ (Barlow et al. 1963, pp. 377–378).

Example 3 (CARA preferences). Set $u(z) = -e^{-rz}$. Letting $v(x, \theta) = -\log[(1-\theta) + \theta e^{r(\ell - x)}]/r$, the certainty equivalent of (x, t) is $v(x, \theta) + w - t$, which represents the same preferences over contracts as $U(x, t, \theta)$.[15] The optimal indemnity in the relaxed problem satisfies the first-order condition $v_x(x, \theta) - \theta - (v_{x\theta}(x, \theta)/\rho(\theta)) = 0$ for each θ, which simplifies to (setting $\xi = e^{r(\ell - x)}$)

$$\theta(1-\theta)[(1-\theta) + \theta\xi](\xi - 1) - \frac{\xi}{\rho(\theta)} = 0. \tag{12}$$

This equation is quadratic in ξ; since $\xi \geq 0$, we take the positive solution

$$\xi(\theta) = \frac{\frac{1}{\rho(\theta)} - \theta(1-\theta)(1-2\theta) + \sqrt{\left(\theta(1-\theta)(1-2\theta) - \frac{1}{\rho(\theta)}\right)^2 + 4\theta^3(1-\theta)^3}}{2\theta^2(1-\theta)}. \tag{13}$$

Since $\xi(\theta) = e^{r(\ell - x(\theta))}$, we have $x(\theta) = \ell - (\log \xi(\theta)/r)$ and $t(\theta) = v(x(\theta), \theta) - v(0, \underline{\theta}) - \int_{\underline{\theta}}^{\theta} v_\theta(x(s), s)\, ds$, which completes the solution to the (relaxed) problem. ◊

4.2 *Complete sorting*

We now determine when the solution to the relaxed problem satisfies constraint (3) by providing conditions on the distribution of types under which the optimal menu exhibits *complete sorting*. In the standard contracting model with quasilinear utility and private values, sufficient conditions for complete sorting can be read easily off the first-order conditions of the problem. In particular, complete sorting follows in the standard model if the hazard rate $\rho(\cdot)$ is increasing in θ. To illustrate the challenge of finding sufficient conditions for complete sorting in our model, suppose for a moment that we suppress the common values aspect of the model by setting the principal's marginal cost of insurance equal to $k = \underline{\theta}$. In that case, the first-order condition (11) becomes

$$(1-k)\theta h'(V(\theta) + \theta\Delta(\theta)) - k(1-\theta)h'(V(\theta) - (1-\theta)\Delta(\theta)) = \int_\theta^{\overline{\theta}} a(s, k) f(s)\, ds / f(\theta), \tag{14}$$

where

$$a(s, k) = (1-k)h'(V(s) + s\Delta(s)) + kh'(V(s) - (1-s)\Delta(s)).$$

Fixing $\Delta(\theta)$ and $V(\theta)$, the left side of (14) is strictly increasing in θ. If the right side were decreasing in θ, then sorting would be complete, and if a were constant, then the right side would be decreasing if the hazard rate were increasing. Unfortunately, when we set $k = \theta$ (common values), the left side of (14) is *not* decreasing. And, in general, monotonicity of the hazard rate does not imply that the right side is increasing.

[15] Since the certainty equivalent is linear in t, it is simplest to solve for $x(\cdot)$ without transforming the variables (as in the standard monopoly model). That is, replace $t(\cdot)$ from the objective function, integrate by parts, and maximize pointwise with respect to $x(\cdot)$, ignoring the monotonicity condition.

Since the solution $\Delta(\cdot)$ to (11) is continuous, the integral on the right side of (11) is differentiable in θ. As the left side has a nonzero derivative with respect to Δ, the implicit function theorem implies that $\Delta(\cdot)$ is C^1 on Θ (except possibly at the endpoints when f is zero there), so we can replace (3) with $\Delta'(\theta) \leq 0$ for all $\theta \in \Theta$.[16]

Differentiate (8) with respect to θ and use (6) to obtain, after some algebra,

$$\Delta'(\theta) = \frac{\lambda(\theta)[f'(\theta)\theta(1-\theta) + f(\theta)(1-2\theta)] - f(\theta)\theta(1-\theta)\lambda'(\theta)}{f(\theta)^2\theta^2(1-\theta)^2[\theta h''(V(\theta)+\theta\Delta(\theta)) + (1-\theta)h''(V(\theta)-(1-\theta)\Delta(\theta))]}. \tag{15}$$

Since $h''(\cdot) > 0$, the denominator of (15) is positive, and the sign of $\Delta'(\theta)$ depends on the sign of the numerator. Notice that, consistent with Theorem 1(v), equation (15) implies that there is *no pooling of types at the top*; i.e., $\Delta'(\overline{\theta}) < 0$.[17]

We now give three sufficient conditions for complete sorting.

Theorem 3 (Complete sorting: sufficient conditions). *The optimal menu completely sorts all types who get some insurance if one of the following conditions holds.*

(i) $\frac{\rho'(\theta)}{\rho(\theta)} > \frac{3\theta-1}{\theta(1-\theta)}$ *for all* $\theta \in [\underline{\theta}, \overline{\theta}]$.

(ii) $f(\cdot)$ *is log-concave and either* $\overline{\theta} \leq 1/2$ *or* $f'(\cdot) \geq 0$.

(iii) $f(\cdot)$ *is* C^1, $f'(\cdot)/f(\cdot)$ *is bounded below on* Θ, *and* ℓ *is sufficiently small (how small depends on the primitives).*

Stiglitz (1977) finds that a sufficient condition for complete sorting is $f'(\theta)/f(\theta) > (3\theta-2)/\theta(1-\theta)$ for all θ. Unfortunately, if $\overline{\theta} > 2/3$ and $f'(\theta) \leq 0$ for some $\theta > 2/3$, then this condition fails. For example, the commonly used uniform distribution fails this condition if $\overline{\theta} > 2/3$. Part (i) substantially extends Stiglitz's sufficient condition. It clearly modifies the familiar monotone hazard rate condition (MHRC): it is weaker than the MHRC for $\theta < 1/3$ and stronger otherwise. In particular, if $\overline{\theta} \leq 1/3$, then the MHRC also implies complete sorting in our model.

Since by part (i), the MHRC implies complete sorting among low enough types and, by Theorem 1, there is no pooling at the top, it is natural to wonder whether the MHRC suffices for complete sorting. We use Example 3 (CARA) to show that even without wealth effects, the MHRC does *not* imply complete sorting.

Example 4 (The MRHC does not imply complete sorting). Refer to (13) in Example 3 and recall that sorting is complete if and only if $\xi(\cdot)$ is strictly decreasing. If $\rho(\cdot)$ were set equal to a constant, then $\xi(\theta)$ would tend to $+\infty$ as θ tends to 1. Since Θ is bounded, the hazard rate $\rho(\cdot)$ cannot be constant on all of Θ, but the observation suggests that if the hazard rate increases slowly enough for large enough θ, then sorting is not complete. To

[16]Since $\Delta(\cdot)$ is continuous, the integrand on the right side of (13) is continuous in s, ensuring that the integral is differentiable in θ. Since the left side is differentiable and $h''(\cdot) > 0$, the hypotheses of the implicit function theorem hold.

[17]If $f(\overline{\theta}) > 0$, then (9) and (10) imply that $\Delta'(\overline{\theta}) < 0$. If $f(\overline{\theta}) = 0$, then $f'(\theta)/f(\theta) \to -\infty$ as $\theta \to \overline{\theta}$, so in either case, $\limsup_{\theta\to\overline{\theta}} \Delta'(\theta) < 0$, implying $\Delta'(\theta) < 0$ for types near $\overline{\theta}$.

confirm the conjecture, let $\overline{\theta} > 2/3$ and let $f(\cdot)$ be any density with an increasing hazard rate but with $\rho'(\hat{\theta}) = 0$ at some $\hat{\theta} > 2/3$. Differentiate either (12) or (13) with respect to θ to find, after some work, that $\xi'(\hat{\theta}) > 0$, so sorting is not complete. For a numerical example, let $f(\cdot)$ be the truncated exponential at $\overline{\theta} < 1$ with parameter η. Then $\rho(\theta) = (\eta e^{-\eta\theta})/(e^{-\eta\theta} - e^{-\eta\overline{\theta}})$ and $\rho'(\theta) > 0$ for all θ. Insert this expression into (13), and set $\eta = 30$ and $\overline{\theta} = 0.9$ to find that $\xi(0.5) = 1.14 < 1.2 = \xi(0.8)$, so that the solution $x(\cdot)$ to the relaxed problem is not increasing everywhere. ◊

Part (ii) of Theorem 2 shows that sorting is complete if the density is log-concave and either the highest type below 1/2 or the density is nondecreasing. It is easy to check that (ii) holds for the class of densities on $[\underline{\theta}, \overline{\theta}]$ given by $f(\theta) = (1+\alpha)\theta^\alpha/(\overline{\theta}^{\alpha+1} - \underline{\theta}^{\alpha+1})$, $\alpha \geq 0$, which includes the uniform distribution that is ruled out by Stiglitz's condition.

Finally, part (iii) shows that if $f(\cdot)$ is C^1 and the likelihood ratio $f'(\cdot)/f(\cdot)$ is bounded below, then there is a region of losses for which sorting is complete. This result does not impose monotonicity of the hazard rate or of the likelihood ratio. Moreover, if f is C^1, the second condition of (iii) holds if the density is positive everywhere.

4.3 *Exclusion*

It follows from (3) and (5) that the set of types that receive some insurance at the optimum is an interval $[\theta_0, \overline{\theta}]$, with $\theta_0 \geq \underline{\theta}$.

We now prove two results on the value of θ_0: one for no type to be excluded ($\theta_0 = \underline{\theta}$) and one for a subset of low types to be excluded ($\theta_0 > \underline{\theta}$).

Proposition 2 (No exclusion and exclusion). *Assume that sorting is complete.*

(i) If $\rho(\widetilde{\theta})\widetilde{\theta}(1-\widetilde{\theta})$ is low enough for $\widetilde{\theta} \in [\underline{\theta}, \overline{\theta})$ or if the agent's risk aversion is low enough on $[w-\ell, w]$, then all types in $[\underline{\theta}, \widetilde{\theta}]$ are excluded in an optimal menu.

(ii) If $f(\underline{\theta})$ is large enough or if $f(\underline{\theta}) > 1/\underline{\theta}(1-\underline{\theta})$ and the agent's risk aversion is large enough on $[w-\ell, w]$, then no type is excluded in an optimal menu.

Intuitively, part (i) shows that a type is excluded if it is close to zero or the hazard rate is low enough there, or if the agent's risk aversion is low enough. Part (ii) shows that no type is excluded (and thus constraint (5) does *not* bind) if there are enough low types in the population or if the agent is risk averse enough.[18]

4.4 *Curvature*

We now come to the most surprising part of the paper: our analysis of the curvature of the premium as a function of coverage. We are particularly interested in whether

[18] The usual way to analyze exclusion is to find the optimal contract for a given θ_0 and then optimize with respect to θ_0. In the quasilinear case, this is straightforward, since one of the variables of the menu can be omitted from the problem. With wealth effects it is more challenging, since both variables must be solved jointly. For this reason, we proceed differently in the proof of Proposition 2(ii).

monopoly insurers offer quantity discounts. The issue is important for at least two reasons. First, firms commonly offer quantity discounts in practice, so it is natural to ask whether a monopolist insurer would. Second, in competitive insurance models, quantity *premia* rather than discounts are the rule, since equilibrium prices equal marginal cost (in many competitive models). Since this curvature property has been used in the empirical literature to test for adverse selection, it is important to know if the implication holds for a monopoly insurer as well.

Stiglitz (1977, pp. 427–428) gives an example in which the optimal premium is an affine function of the coverage amount: imposing CARA preferences, he derives a functional form for the density that supports an affine premium as a solution to the principal's problem. Unfortunately, the implied density is U-shaped, so that the mass is concentrated in the tails. His example shows that the premium need not be strictly convex in the quantity bought. And if the intercept of his affine premium were positive, it would also show that quantity discounts are consistent with adverse selection for a monopolist insurer (though for an implausible type distribution). We pin down the shape of the premium schedule exactly under mild conditions, and we argue that quantity discounts are consistent with plausible preferences and type distributions.

Let $(x(\theta), t(\theta))_{\theta\in\Theta}$ be an optimal menu. Since the coverage cannot increase unless the premium increases, there is an increasing function $T(\cdot)$ on $[x(\underline{\theta}), \ell]$ such that $t(\theta) = T(x(\theta))$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$.[19] We want to know when $T(x)/x$ is nonincreasing. It is simpler to determine when $T(\cdot)$ is concave, so we begin with that issue. To simplify the notation we set $u_\ell = u(w-\ell+x(\theta)-t(\theta))$ and $u_n = u(w-t(\theta))$, where the subscript denotes that the utility function is evaluated at the level of wealth in the loss (ℓ) or no-loss (n) state for type θ in the optimal menu.

It is easy to show that $T(\cdot)$ cannot be concave if a positive measure of agents pool at any (x, t) with $x > 0$.[20] So we assume from now on that the optimal menu sorts types completely (just as it is done for curvature results in the standard monopoly model).

Since sorting is complete, $x(\cdot)$ is strictly increasing, so it has an inverse, call it $z(\cdot)$ (i.e., $\theta = z(x)$). We can now describe an optimal menu as a *nonlinear premium schedule* $T(x) = t(z(x))$. By the first-order condition, the slope of $T(\cdot)$ at $x(\theta)$ equals type-θ's marginal rate of substitution of x for t (see (45) in Section A.8):

$$T'(x(\theta)) = -U_x(x(\theta), t(\theta), \theta)/U_t(x(\theta), t(\theta), \theta). \tag{16}$$

We now give a necessary and sufficient condition for $T(\cdot)$ to be strictly concave locally.

LEMMA 3 (Curvature). *Let $T(\cdot)$ be an optimal nonlinear premium schedule that completely sorts types. We have $T''(x(\theta)) < 0$ if and only if*

$$\frac{f'(\theta)}{f(\theta)} > \frac{3\theta - 2 + c(\theta)}{\theta(1-\theta)}, \tag{17}$$

[19]If $x(\underline{\theta}) > 0$, we extend $T(\cdot)$ to all of $[0, \ell]$ by setting $T(x) = R(x)$, where $R(x)$ solves $U(x, R(x), \underline{\theta}) = U(0, 0, \underline{\theta})$ on $[0, x(\underline{\theta})]$. If $x(\underline{\theta}) > 0$, then $T(\cdot)$ is concave on $[0, x(\underline{\theta})]$ and differentiable at $x(\underline{\theta})$.

[20]Suppose that $x(\theta_0) = x(\theta_1) = \widetilde{x} > 0$ with $\theta_1 > \theta_0$. Then $T'(\widetilde{x}^-) \le \theta_0 u'_\ell/(\theta_0 u'_\ell + (1-\theta_0)u'_n) < \theta_1 u'_\ell/(\theta_1 u'_\ell + (1-\theta_1)u'_n) \le T'(\widetilde{x}^+)$, so $T(\cdot)$ cannot be concave.

where $c(\theta) = \theta u_n'' u_\ell'^2 / [\theta u_n'' u_\ell'^2 + (1-\theta) u_\ell'' u_n'^2]$.

Since $c(\theta) \in (0,1)$ for all θ, Lemma 3 implies that the premium schedule is concave if $f'(\theta)/f(\theta) > (3\theta - 1)/\theta(1-\theta)$ for all $\theta \in \Theta$. The condition holds, for example, if $f(\cdot)$ is uniform with $\overline{\theta} < 1/3$. This simple uniform example already shows that, unlike a competitive insurer, a monopoly insurer can optimally give global quantity discounts.

An objection to Lemma 3 is that $c(\cdot)$ is endogenous, but in some cases, it gives us a complete description of the curvature of $T(\cdot)$.

Example 5 (Uniform density, log utility). Let $f(\cdot)$ be uniform on $[0, \overline{\theta}]$ with $\overline{\theta} \geq \frac{1}{2}$. By Theorem 3, sorting is complete. For $u(\cdot) = \log(\cdot)$, the function $c(\cdot)$ from Lemma 3 simplifies to the identity function $c(\theta) = \theta$. Since $f'(\theta) = 0$ for all θ, (17) implies that $T(\cdot)$ is backward-S-shaped, concave on $[0, x(\frac{1}{2})]$, and convex on $[x(\frac{1}{2}), \ell]$. Thus $T(\cdot)$ exhibits quantity discounts, at least for small coverage levels. And if $f(\cdot)$ is uniform on just $[0, \frac{1}{2}]$, then $T(\cdot)$ is globally concave and exhibits quantity discounts globally. ◊

The curvature in Example 5 holds far more generally. If $f(\cdot)$ is log-concave, the left side of (17) is decreasing. If the right side were increasing as in the example, the conclusion would follow. Unfortunately, since $c(\cdot)$ is endogenous, it is not easy to find conditions that ensure that the right side of (17) is increasing. But if DARA holds, we show that it crosses the left side at most once from below.

Theorem 4 (Backward-S-shaped premium). *Let $T(\cdot)$ be an optimal schedule that completely sorts types, suppose that f is log-concave, and that DARA holds.*

(i) There is an $\hat{x} \in [x(\underline{\theta}), \ell]$ such that $T(\cdot)$ is concave below $\hat{x}$ and convex above $\hat{x}$.

(ii) If $f'(\cdot)$ takes positive and negative values, then $\hat{x} > x(\underline{\theta})$ if $\underline{\theta} < 1/3$ and $\hat{x} < \ell$ if $\overline{\theta} > 2/3$.

Proof. (i) Denote the right side of (17) by $g(\theta)$. We first show that $c'(\theta) \geq 0$ implies that $g'(\theta) > 0$. Differentiate g to find that

$$g'(\theta) = \frac{c'(\theta)}{\theta(1-\theta)} + \frac{3\theta^2 - 4\theta + 2 - c(\theta)(1-2\theta)}{\theta^2(1-\theta)^2}.$$

Since $c(\theta) \in (0,1)$, it follows that $3\theta^2 - 4\theta + 2 - c(\theta)(1-2\theta) > 2/3 > 0$. Therefore, $c'(\theta) \geq 0$ implies that $g'(\theta) > 0$.

We have $T'(x) = \theta u_\ell' / [(1-\theta) u_n' + \theta u_\ell']$ ((16)). Rearrange to find $(1-\theta) u_n' / \theta u_\ell' = (1/T') - 1 > 0$ and use the equality to rewrite $c(\theta)$ as

$$c(\theta) = \frac{1}{1 + \frac{r_\ell}{r_n}\left(\frac{1}{T'} - 1\right)},$$

where r_i is the Arrow–Pratt risk aversion measure in state $i = \ell, n$, evaluated at the contract for type θ. Differentiate $c(\cdot)$ to find that

$$c'(\theta) = -\Lambda\left[\frac{\partial \frac{r_\ell}{r_n}}{\partial\theta}\left(\frac{1}{T'} - 1\right) - \frac{r_\ell}{r_n}\frac{T'' x'}{T'^2}\right], \tag{18}$$

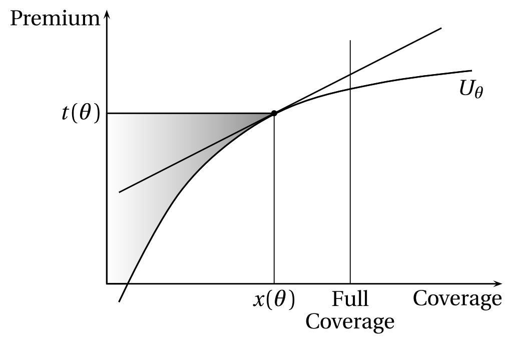

Figure 1. If an interior type gets close enough to its first-best contract, then the premium cannot be convex below or concave above that type's coverage.

where $\Lambda = (1 + \frac{r_\ell}{r_n}(\frac{1}{T'} - 1))^{-2}$. Since the menu is increasing in θ, DARA implies that $\partial \frac{r_\ell}{r_n}/\partial\theta \leq 0$. By (18), if $T''(x(\theta_0)) \geq 0$, then $c'(\theta_0) \geq 0$ and so $g'(\theta_0) > 0$. Thus, $g(\cdot)$ crosses the decreasing function $f'(\cdot)/f(\cdot)$ at most once from below, so there is an interval $(\hat{\theta}, \overline{\theta}]$ with $T(\cdot)$ convex on the interval $[x(\hat{\theta}), x(\overline{\theta})]$ and concave otherwise. Setting $\hat{x} = x(\hat{\theta})$ completes the proof that $T(\cdot)$ is backward-S-shaped.

(ii) The result follows from (17) and $c(\theta) \in (0, 1)$. □

Theorem 4 is the most surprising result of the paper: despite the complications of common values and wealth effects, it holds under the weak and commonly imposed assumptions of a log-concave density and DARA. We are not aware of such a curvature result in other monopoly pricing models.[21]

Few papers in the monopoly pricing literature give any intuition for curvature results. For some intuition on the role of log-concavity of the density of types in our case, consider Figure 1, which shows a contract for an interior type θ. By (IC), optimal menus are monotone, so contracts given to lower types must lie in the shaded region. If the contract given to type θ is close enough to its first best (the zero-surplus full-insurance contract for that type), then $T(\cdot)$ cannot be convex on $[0, x(\theta)]$. To see this graphically, notice that the indifference curve of this type passes close to the origin, since this type gets little surplus. But strict concavity of indifference curves then rules out convexity of $T(\cdot)$ on $[0, x(\theta)]$. Working backward from this interior type to lower types, incentive compatibility and participation almost force the premium to be concave for the quantities associated with those types. Regarding the interval $[x(\theta), \ell]$, recall that, by (16), $T'(x(\theta))$ equals the slope of the indifference curve of θ at $(x(\theta), t(\theta))$. If $T(\cdot)$ were concave on $[x(\theta), \ell]$, then $T'(x) \leq T'(x(\theta))$ for all $x \in [x(\theta), \ell]$. But if $x(\theta)$ is close to ℓ, then (see (16)) $T'(x(\theta)) \approx \theta < \overline{\theta} = T'(x(\overline{\theta}))$, a contradiction. Why should an interior type have a contract that is close to its first best? First, interior types are relatively more likely if

[21] Spence (1977) considers a nonlinear pricing problem of allocating a good in fixed supply and finds that a tariff chosen to maximize a weighted sum of utilities can be backward-S-shaped. But his curvature result is an artifact of the weights used: with equal weights on types, the tariff is affine. He also examines the tariff chosen by a monopolist when $U = \theta u(x) - t$ with $u(\cdot)$ strictly increasing and concave; the tariff is globally concave under the MHRC.

$f(\cdot)$ is log-concave (with $f'(\cdot)$ changing signs); second, if the the support is wide enough, first-best profit is maximized at some interior type. Since some interior types are both more likely and more profitable than extreme types, it is plausible that optimal menus push contracts for some interior types close to their first-best.

4.4.1 *Private versus common values* It is instructive to compare the curvature result in Theorem 4 with that in Maskin and Riley (1984, p. 172). They assume that type-θ buyer's preferences over quantity–payment pairs (x, t) are of the form $U(x, t, \theta) = v(x, \theta) - t$, where $v(x, \theta) = \int_0^x p(q, \theta)\, dq$, and $p(\cdot, \theta)$ is decreasing in q and $p(q, \cdot)$ is increasing in θ. The cost of selling x units to any type is kx, where k is a positive constant. As already noted, theirs is a private-values model with no wealth effects on demand, in contrast to our insurance model.

Recall that in Maskin and Riley (1984) there are always quantity discounts for the highest types; this is not so for a monopoly insurer. To see why, differentiate $t(\theta)/x(\theta)$ with respect to θ to find that under complete sorting,

$$\frac{d(t(\theta)/x(\theta))}{d\theta} = \frac{x'(\theta)}{x(\theta)}\left(\frac{t'(\theta)}{x'(\theta)} - \frac{t(\theta)}{x(\theta)}\right) \geq \frac{x'(\theta)}{x(\theta)}\left(\theta - \frac{t(\theta)}{x(\theta)}\right) = -\frac{x'(\theta)}{x(\theta)^2}\pi(\theta),$$

with equality if and only if $\theta = \overline{\theta}$. The inequality comes from $t'(\theta)/x'(\theta) = \theta u'_\ell/(\theta u'_\ell + (1-\theta)u'_n) \geq \theta$, with equality if and only if $\theta = \overline{\theta}$ (see (45) in Section A.8). If $x'(\theta)\pi(\theta) < 0$, then revenue per unit, $t(\cdot)/x(\cdot)$, must be rising at θ. And in a small enough neighborhood including the highest type, $t(\cdot)/x(\cdot)$ is decreasing if and only if profit is positive for the highest type. But even in the two-type case, profit from the highest type can be *negative* in an optimal insurance menu. In contrast, in Maskin and Riley (1984), profit from *every* type is nonnegative and is *always* positive for the highest type, which explains the markedly different result we obtain.

They also give mild assumptions under which the optimal price is concave in quantity (Maskin and Riley 1984, Proposition 6, p. 185).[22] To clarify the role that common values play in the curvature of the premium, consider the CARA case (Example 3). As in the derivation of (14), suppress common values by setting the insurer's marginal cost of coverage at $k = \underline{\theta}$, so under CARA we are back in a standard monopoly pricing model with quasilinear utility and private values. It is easy to confirm that Assumptions 1–3 of Maskin and Riley (1984) hold, so by their (25), the necessary and sufficient condition for the premium to be strictly concave in coverage is that, for all θ,

$$\frac{\rho'(\theta)}{\rho(\theta)} + \left[\frac{v_{xx\theta}(x(\theta), \theta)}{v_{xx}(x(\theta), \theta)} - \frac{v_{x\theta\theta}(x(\theta), \theta)}{v_{x\theta}(x(\theta), \theta)}\right] > 0. \tag{19}$$

After tedious algebra, (19) becomes $\rho'(\theta)/\rho(\theta) > (2\theta - 1)/(\theta(1-\theta))$ for all θ. If a density satisfies this inequality, the premium is *globally concave* (implying quantity discounts) in the CARA case with private values. For example, let the type distribution be uniform

[22]Figure 1 provides some intuition for concavity in their case: if the *highest* type gets close enough to its first-best contract, then the nonlinear price is globally concave. Since higher types generate more profit under private values, this property—together with the MHRC—suggests that it is plausible that the highest type gets a contract close to its first best.

on $[0, \overline{\theta}]$. Then $\rho'(\theta)/\rho(\theta) = 1/(\overline{\theta} - \theta) > (2\theta - 1)/(\theta(1-\theta))$ for all θ and (19) holds. But with common values, we know from Theorem 4 that if $f(\cdot)$ is log-concave and $\overline{\theta} > 2/3$, the premium is *convex* in coverage for $\theta \in (2/3, \overline{\theta}]$.

4.4.2 *Using curvature to test for adverse selection in insurance* There is a growing empirical literature that tests for the presence of adverse selection in insurance. Most of it focuses on the prediction that *riskier* types buy *more* coverage, an implication of most competitive models (following Rothschild and Stiglitz 1976) as well as our monopoly model. For example, Puelz and Snow (1994) find evidence of this monotonicity property in automobile insurance in the United States (but Dionne et al. 2001 fail to find monotonicity using the same data but a different method). By contrast, Cawley and Philipson (1999) do not find monotonicity in life insurance data from the United States,[23] and Chiappori and Salanié (2000) do not find monotonicity in automobile insurance in France (but Cohen 2005 finds monotonicity in French auto insurance using data for more experienced drivers). Finkelstein and Poterba (2004) do not find evidence of monotonicity between coverage amount and risk type in the U.K. annuity market, although they do find evidence of adverse selection on other contract dimensions. Chiappori et al. (2006) find evidence of monotonicity for (again) automobile insurance in France; they also provide a detailed discussion of monotonicity and its robustness across different settings.

A second test is based on *curvature*. In Rothschild and Stiglitz (1976), the equilibrium premium for a type θ is $t(\theta) = \theta x(\theta)$, where $x(\cdot)$ is the (increasing) equilibrium indemnity function. Hence $t(\theta)/x(\theta) = \theta$ and there are *always* quantity premia.[24] Several prominent papers use the quantity premia implication to test for adverse selection. Puelz and Snow (1994) find evidence of quantity premia in automobile insurance; Finkelstein and Poterba (2004) find evidence of mild quantity discounts together with a slightly increasing marginal price—a shape consistent with our Theorem 4, but not with competitive models of adverse selection.[25]

Perhaps the best known (certainly the most widely cited) paper that tests the curvature prediction is Cawley and Philipson (1999). They regress the premium on a quadratic function of the coverage amount and find that the coefficient on the squared term is zero, and that the intercept is positive.[26] They conclude that the estimated affine function—which implies quantity *discounts*—is evidence *against* adverse selection in life insurance. Theorem 4 shows that quantity premia are not an implication of adverse selection as such, but of the *joint* imposition of adverse selection and (some form

[23] He (2009) argues that selection bias accounts for the failure of monotonicity in Cawley and Philipson (1999): they use a cross section of the data and so exclude people who already bought policies, but died before the sample period. Using the same data, but taking into account this possible selection bias, He (2009) finds that monotonicity holds: riskier types buy more coverage.

[24] See Wilson (1993, pp. 382–384), for another insurance example with quantity premia.

[25] Quantity discounts are sometimes explained informally by a fixed cost of providing a good or writing a contract. For a monopolist, a fixed cost just affects whether a contract is offered, not its shape.

[26] Finkelstein and Poterba (2004) also regress payment on a quadratic function of the annuity amount. They find that the coefficient on the squared term is positive, but small in magnitude, and the intercept is positive, consistent with quantity discounts.

of) perfect competition. Indeed, even with a convex segment at the end, our premium shape could exhibit *global* quantity discounts.

Since recent papers do find evidence of market power in insurance,[27] researchers should be cautious about using curvature to rule out adverse selection. Absent evidence on competition, we think that the main focus of tests for adverse selection should be on the monotonicity prediction that those who buy more coverage suffer more losses.

5. Conclusion

Stiglitz (1977) introduced the insurance model that we examine and solved the two-type case with an illuminating graphical analysis that is now a textbook standard. Despite the importance of adverse selection in insurance and well known problems with its competitive provision, the monopoly case has received surprisingly little attention. We have derived several important properties of optimal menus in this setting, including a sharp result on curvature that is also of empirical interest.

Insurance markets surely lie in between competition and monopoly. Monopoly seems to be the right place to start thinking about noncompetitive markets: for insurance with adverse selection, there is no agreement on what a good model of competition is, let alone of oligopoly. Clearly, more work is needed on models of imperfectly competitive markets with adverse selection.

There are several other extensions worth pursuing. For example, one could allow for more than one loss amount. If the private information is still purely about likelihood of a loss, not its magnitude, then it is not hard to prove that the principal will offer a menu of deductible contracts. Since a deductible contract is still two-dimensional, many of our proofs extend to this case. A multiperiod version of the model raises issues such as renegotiation and experience rating, and should reveal further implications of adverse selection for insurance. Finally, we have extended some general properties of optimal menus to the cases of nonexpected utility and to private information about risk preferences (rather than risk), but much more work remains to gain a complete understanding of these cases.

Appendix

A.1 *Proof of Theorem 1 (Properties of an optimal menu)*

In the proof of Theorem 1, we use the following result:

Lemma 4 (Indirect utility function). *Let $(x(\cdot), t(\cdot))$ be a bounded menu that satisfies (IC), with $x(\theta) \leq \ell$ for all $\theta \in \Theta$, and let $V(\theta) = U(x(\theta), t(\theta), \theta)$ be the indirect utility of type θ from that menu. Then $V(\cdot)$ is decreasing and continuous on Θ.*

[27] See Chiappori et al. (2006), Cohen and Einav (2007), and Dafny (2010). For instance, Cohen and Einav (2007) estimate the demand for insurance for a new entrant into the Israeli automobile insurance market. They argue that this firm has market power and that a monopoly insurance model better describes this situation than a competitive one.

Proof. Let $\theta' > \theta$. We have $V(\theta) \geq U(x(\theta'), t(\theta'), \theta) \geq V(\theta')$, where the first inequality follows from (IC) and the second follows from $x(\theta') \leq \ell$. Hence $V(\theta)$ is decreasing in θ.

Monotonicity implies that the left and right limits exist at any $\theta \in \Theta$. Let $\theta' \in \Theta$. We show that the left and right limits of $V(\cdot)$ are equal at $\theta = \theta'$. Consider any sequence θ_n approaching θ' from below, and let $t_n = t(\theta_n)$ and $x_n = x(\theta_n)$. We have

$$0 \geq V(\theta') - V(\theta_n) \geq U(x_n, t_n, \theta') - V(\theta_n), \tag{20}$$

where the first inequality follows from monotonicity of $V(\cdot)$ and the second follows from (IC). But $U(x_n, t_n, \theta') - V(\theta_n) = (\theta' - \theta_n)(u(w - \ell - t_n + x_n) - u(w - t_n))$. Since $(x(\cdot), t(\cdot))$ is bounded and $u(\cdot)$ is continuous, $U(x_n, t_n, \theta') - V(\theta_n)$ tends to 0 as θ_n tends to θ', so by (20), $V(\cdot)$ is left-continuous at $\theta = \theta'$. Now consider any sequence θ_n approaching θ' from above. For every n,

$$0 \geq V(\theta_n) - V(\theta') \geq U(x(\theta'), t(\theta'), \theta_n) - V(\theta'),$$

where the first inequality follows from monotonicity in θ and the second follows from (IC). But again $U(x(\theta'), t(\theta'), \theta_n) - V(\theta')$ tends to zero as θ_n tends to θ' and so $V(\theta_n)$ converges to $V(\theta')$ and so $V(\cdot)$ is right-continuous at $\theta = \theta'$. Since θ' was arbitrary, it follows that $V(\cdot)$ is continuous on Θ. □

We prove Theorem 1 in several steps. Most are by contraposition: we show that if the property fails for a positive measure of types in a feasible menu, then there is another feasible menu with higher profit (relying on Lemma 1).

For a feasible menu $(x(\cdot), t(\cdot))$, let $R(\cdot)$ be its *revenue function*: for $x \in \mathbb{R}$, $R(x)$ is the (possibly extended) real number

$$R(x) = \sup\{\tau \in \mathbb{R} \mid V(\theta) = U(x, \tau, \theta) \text{ for some } \theta \in \Theta\}. \tag{21}$$

This function is the upper envelope of the indifference curves that pass through contracts in the menu. Clearly it is increasing.

(i) *No overinsurance.* Let $(x(\cdot), t(\cdot))$ be a feasible (hence increasing) menu with $x(\theta) > \ell$ on a set $\Theta_+ \subset \Theta$ of positive measure. Since $x(\cdot)$ is increasing, Θ_+ is the intersection of Θ with an interval with right endpoint $\overline{\theta}$. Let R be the revenue function ((21)) for this menu. Pool all types in Θ_+ at $(\ell, R(\ell))$ and leave the contracts of all other types unchanged. By the definition of R, the new menu is feasible, and by Lemma 1, expected profit increases for all types in the positive measure set Θ_+, so that the original menu does not solve the principal's problem.

(ii) *Premium, indemnity, and net indemnity are nonnegative.* It is enough to show that $t(\theta) \geq 0$ for almost all types: if that condition holds, then (P) implies that $x(\theta) \geq 0$ and $x(\theta) - t(\theta) \geq 0$ for almost all types. Let $(x(\cdot), t(\cdot))$ be a feasible menu with $t(\theta) < 0$ on a set $\Theta_- \subset \Theta$ of positive measure. By part (i), we may take $x(\overline{\theta}) \leq \ell$. By monotonicity, Θ_- is the intersection of Θ and a left-closed interval with left endpoint $\underline{\theta}$. Now pool all types in Θ_- at the contract $(x', 0)$, where x' is the minimum of ℓ and the infimum of $\{\chi \in \mathbb{R}_+ \mid V(\theta) = U(\chi, 0, \theta) \text{ for some } \theta \in \Theta\}$ (if this set is empty, the infimum equals infinity), and leave the contracts for all other types unchanged. The resulting menu satisfies (P) and (IC). Since the original menu satisfies (IC), $x(\theta) < x' \leq \ell$ for each $\theta \in \Theta_-$. Also, by the definition of x', the move makes each type in Θ_- at least weakly worse off. By Lemma 1, the change increases expected profit.

By (i) and (ii), any solution to the principal's problem is almost everywhere equal to a solution that is nonnegative with $x(\theta) \leq \ell$. For the rest of the proof, it suffices to consider only feasible menus that are nonnegative with $x(\theta) \leq \ell$.

(iii) *Participation binds at the bottom*. Let $(x(\cdot), t(\cdot))$ be a nonnegative feasible menu with $x(\theta) \leq \ell$ and $U(x(\underline{\theta}), t(\underline{\theta}), \underline{\theta}) - U(0, 0, \underline{\theta}) = K > 0$. Change the menu to reduce the utility of each type by K in each state. Clearly (IC) continues to hold. Since (P) holds for the lowest type in the new menu, it holds for every type in the new menu by the SSCP, and since the premium increases and the net indemnity goes down for every type, profit from each type increases.

(iv) *Full insurance at the top*. Let $(x(\cdot), t(\cdot))$ be a nonnegative feasible menu with $x(\overline{\theta}) < \ell$.[28] Let τ and $\widetilde{\tau}$ solve $\tau - \overline{\theta}\ell = t(\overline{\theta}) - \overline{\theta}x(\overline{\theta})$ and $U(\ell, \widetilde{\tau}, \overline{\theta}) = V(\overline{\theta})$. In words, (ℓ, τ) is a contract with the same expected profit from $\overline{\theta}$ as $(x(\overline{\theta}), t(\overline{\theta}))$, and $(\ell, \widetilde{\tau})$ is a contract that is indifferent to $(x(\overline{\theta}), t(\overline{\theta}))$ for type $\overline{\theta}$.

Notice that $\widetilde{\tau} > \tau$. Let $\overline{\tau} = (\tau + \widetilde{\tau})/2 > \tau$ and consider $(\ell, \overline{\tau})$. By construction, $\overline{\tau} - \overline{\theta}\ell > t(\overline{\theta}) - \overline{\theta}x(\overline{\theta})$. Since $x(\overline{\theta}) < \ell$, we have $\overline{\tau} - \theta\ell > t(\overline{\theta}) - \theta x(\overline{\theta})$ for every $\theta \in \Theta$. By Lemma 1, $x(\theta) < \ell$, and by the monotonicity of the menu, $t(\overline{\theta}) - \theta x(\overline{\theta}) > t(\theta) - \theta x(\theta)$ for every $\theta \in \Theta$. So for each type θ, expected profit at $(\ell, \overline{\tau})$ is higher than at $(x(\theta), t(\theta))$.

Define $m(\theta) = U(\ell, \overline{\tau}, \theta) - V(\theta)$. It follows that $m(\overline{\theta}) > 0$ (by definition of $\overline{\tau}$), and $m(\cdot)$ is continuous (by Lemma 4) and increasing in θ (by the single-crossing property). Hence, the set of types such that $m(\theta) \geq 0$ is of the form $[\overline{\theta} - \delta, \overline{\theta}] \cap \Theta$, with $\delta \in (0, \overline{\theta} - \underline{\theta})$, which is of positive measure. Now pool all types in this set at $(\ell, \overline{\tau})$, leaving the contracts of the other types unchanged. The new contract is feasible by construction and gives higher expected profit than the menu with $x(\overline{\theta}) < \ell$.

(v) *No pooling at the top*. Suppose that the set of types below $\overline{\theta}$ with $x = \ell$ is of positive measure for some nonnegative feasible menu $(x(\theta), t(\theta))_{\theta \in \Theta}$ with $x(\theta) \leq \ell$ for all $\theta \in \Theta$. By (IC), any such type is charged the same payment, call it τ. Let $\hat{\theta}$ be the infimum of the set of types that are indifferent between $(x(\theta), t(\theta))$ and (ℓ, τ). W.l.o.g., we can set $x(\theta) = \ell$ and $t(\theta) = \tau$ for all types $\theta \geq \hat{\theta}$, since expected profit does not fall, and (IC) and (P) still hold. With this change, the set of types receiving $x = \ell$ is $[\hat{\theta}, \overline{\theta}] \cap \Theta$. For all other types, $V(\theta) > U(\ell, \tau, \theta)$ and $x(\theta) < \ell$.

Fix $\varepsilon \in (0, \overline{\theta} - \hat{\theta})$. For each $\delta > 0$, let (x_δ, t_δ) be the contract that leaves type $\hat{\theta}$ indifferent between (ℓ, τ) and (x_δ, t_δ), and leaves type $\hat{\theta} + \varepsilon$ indifferent between $(\ell, \tau + \delta)$ and (x_δ, t_δ).[29] Formally, (x_δ, t_δ) solves (see Figure 2)

$$U(\ell, \tau, \hat{\theta}) = U(x_\delta, t_\delta, \hat{\theta})$$
$$U(\ell, \tau + \delta, \hat{\theta} + \varepsilon) = U(x_\delta, t_\delta, \hat{\theta} + \varepsilon).$$

[28] We could finish off the proof as follows. If there is no mass point at $\overline{\theta}$, then there is nothing to prove: the menu is almost everywhere equal to a menu with full insurance at the top. If there is a mass point at $\overline{\theta}$, then the principal could set $x(\overline{\theta}) = \ell$, leaving the contract for other types unchanged; the new menu is feasible and the principal is better off by Lemma 1. We show a *stronger* result: the essential supremum of $x(\cdot)$ must equal ℓ for any solution; otherwise, the principal can strictly improve upon it.

[29] It is possible, but irrelevant, that $\hat{\theta} + \varepsilon$ is not in Θ.

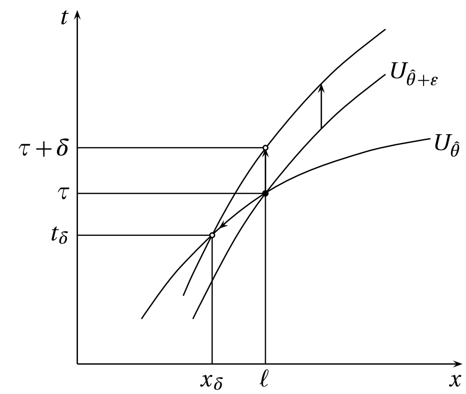

Figure 2. The menu constructed in the proof of the no pooling at the top property.

For every $\varepsilon \in (0, \overline{\theta} - \hat{\theta})$, there is a $d(\varepsilon) > 0$ such that if $\delta \in [0, d(\varepsilon)]$, a solution exists. Define a menu of contracts $(\hat{x}(\theta), \hat{t}(\theta))_{\theta \in \Theta}$ as

$$(\hat{x}(\theta), \hat{t}(\theta)) = \begin{cases} (\ell, \tau + \delta) & \text{if } \theta \in (\hat{\theta} + \varepsilon, \overline{\theta}] \cap \Theta \\ (x_\delta, t_\delta) & \text{if } \theta \in [\hat{\theta}, \hat{\theta} + \varepsilon] \cap \Theta \\ (\widetilde{x}_\delta(\theta), \widetilde{t}_\delta(\theta)) & \text{if } \theta \in [\underline{\theta}, \hat{\theta}) \cap \Theta, \end{cases}$$

where $(\widetilde{x}_\delta(\theta), \widetilde{t}_\delta(\theta)) = (x_\delta, t_\delta)$ if type θ strictly prefers (x_δ, t_δ) to $(x(\theta), t(\theta))$ and $(x(\theta), t(\theta))$ otherwise. The menu $(\hat{x}(\theta), \hat{t}(\theta))_{\theta \in \Theta}$ satisfies (IC) and also satisfies (P) if $\delta \in [0, \tau(\varepsilon) - \tau]$, where $\tau(\varepsilon)$ is given by $U(\ell, \tau(\varepsilon), \hat{\theta} + \varepsilon) = U(0, 0, \hat{\theta} + \varepsilon)$. (Since $U(\ell, \tau, \hat{\theta}) \geq U(0, 0, \hat{\theta})$, the SSCP implies that $\tau(\varepsilon) > \tau$.)

The expected profit from menu $(\hat{x}(\theta), \hat{t}(\theta))_{\theta \in \Theta}$ is

$$\int_{(\hat{\theta}+\varepsilon, \overline{\theta}]} [\tau + \delta - \theta \ell]\, dF(\theta) + \int_{[\hat{\theta}, \hat{\theta}+\varepsilon]} [t_\delta - \theta x_\delta]\, dF(\theta) + \int_{[\underline{\theta}, \hat{\theta})} [\widetilde{t}_\delta(\theta) - \theta \widetilde{x}_\delta(\theta)]\, dF(\theta). \quad (22)$$

We next show that the first two integrals are each differentiable in δ at $\delta = 0$ and that, for small enough $\varepsilon > 0$, the derivative of the sum is positive. The derivative of the first integral at $\delta = 0$ is $1 - F(\hat{\theta} + \varepsilon) > 0$ for every $\varepsilon \in [0, \overline{\theta} - \hat{\theta})$.

Consider the second integral. Since u is C^1 with $u' > 0$, U_x/U_t is strictly increasing in θ, so x_δ and t_δ are continuously differentiable in δ for every $\delta \in I(\varepsilon) =: [0, \min\{\tau(\varepsilon) - \tau, d(\varepsilon)\}]$. The derivative $t'_\delta - \theta x'_\delta$ with respect to δ is bounded on $I(\varepsilon) \times \Theta$, and some algebra reveals that

$$0 \geq t'_0 - \theta x'_0 = \frac{\hat{\theta} - \theta}{\varepsilon} \geq -1 \quad (23)$$

for $\theta \in [\hat{\theta}, \hat{\theta} + \varepsilon]$. So the derivative of the second integral in (22) with respect to δ exists at $\delta = 0$, equals $\int_{[\hat{\theta}, \hat{\theta}+\varepsilon]} (\hat{\theta} - \theta)/\varepsilon\, dF(\theta)$, and satisfies

$$0 \geq \int_{[\hat{\theta}, \hat{\theta}+\varepsilon]} \frac{\hat{\theta} - \theta}{\varepsilon}\, dF(\theta) \geq F(\hat{\theta}) - F(\hat{\theta} + \varepsilon).$$

Since $F(\cdot)$ is right-continuous, the last expression tends to zero as ε tends to zero, so the derivative of the second term in (22) at $\delta = 0$ converges to zero as ε converges to 0. So the first two integrals are differentiable in δ at $\delta = 0$ and the sum of the derivatives is positive for small enough $\varepsilon > 0$. For the rest of the proof, fix such a ε.

Consider the third integral in (22). Let $\Theta^+(\delta)$ be the set of types in $[\underline{\theta}, \hat{\theta}) \cap \Theta$ that strictly prefer (x_δ, t_δ) to $(x(\theta), t(\theta))$. Since $(x_\delta, t_\delta) \to (\tau, \ell)$, $V(\theta) > U(\ell, \tau, \theta)$ for all $\theta \in [\underline{\theta}, \hat{\theta}) \cap \Theta$, and V and U are continuous, it follows that $\Theta^+(\cdot)$ is a decreasing sequence of sets with empty intersection, so $\lim_{\delta \to 0} \int_{\Theta^+(\delta)} dF(\theta) = 0$. For $\theta \in [\underline{\theta}, \hat{\theta}) \cap \Theta$,

$$\frac{\widetilde{t}_\delta(\theta) - \theta \widetilde{x}_\delta(\theta) - t(\theta) + \theta x(\theta)}{\delta} \geq \frac{t_\delta - \theta x_\delta - \tau + \theta \ell}{\delta} \tag{24}$$

since, by Lemma 1, $\tau - \theta\ell > t(\theta) - \theta x(\theta)$, and the right side is negative. (The left side is either 0 or greater than the right side when nonzero.) Integrate over $[\underline{\theta}, \hat{\theta})$ and use (24) to find

$$\begin{aligned} \int_{[\underline{\theta}, \hat{\theta})} \frac{\widetilde{t}_\delta(\theta) - \widetilde{x}_\delta(\theta) - t(\theta) + \theta x(\theta)}{\delta}\, dF &= \int_{\Theta^+(\delta)} \frac{\widetilde{t}_\delta(\theta) - \theta \widetilde{x}_\delta(\theta) - t(\theta) + \theta x(\theta)}{\delta}\, dF \\ &\geq \int_{\Theta^+(\delta)} \frac{t_\delta - \theta x_\delta - \tau + \theta \ell}{\delta}\, dF. \end{aligned}$$

Since t_δ and x_δ are differentiable in δ on $I(\varepsilon)$, the mean value theorem implies that the integrand in the last integral equals $t'_{\delta(\varepsilon,\theta)} - \theta x'_{\delta(\varepsilon,\theta)}$ for some $\delta(\varepsilon, \theta) \in I(\varepsilon)$, and since the derivatives are continuous on $I(\varepsilon)$, the integrand is bounded below by some number M. So the last integral is at least as large as $M \int_{\Theta^+(\delta)} dF$, which tends to 0 as δ converges to 0. Thus

$$\liminf_{\delta \to 0^+} \int_{[\underline{\theta}, \hat{\theta})} \frac{\widetilde{t}_\delta(\theta) - \theta \widetilde{x}_\delta(\theta) - t(\theta) - \theta x(\theta)}{\delta}\, dF(\theta) \geq 0. \tag{25}$$

Letting $\Pi(\delta)$ equal (22), the expected profit from menu $(\hat{x}(\cdot), \hat{t}(\cdot))$, we have shown that

$$\liminf_{\delta \to 0^+} \frac{\Pi(\delta) - \Pi(0)}{\delta} > 0,$$

implying that the original menu does not solve the principal's problem.

(vi) *Positive profit*. Let $\varepsilon > 0$ and consider a menu in which each type chooses between $(0, 0)$ or $(\ell, \overline{\theta}\ell + \varepsilon)$ to maximize expected utility. Clearly, expected profit is positive from any type who chooses $(\ell, \overline{\theta}\ell + \varepsilon)$. Since $\overline{\theta} < 1$ and the agent is strictly risk averse, $U(\ell, \overline{\theta}\ell, \overline{\theta}) > U(0, 0, \overline{\theta})$, and since U is continuous, it follows that $U(\ell, \overline{\theta}\ell + \varepsilon, \theta) > U(0, 0, \overline{\theta})$ on a neighborhood of $(\theta, \varepsilon) = (\overline{\theta}, 0)$. So for $\varepsilon > 0$ small enough, a positive measure of types strictly prefer $(\ell, \overline{\theta}\ell + \varepsilon)$ to $(0, 0)$.

A.2 *Proof of Theorem 1 for the general insurance model*

We denote by $U(x, t, \theta)$ a type-$\theta \in \Theta \subset (0, 1)$ agent's utility for a contract (x, t), but we do not assume that θ is the probability of a loss or that the agent has expected utility preferences. Set $D = \{(x, t) \mid (w - t, w - \ell - t + x) \geq 0\}$ and $D_+ = D \cap \mathbb{R}^2_+$.

THEOREM 5. *Suppose that the principal's unit cost of selling to type θ equals $k(\theta)$, with $k(\cdot)$ nondecreasing and continuous on Θ. Let the function $U(\cdot,\cdot,\cdot): D \times [0,1]$ satisfy the following assumptions:*

(a) $U(\cdot,\cdot,\theta)$ is strictly quasiconcave and continuous in (x,t) on D for each θ, and it is strictly increasing in x and strictly decreasing in t.

(b) $\{U(x,t,\cdot)\}_{(x,t)\in C}$ is uniformly equicontinuous for any compact subset C of D.[30]

(c) For each $(x,t)\in D$, $U(x,t,\cdot)$ is monotone (increasing or decreasing) in θ.

(d) The unique Pareto optimum quantity is ℓ for every $\theta\in\Theta$.

(e) U satisfies the strict single-crossing property (SSCP) in (x,t) and θ on D or it satisfies the SSCP on D_+ and contracts are restricted to lie in D_+.

Then the conclusion of Theorem 1, parts (i)–(iv) and (vi) hold. If, in addition, U is C^1 with $U_t<0$ and satisfies the Spence–Mirrlees SSCP that $-U_x/U_t$ is strictly increasing in θ, then the conclusion of Theorem 1(v) holds.

It is easy to verify that these assumptions are satisfied in our model. Conditions (a) and (c)–(e) clearly hold. Regarding (b), for any compact set of contracts C, $|U(x,t,\theta')-U(x,t,\theta'')| = |u(w-\ell+x-t)-u(w-t)||\theta'-\theta''| \le (\max_{(x,t)\in C}|u(w-\ell+x-t)-u(w-t)|)|\theta'-\theta''| = M|\theta'-\theta''|$ and uniform equicontinuity follows.

We repeatedly invoke Lemma 4 in the proof of Theorem 1, but our proof of the lemma uses linearity of expected utility in the loss chance. We use assumption (b) to extend the conclusion of Lemma 4 (continuity of V) to the general case.[31]

PROOF OF THEOREM 5. It is easy to check that, after we change the principal's unit cost of selling from θ to $k(\theta)$, the only changes we need to make to the proof of Theorem 1 are to parts (iii) (Participation binds at the bottom) and (v) (No pooling at the top). For part (v), some calculations change, notably that (23) now becomes

$$0 \ge t_0' - k(\theta)x_0' = \frac{k(\hat\theta)-k(\theta)}{k(\hat\theta+\varepsilon)-k(\hat\theta)} \ge -1,$$

but the changes are straightforward to confirm.

The argument for part (iii) exploits the assumption that the agent has expected utility preferences. In the general case, the argument is slightly more involved. Let $(x(\cdot),t(\cdot))$ be a nonnegative feasible menu with $U(x(\underline{\theta}),t(\underline{\theta}),\underline{\theta}) > U(0,0,\underline{\theta})$ and $x(\theta)\le\ell$ for all $\theta\in\Theta$. As just observed, the indirect utility $V(\cdot)$ of the menu is continuous in θ.

[30]A family $\mathcal{F}$ of functions f defined on an interval $[a,b]$ is uniformly equicontinuous if, given any $\varepsilon>0$, there exists a $\delta>0$ such that $|x'-x''|<\delta$ implies $|f(x')-f(x'')|<\varepsilon$ for all $x',x''\in[a,b]$ and all $f\in\mathcal{F}$.

[31]Uniform equicontinuity is used in the last step of the proof of the left- and right-continuity of $V(\cdot)$, i.e., the proof of $\lim_{\theta_n\to\theta'_-}U(x_n,t_n,\theta')-V(\theta_n) = \lim_{\theta_n\to\theta'_+}U(x_n,t_n,\theta')-V(\theta_n)=0$. To see this, let $\varepsilon>0$. Then there is a $\delta>0$ such that $|\theta'-\theta_n|<\delta$ implies $|U(x,t,\theta')-U(x,t,\theta_n)|<\varepsilon$ for any (x,t) in C. Thus, $|U(x_n,t_n,\theta')-V(\theta_n)| \le \max_{(x,t)\in C}|U(x,t,\theta')-U(x,t,\theta_n)|<\varepsilon$ for n large enough.

Let $\delta > 0$ satisfy $U(x(\underline{\theta}), t(\underline{\theta}) + \delta, \underline{\theta}) = U(0, 0, \underline{\theta})$. Define $R_1(x)$ on $[x(\underline{\theta}), \ell]$ by

$$R_1(x) = \max\{R(x), r(x)\},$$

where $R(\cdot)$ is the revenue function (21) for the menu $(x(\cdot), t(\cdot))$ and $r(\cdot)$ is defined implicitly by

$$U(x, r(x), \underline{\theta}) = U(x(\underline{\theta}), t(\underline{\theta}) + \delta, \underline{\theta}).$$

Now define a new menu $(x_1(\cdot), t_1(\cdot))$ as follows. Set $(x_1(\underline{\theta}), t_1(\underline{\theta})) = (x(\underline{\theta}), t(\underline{\theta}) + \delta)$. For $\theta \in \Theta - \{\underline{\theta}\}$, set $(x_1(\theta), t_1(\theta)) = (x(\theta), t(\theta))$ if $x(\theta)$ maximizes $U(x, R_1(x), \theta)$ on $[x(\underline{\theta}), \ell]$; otherwise set $(x_1(\theta), t_1(\theta)) = (x', R_1(x'))$ for any maximizer x' of $U(x, R_1(x), \theta)$ on $[x(\underline{\theta}), \ell]$. In the second case, notice that $x(\theta) < x' \le \ell$ and $U(x', R_1(x'), \theta) \le V(\theta)$, so by Lemma 1, expected profit rises from any type for which $(x_1(\theta), t_1(\theta)) \neq (x(\theta), t(\theta))$. Let Θ_1 denote the set of types for which $(x_1(\theta), t_1(\theta)) \neq (x(\theta), t(\theta))$.

We now show that Θ_1 has positive measure. Since $\underline{\theta}$ is the left endpoint of the support, the set $[\underline{\theta}, \underline{\theta} + \nu) \cap \Theta$ has positive measure for every $\nu > 0$. It suffices to show that for ν small enough, $[\underline{\theta}, \underline{\theta} + \nu) \cap \Theta \subset \Theta_1$.

Let $V_1(\cdot)$ be the indirect utility for the menu $(x_1(\cdot), t_1(\cdot))$. Since $V_1(\cdot)$ and $V(\cdot)$ are continuous and $V(\underline{\theta}) > V_1(\underline{\theta})$, there is a $\nu^* > 0$ such that if $\theta \in \Theta \cap [\underline{\theta}, \underline{\theta} + \nu^*)$, then $V(\theta) > V_1(\theta)$. Thus, $[\underline{\theta}, \underline{\theta} + \nu^*) \cap \Theta \subset \Theta_1$, and Θ_1 has positive measure. □

Remark. Theorem 5 also covers (with minor changes) the quasilinear setting of Maskin and Riley (1984) with the addition of common values, i.e., $U(x, t, \theta) = v(x, \theta) - t$ and $\pi(\theta) = t - k(\theta)x$, where $k(\cdot)$ is strictly increasing in θ. Assuming that the first-best quantity $x^*(\cdot)$ is strictly increasing in θ, Theorem 1 can be adapted to this case. For instance, no overinsurance becomes no quantity overprovision and its proof follows along the same lines; nonnegativity is assumed in this setting; no surplus at the bottom is trivial due to linearity in t; the proof of efficiency at the top holds almost unaltered after replacing ℓ with $x^*(\overline{\theta})$, and similarly with the proof of no pooling at the top; finally, positive profit holds if, e.g., the highest type generates positive surplus.

A.3 *Proof of Theorem 2*

Since (iii) follows from (ii), we just prove (i) and (ii). Fix a feasible menu $(x(\theta), t(\theta))_{\theta \in \Theta}$, with $0 \le x(\theta) \le \ell$ for all $\theta \in \Theta$. (By Theorem 1, any optimal menu satisfies these conditions almost everywhere.) Recall that by Lemma 1, $t(\theta) - \theta' x(\theta) \ge t(\theta') - \theta' x(\theta')$ if $\theta > \theta'$; i.e., profit increases if a type takes the contract offered to a higher type.

(i) *The principal prefers a larger loss size.* Let $\ell < \widetilde{\ell} < w$ and let $\widetilde{U}(x, t, \theta)$ be the expected utility of a type-θ agent for contract (x, t) when the loss is $\widetilde{\ell}$ (and let $U(x, t, \theta)$ be the expected utility of a type-θ agent for (x, t) when the loss is ℓ). Fix $\theta' \in \Theta$, and let (χ, τ) be any nonnegative contract bounded above by $(x(\theta'), t(\theta'))$ and no better than $(x(\theta'), t(\theta'))$ for type θ' when the loss is ℓ. Formally,

(a) $(0, 0) \le (\chi, \tau) \le (x(\theta'), t(\theta'))$; and

(b) $U(\chi, \tau, \theta') \le U(x(\theta'), t(\theta'), \theta')$.

These inequalities hold if we set (χ, τ) equal to any *lower* type's contract in the menu $(x(\theta), t(\theta))_{\theta\in\Theta}$, or to the null contract $(0, 0)$.

Claim 1. Inequalities (a) and (b) imply $\widetilde{U}(x(\theta'), t(\theta'), \theta') \geq \widetilde{U}(\chi, \tau, \theta')$, and the inequality is strict if $(\chi, \tau) \neq (x(\theta'), t(\theta'))$.

If $(\chi, \tau) = (x(\theta'), t(\theta'))$, there is nothing to prove, so suppose that $(\chi, \tau) \neq (x(\theta'), t(\theta'))$, which, by (a) and (b), implies that $\chi - \tau < x(\theta') - t(\theta')$. To simplify notation, set $u_\ell(\theta') = u(w - \ell + x(\theta') - t(\theta'))$, $u_n(\theta') = u(w - t(\theta'))$, $u_n = u(w - \tau)$, $u_\ell = u(w - \ell + \chi - \tau)$, $\widetilde{u}_\ell(\theta') = u(w - \widetilde{\ell} + x(\theta') - t(\theta'))$, and $\widetilde{u}_\ell = (w - \widetilde{\ell} + \chi - \tau)$. Rewrite the inequality $U(x(\theta'), t(\theta'), \theta') \geq U(\chi, \tau, \theta')$ as

$$u_\ell(\theta') - u_\ell \geq \frac{1-\theta'}{\theta'}(u_n - u_n(\theta')). \tag{26}$$

The strict concavity of $u(\cdot)$ and the inequality $\chi - \tau < x(\theta') - t(\theta')$ imply that $\widetilde{u}_\ell(\theta') - \widetilde{u}_\ell > u_\ell(\theta') - u_\ell$ and so, by (26), that $(1 - \theta')(\widetilde{u}_n(\theta') - \widetilde{u}_n) + \theta'(\widetilde{u}_\ell(\theta') - \widetilde{u}_\ell) > 0$ or, equivalently, $\widetilde{U}(x(\theta'), t(\theta'), \theta') > \widetilde{U}(\chi, \tau, \theta')$, which proves Claim 1.

Since, for each $\theta, \theta' \in \Theta$ with $\theta < \theta'$ and $(x(\theta), t(\theta)) \neq (x(\theta'), t(\theta'))$, (a) and (b) hold for $(\chi, \tau) = (x(\theta), t(\theta))$ and for $(\chi, \tau) = (0, 0)$, it follows that $(x(\theta), t(\theta))_{\theta\in\Theta}$ continues to satisfy all the *downward* incentive and participation constraints when the loss equals $\widetilde{\ell}$.

Let C be the closure of the set $\{(x(\theta), t(\theta)) \mid \theta \in \Theta\} \cup \{(0, 0)\}$. Consider the problem of choosing a contract in C to maximize $\widetilde{U}(\cdot, \theta)$. (A solution exists since C is compact.) Since $(x(\theta), t(\theta))_{\theta\in\Theta}$ satisfies (IC) when the loss is ℓ, any maximizer (x, t) of $\widetilde{U}(\cdot, \cdot, \theta)$ on C satisfies $U(x, t, \theta) \leq U(x(\theta), t(\theta), \theta)$ (the new choice cannot increase the *pre-change* expected utility for θ). Moreover, C is ordered by the usual vector inequality $\geq$ on $\mathbb{R}^2$. By Claim 1, $(x, t) \geq (x(\theta), t(\theta))$.[32] Hence, by Lemma 1, profit from the new contract does not fall. (By Theorem 1, coverage levels in C are w.l.o.g. bounded above by ℓ.) Consider a menu defined by choosing, for each $\theta \in \Theta$, any maximizer of $\widetilde{U}(\cdot, \cdot, \theta)$. It satisfies (IC) and (P) when the loss is $\widetilde{\ell}$, and is at least as profitable as the original menu.

(ii) *The principal prefers a more risk averse agent.* Let $\widetilde{u}(\cdot)$ be more risk averse than $u(\cdot)$; i.e., $\widetilde{u}(\cdot) = \phi(u(\cdot))$ for some strictly increasing and strictly concave function $\phi(\cdot)$. Denote by $\widetilde{U}(x, t, \theta)$ the expected utility of a type-θ agent with von Neumann–Morgenstern utility function $\widetilde{u}(\cdot)$.

As before, fix $\theta' \in \Theta$ and let (χ, τ) be any point satisfying (a) and (b) from the proof of (i). We will show that $\widetilde{U}(x(\theta'), t(\theta'), \theta') \geq \widetilde{U}(\chi, \tau, \theta')$.[33] As in (i), we can suppose that $\chi - \tau < x(\theta') - t(\theta')$. Let $u_\ell(\theta')$, $u_n(\theta')$, u_n, and u_ℓ be defined as before, and set $\Delta\phi_i =$

[32] As the loss increases, the maximizers on C *strongly increase* in the sense of Shannon (1995, pp. 215–216). Claim 1 shows that U satisfies the strict single-crossing property in (x, t) and ℓ, so the conclusion in this sentence also follows from her Theorem 4.

[33] This inequality follows from familiar comparative statics results, for example, Theorem 1 in Jewitt (1987). We nonetheless include its simple proof.

$\phi(u_i(\theta')) - \phi(u_i)$ and $\Delta u_i = u_i(\theta') - u_i$, for $i = \ell, n$. Then

$$\begin{aligned}\widetilde{U}(x(\theta'), t(\theta'), \theta') - \widetilde{U}(\chi, \tau, \theta') &= \theta\Delta\phi_\ell + (1-\theta)\Delta\phi_n \\ &= \theta\frac{\Delta\phi_\ell}{\Delta u_\ell}\Delta u_\ell + (1-\theta)\frac{\Delta\phi_n}{\Delta u_n}\Delta u_n \\ &> \frac{\Delta\phi_n}{\Delta u_n}(\theta\Delta u_\ell + (1-\theta)\Delta u_n) \\ &\geq 0,\end{aligned}$$

where the first inequality follows from the strict concavity of $\phi(\cdot)$, and the second follows from the monotonicity of $\phi(\cdot)$ and $U(x(\theta'), t(\theta'), \theta) \geq U(\chi, \tau, \theta)$. Hence $(x(\theta), t(\theta))_{\theta\in\Theta}$ satisfies the downward incentive compatibility and participation constraints after the increase in risk aversion. As in the proof of (i), now let each type choose a best contract in C (the closure of the original menu in $\mathbb{R}^2$ and $(0,0)$). Any such menu satisfies (IC) and (P) after the agent becomes more risk averse and has at least as much profit as $(x(\theta), t(\theta))_{\theta\in\Theta}$. Hence, an increase in risk aversion cannot lower the principal's expected profit.

A.4 *Square-root utility example*

Principal's problem. Let us denote a menu (x_1, t_1, x_2, t_2). It is easy to show that in this case the downward incentive constraint is binding. Also, using Theorem 1, $x_2 = \ell$, $x_i \geq 0$ (so $t_i \geq 0$), $i = 1, 2$, and (P) binds for θ_1. The principal's problem becomes

$$\max_{x_1, t_1, t_2} f_1(t_1 - \theta_1 x_1) + f_2(t_2 - \theta_2 \ell)$$

subject to

$$\begin{aligned}\theta_1\sqrt{w-\ell+x_1-t_1} + (1-\theta_1)\sqrt{w-t_1} &= \theta_1\sqrt{w-\ell} + (1-\theta_1)\sqrt{w} \\ \sqrt{w-t_2} &= \theta_2\sqrt{w-\ell+x_1-t_1} + (1-\theta_2)\sqrt{w-t_1},\end{aligned}$$

and $0 \leq x_1 \leq \ell$ (or $t_1 \geq 0$ and $x_1 \leq \ell$), which we ignore for now (and check ex post).

Candidate solution. After some algebra, we obtain the expressions

$$t_1 = w - \kappa(f_1, f_2, \theta_1, \theta_2)^2\overline{U}(\theta_1)^2 \tag{27}$$

$$x_1 = \ell + \left(-\kappa(f_1, f_2, \theta_1, \theta_2)^2 + \left(\frac{1}{\theta_1} - \frac{(1-\theta_1)}{\theta_1}\kappa(f_1, f_2, \theta_1, \theta_2)\right)^2\right)\overline{U}(\theta_1)^2 \tag{28}$$

$$t_2 = w - \left(\frac{\theta_2}{\theta_1}(1-(1-\theta_1)\kappa(f_1, f_2, \theta_1, \theta_2)) + (1-\theta_2)\kappa(f_1, f_2, \theta_1, \theta_2)\right)^2\overline{U}(\theta_1)^2, \tag{29}$$

where $\overline{U}(\theta_1) = \theta_1\sqrt{w-\ell} + (1-\theta_1)\sqrt{w}$ and

$$\kappa(f_1, f_2, \theta_1, \theta_2) = \frac{f_1\theta_1(1-\theta_1) + f_2\theta_2(\theta_2-\theta_1)}{f_1\theta_1(1-\theta_1) + f_2(\theta_2-\theta_1)^2} \in \left(1, \frac{\theta_2}{\theta_2-\theta_1}\right). \tag{30}$$

Along with $x_2 = \ell$, (27)–(29) is an optimal menu if $t_1 \geq 0$ and $x_1 \leq \ell$. The bounds given in (30) imply $x_1 \leq \ell$. Thus, our candidate solution is an actual one if $t_1 \geq 0$.

Comparative statics with respect to wealth. It is easy to show that t_1 and x_1 are decreasing in w, while x_2 is independent of it since it equals ℓ. In turn, the change in t_2 is given by (we omit the arguments of κ to simplify the notation)

$$\frac{\partial t_2}{\partial w} = 1 - \left(\frac{\theta_2}{\theta_1}(1-(1-\theta_1)\kappa) + (1-\theta_2)\kappa\right)^2 \overline{U}(\theta_1)(\theta_1(w-\ell)^{-0.5} + (1-\theta_1)w^{-0.5}). \quad (31)$$

We now show that for an open set of parameters, this derivative is positive when loss size is sufficiently small. To see this, let $\ell = \alpha w$ for $\alpha \in (0,1)$. Then

$$\overline{U}(\theta_1)(\theta_1(w-\ell)^{-0.5} + (1-\theta_1)w^{-0.5}) = (\theta_1(1-\alpha)^{0.5} + (1-\theta_1))(\theta_1(1-\alpha)^{-0.5} + (1-\theta_1)),$$

which does not depend on w and tends to 1 as α vanishes. Since $(\theta_2/\theta_1)(1-(1-\theta_1)\kappa) + (1-\theta_2)\kappa < 1$ as $\kappa > 1$, the result follows.

Omitted constraint. We have ignored the constraint $t_1 \geq 0$. We now show that $t_1 \geq 0$ and $\partial t_2/\partial w > 0$ can both hold when α is sufficiently small. Rewrite $t_1 \geq 0$ as

$$t_1 = w(1 - \kappa^2(\theta_1\sqrt{1-\alpha} + 1 - \theta_1)^2) \geq 0. \quad (32)$$

Simple algebraic manipulation reveals that (32) holds if and only if $\kappa \leq \hat{\kappa}(\alpha)$, while (31) holds if and only if $\kappa > \widetilde{\kappa}(\alpha)$. The functions $\hat{\kappa}(\cdot)$ and $\widetilde{\kappa}(\cdot)$ are continuously differentiable, $\hat{\kappa}(0) = \widetilde{\kappa}(0) = 1$, $\hat{\kappa}'(\alpha) > 0$ for all $\alpha \in (0,1)$, and $\widetilde{\kappa}'(0) = 0$. Thus, $\hat{\kappa}(\alpha) > \widetilde{\kappa}(\alpha)$ for α sufficiently small and both inequalities are satisfied. Since for any θ_1 and θ_2 we can make κ arbitrarily close to 1 as f_1 tends to 1, it follows that there exists an open set of parameters $(f_1, f_2, \theta_1, \theta_2, \ell, w)$ such that $t_1 \geq 0$ and $\partial t_2/\partial w > 0$.

A.5 *Complete sorting lemma*

Lemma 5 (Complete sorting). *For $\theta \in [\underline{\theta}, \overline{\theta})$, $\Delta'(\theta) < 0$ if and only if*

$$\frac{f'(\theta)}{f(\theta)} > \frac{3\theta - 2 - b(\theta)}{\theta(1-\theta)}, \quad (33)$$

where $b(\theta) = h'(V(\theta) - (1-\theta)\Delta(\theta))/[h'(V(\theta) + \theta\Delta(\theta)) - h'(V(\theta) - (1-\theta)\Delta(\theta))]$. If (33) holds for all $\theta \in \Theta$, then the optimal menu sorts all types who obtain insurance.

Proof. Using (8) and (9) to eliminate $\lambda'(\theta)$, rewrite the numerator of (15) as $\lambda(\theta)B(\theta)$, where

$$B(\theta) = f(\theta)\left[\frac{f'(\theta)}{f(\theta)}\theta(1-\theta) + (1-3\theta) + \frac{1}{1 - \frac{h'(V(\theta)-(1-\theta)\Delta(\theta))}{h'(V(\theta)+\theta\Delta(\theta))}}\right].$$

Since

$$\frac{1}{1 - \frac{h'(V(\theta)-(1-\theta)\Delta(\theta))}{h'(V(\theta)+\theta\Delta(\theta))}} = 1 + \frac{1}{\frac{h'(V(\theta)+\theta\Delta(\theta))}{h'(V(\theta)-(1-\theta)\Delta(\theta))} - 1}$$

and $\lambda(\theta) < 0$ for $\theta < \overline{\theta}$, it follows that $\Delta'(\theta) < 0$ *if and only if*

$$\frac{f'(\theta)}{f(\theta)}\theta(1-\theta) + (2-3\theta) + \frac{h'(V(\theta)-(1-\theta)\Delta(\theta))}{h'(V(\theta)+\theta\Delta(\theta)) - h'(V(\theta)-(1-\theta)\Delta(\theta))} > 0$$

or, equivalently,

$$\frac{f'(\theta)}{f(\theta)} > \frac{3\theta - 2 - b(\theta)}{\theta(1-\theta)},$$

where $b(\theta) = h'(V(\theta)-(1-\theta)\Delta(\theta))/[h'(V(\theta)+\theta\Delta(\theta)) - h'(V(\theta)-(1-\theta)\Delta(\theta))]$. □

A.6 *Proof of Theorem 3 (Complete sorting: sufficiency)*

(i) Since $\rho'(\theta)/\rho(\theta) = \rho(\theta) + f'(\theta)/f(\theta)$, we must show that the condition

$$\frac{f'(\theta)}{f(\theta)} > \frac{3\theta - 1}{\theta(1-\theta)} - \frac{f(\theta)}{1-F(\theta)} \tag{34}$$

is sufficient for complete sorting. Fix $\hat{\theta} \in [\underline{\theta}, \overline{\theta})$. We first claim that if $\Delta'(\tau) < 0$ for all $\tau \in (\hat{\theta}, \overline{\theta})$ and condition (34) holds, then $\Delta'(\hat{\theta}) < 0$. To establish this claim, we show that

$$b(\hat{\theta}) > -1 + \frac{f(\hat{\theta})}{1-F(\hat{\theta})}\hat{\theta}(1-\hat{\theta}), \tag{35}$$

implying that condition (33) holds at $\hat{\theta}$.

Let $h_n'(\theta) = h'(V(\theta)+\theta\Delta(\theta))$. Since $\Delta'(\tau) < 0$ and $f(\tau) > 0$ for all $\tau \in (\hat{\theta}, \overline{\theta})$,

$$f(\tau)h_n'(\hat{\theta}) > f(\tau)h_n'(\tau) > f(\tau)a(\tau) \tag{36}$$

for all $\tau \in (\hat{\theta}, \overline{\theta})$, with equalities at $\overline{\theta}$, where $a(\cdot)$ is defined in (11).

Integrate both sides of (36) from $\hat{\theta}$ to $\overline{\theta}$ and divide by $1 - F(\hat{\theta})$ to obtain

$$h_n'(\hat{\theta}) > \frac{1}{1-F(\hat{\theta})}\int_{\hat{\theta}}^{\overline{\theta}} a(\tau)f(\tau)\,d\tau = (\Delta h)'(\hat{\theta})\frac{\hat{\theta}(1-\hat{\theta})f(\hat{\theta})}{1-F(\hat{\theta})}, \tag{37}$$

where $(\Delta h)'(\theta) = h'(V(\theta)+\theta\Delta(\theta)) - h'(V(\theta)-(1-\theta)\Delta(\theta))$ and we have used (11). Add $-(\Delta h)'(\hat{\theta})$ to both sides of (37) and rearrange to get (35), so that (33) holds at $\hat{\theta}$.

It now follows that, under condition (34), $\Delta'(\theta) < 0$ for all $\theta \in \Theta$: if $\Delta'(\cdot) \geq 0$ somewhere, then there would be a largest $\theta \in [\underline{\theta}, \overline{\theta})$ with $\Delta'(\theta) \geq 0$ (since $\Delta'(\cdot)$ is continuous and $\limsup_{\theta \to \overline{\theta}} \Delta'(\theta) < 0$). By the claim, condition (34) would fail.

(ii) Let $f'(\cdot)/f(\cdot)$ be decreasing and suppose that sorting is not complete. Then $\Delta'(\theta) \geq 0$ for some θ. We will show that $\overline{\theta} \geq 1/2$ and that $f'(\cdot)$ is sometimes negative.

Since $\limsup_{\theta \to \overline{\theta}} \Delta'(\theta) < 0$ and Δ' is continuous on $(\underline{\theta}, \overline{\theta})$, there is a largest type $\hat{\theta} \in \Theta$ with $\Delta'(\hat{\theta}) = 0$. In addition, $\Delta'(\theta) < 0$ for all $\theta \in (\hat{\theta}, \overline{\theta})$.

From Lemma 3 the sign of $-\Delta'(\cdot)$ is the same as the sign of

$$g(\theta) = \frac{f'(\theta)}{f(\theta)} - \frac{(3\theta - 2 - b(\theta))}{\theta(1-\theta)}. \tag{38}$$

Moreover,

$$b'(\theta) = -\Delta'(\theta)\left[\frac{(1-\theta)h''_\ell(h'_n - h'_\ell) + h'_\ell h''_\ell(1-\theta) + h''_n h'_\ell \theta}{(h'_n - h'_\ell)^2}\right],$$

so that $b'(\hat{\theta}) = 0$. Since $\Delta'(\theta) < 0$ for all $\theta > \hat{\theta}$ and since $g(\hat{\theta}) = 0$, we must have $g'(\hat{\theta}) \geq 0$. Since $f'(\cdot)/f(\cdot)$ is decreasing, the second fraction *cannot be increasing* at $\hat{\theta}$. But

$$\frac{\partial}{\partial\theta}\frac{3\theta - 2 - b(\theta)}{\theta(1-\theta)} = \frac{-\theta(1-\theta)b'(\theta) + 3\theta^2 + (1-2\theta)(2+b(\theta))}{\theta^2(1-\theta)^2}.$$

Since $b'(\hat{\theta}) = 0$ and $b(\theta) > 0$, $g'(\hat{\theta}) \geq 0$ implies that $\hat{\theta} > 1/2$, so $\overline{\theta} > 1/2$.

To show that $f'(\cdot)$ must sometimes be negative, rewrite (38) as $\widetilde{g}(\theta) = (1-\theta)g(\theta) = (1-\theta)(f'(\theta)/f(\theta)) - (3\theta - 2 - b(\theta))/\theta$. Since $\Delta'(\theta) < 0$ for all $\theta > \hat{\theta}$ and since $\widetilde{g}(\hat{\theta}) = 0$, we must have $\widetilde{g}'(\hat{\theta}) \geq 0$. But since $b'(\hat{\theta}) = 0$, the fraction $(3\theta - 2 - b(\theta))/\theta$ is strictly increasing in a neighborhood of $\hat{\theta}$, which implies that $(1-\theta)f'(\theta)/f(\theta)$ must be increasing in a neighborhood of $\hat{\theta}$, so $f'(\cdot)$ is negative on a neighborhood of $\hat{\theta}$.

(iii) Notice that $b(\theta) \geq b_0$ for every θ, where $b_0 = h'(u(w-\ell))/[h'(u(w)) - h'(u(w-\ell))] > 0$. By Lemma 5, complete sorting follows if $f'(\theta)/f(\theta) > (3\theta - 2 - b_0)/\theta(1-\theta)$. Since $\lim_{\ell\to 0} u(w-\ell) = u(w)$, $\lim_{\ell\to 0}\Delta_0 = 0$. Thus, $\lim_{\ell\to 0} b_0 = \infty$. As the ratio $f'(\cdot)/f(\cdot)$ is bounded below, there is a threshold for the loss, $\hat{\ell} > 0$, such that (33) is satisfied for all types if $\ell \in (0, \hat{\ell})$.

A.7 *Proof of Proposition 2 (No exclusion and exclusion)*

(i) Type $\widetilde{\theta}$ is excluded from the optimal menu of contracts if $\Delta(\widetilde{\theta}) = \Delta_0$. From (11), we must show that the marginal benefit of providing insurance to $\widetilde{\theta}$ starting from no insurance is less than the marginal cost of doing so. Formally,

$$f(\widetilde{\theta})\widetilde{\theta}(1-\widetilde{\theta})\big[h'(u(w)) - h'(u(w-\ell))\big] < \int_{\widetilde{\theta}}^{\overline{\theta}} a(s)f(s)\,ds. \tag{39}$$

Assume first that $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$. Then $\int_{\widetilde{\theta}}^{\overline{\theta}} a(s)f(s)\,ds > a(\overline{\theta})(1-F(\widetilde{\theta}))$ and (39) holds if $f(\widetilde{\theta})\widetilde{\theta}(1-\widetilde{\theta})[h'(u(w)) - h'(u(w-\ell))] < a(\overline{\theta})(1-F(\widetilde{\theta}))$. Since $a(\overline{\theta}) = h'(U(\overline{\theta})) \geq h'((1-\overline{\theta})u(w) + \overline{\theta}u(w-\ell))$, type $\widetilde{\theta}$ is excluded from the optimal menu if

$$\frac{f(\widetilde{\theta})}{(1-F(\widetilde{\theta}))} < \frac{h'((1-\overline{\theta})u(w) + \overline{\theta}u(w-\ell))}{\widetilde{\theta}(1-\widetilde{\theta})[h'(u(w)) - h'(u(w-\ell))]}. \tag{40}$$

Since sorting is complete, any $\theta \leq \widetilde{\theta}$ is excluded as well.

Without imposing $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$, a similar, but stronger, sufficient condition for exclusion holds by replacing the numerator of (40) with $h'(u(w-\ell))$. Then the right side of (40) tends to ∞ if risk aversion tends to 0 uniformly on $[w-\ell, w]$, so (40) holds if the agent's risk aversion is low enough on $[w-\ell, w]$.

(ii) Type $\underline{\theta}$ is not excluded from the optimal menu of contracts if $\Delta(\underline{\theta}) > \Delta_0$. From (11) and the concavity of the optimal control problem, this is tantamount to showing that the marginal benefit of providing insurance to $\underline{\theta}$ starting from no insurance is greater than the marginal cost of doing so. Formally,

$$f(\underline{\theta})\underline{\theta}(1-\underline{\theta})\big[h'(u(w))-h'(u(w-\ell))\big] > \int_{\underline{\theta}}^{\overline{\theta}} a(s)f(s)\,ds. \tag{41}$$

Assume first that $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$. It is easy to show that in this case $a'(\cdot) < 0$. Hence, $\int_{\underline{\theta}}^{\overline{\theta}} a(s)f(s)\,ds < a(\underline{\theta})$ and (41) holds if $f(\underline{\theta})\underline{\theta}(1-\underline{\theta})[h'(u(w))-h'(u(w-\ell))] > a(\underline{\theta})$. Since $a(\underline{\theta}) < (1-\underline{\theta})h'(u(w))+\underline{\theta}h'(u(w-\ell))$, it follows that

$$f(\underline{\theta})\underline{\theta}(1-\underline{\theta})\big[h'(u(w))-h'(u(w-\ell))\big] > (1-\underline{\theta})h'(u(w))+\underline{\theta}h'(u(w-\ell)),$$

which is equivalent to

$$f(\underline{\theta}) > \frac{(1-\underline{\theta})h'(u(w))+\underline{\theta}h'(u(w-\ell))}{\underline{\theta}(1-\underline{\theta})[h'(u(w))-h'(u(w-\ell))]}. \tag{42}$$

Thus, type $\underline{\theta}$ is not excluded from the optimal menu if $f(\underline{\theta})$ is greater than the right side of (42). Since sorting is complete, no type is excluded.

Without imposing $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$, a stronger sufficient condition holds, with the numerator on the right side of (42) replaced with $h'(u(w))$. Then the right side of (42) tends to 1 if risk aversion tends to ∞ uniformly on $[w-\ell, w]$. If $f(\underline{\theta}) > 1/\underline{\theta}(1-\underline{\theta})$, then (42) holds if the agent's risk aversion is sufficiently high on $[w-\ell, w]$.

A.8 *Proof of Lemma 3 (Curvature)*

Let $(V(\theta), \Delta(\theta))$ solve the optimal control problem with $\Delta'(\cdot) < 0$ everywhere. Since $u_n(\theta) = V(\theta) + \theta\Delta(\theta)$ for all θ, we can use (1) and (2) to recover the optimal menu $(x(\theta), t(\theta))_{\theta\in\Theta}$.

Recall that $V'(\theta) = -\Delta(\theta)$, so $u_n'(\theta) = \theta\Delta'(\theta)$. Differentiate (1) and (2) to get

$$t'(\theta) = -\frac{\theta\Delta'(\theta)}{u'(w-t(\theta))} \tag{43}$$

$$x'(\theta) = -\frac{\Delta'(\theta)[(1-\theta)u'(w-t(\theta))+\theta u'(w-\ell+x(\theta)-t(\theta))]}{u'(w-\ell+x(\theta)-t(\theta))u'(w-t(\theta))}, \tag{44}$$

where we have used $u_n(\theta) = u(w-t(\theta))$ and $h'(\cdot) = 1/u'(h(\cdot))$.

Since (by assumption) sorting is complete, we have $x'(\cdot) > 0$, so the inverse of $x(\cdot)$, call it $z(\cdot)$, is well defined (i.e., $\theta = z(x)$). We can now represent the optimal mechanism as a *nonlinear premium schedule* $T(x) = t(z(x))$. Then (43), (44), and $\theta = z(x)$ give

$$T'(x) = t'(z(x))z'(x) = \frac{t'(z(x))}{x'(z(x))} = \frac{\theta u'(w-\ell+x(\theta)-t(\theta))}{(1-\theta)u'(w-t(\theta))+\theta u'(w-\ell+x(\theta)-t(\theta))}. \tag{45}$$

Differentiate $T'(\cdot)$ and use $\theta = z(x)$ to find, after some algebra,

$$T''(x) = \frac{1}{(\theta u_\ell'+(1-\theta)u_n')^2}\left\{\frac{u_n'u_\ell'}{x'(\theta)} + \theta(1-\theta)\left[u_n''u_\ell'\frac{t'(\theta)}{x'(\theta)} + u_\ell''u_n'\left(1-\frac{t'(\theta)}{x'(\theta)}\right)\right]\right\}. \tag{46}$$

Insert (43) and (44) into (46), and manipulate the resulting expression to reveal that $T''(x) < 0$ if and only if $\Delta'(\theta) < (\theta(1-\theta)[\theta \frac{u_n''}{u_n'^2} + (1-\theta)\frac{u_\ell''}{u_\ell'^2}])^{-1}$. Use $h' = 1/u'$, $h'' = u''/u'^3$, (8), and (9) to rewrite (15) as

$$\Delta'(\theta) = \frac{(\frac{1}{u_n'} - \frac{1}{u_\ell'})\Omega + E[\frac{1}{u'}]}{\theta(1-\theta)\big[\theta \frac{u_n''}{u_n'^3} + (1-\theta)\frac{u_\ell''}{u_\ell'^3}\big]},$$

where $\Omega = \theta(1-\theta)\frac{f'}{f} + 1 - 2\theta$ and $E[\frac{1}{u'}] = \theta(1/u_\ell') + (1-\theta)(1/u_n')$. Now find that $T''(x) < 0$ if and only if

$$\theta(1-\theta)\frac{f'}{f} + 1 - 2\theta > \frac{\big[\theta \frac{u_n''}{u_n'^3} + (1-\theta)\frac{u_\ell''}{u_\ell'^3}\big] - E\big[\frac{1}{u'}\big]\big[\theta \frac{u_n''}{u_n'^2} + (1-\theta)\frac{u_\ell''}{u_\ell'^2}\big]}{\big[\theta \frac{u_n''}{u_n'^2} + (1-\theta)\frac{u_\ell''}{u_\ell'^2}\big]\big(\frac{1}{u_n'} - \frac{1}{u_\ell'}\big)}. \tag{47}$$

The numerator on the right side of (47) simplifies to $(\frac{1}{u_n'} - \frac{1}{u_\ell'})[\theta^2 \frac{u_n''}{u_n'^2} - (1-\theta)^2 \frac{u_\ell''}{u_\ell'^2}]$, and the entire right side simplifies to $\theta - 1 + c(\theta)$. Rearrange to get the result.

References

Arnott, Richard and Joseph E. Stiglitz (1988), "Randomization with asymmetric information." *Rand Journal of Economics*, 19, 344–362. [581]

Barlow, Richard E., Albert W. Marshall, and Frank Proschan (1963), "Properties of probability distributions with monotone hazard rate." *Annals of Mathematical Statistics*, 34, 375–389. [583]

Biais, Bruno, David Martimort, and Jean-Charles Rochet (2000), "Competing mechanisms in a common value environment." *Econometrica*, 68, 799–837. [574]

Cawley, John and Tomas Philipson (1999), "An empirical examination of information barriers to trade in insurance." *American Economic Review*, 89, 827–846. [573, 591]

Chade, Hector and Virginia N. Vera de Serio (2011), "Wealth effects and agency costs." Unpublished paper. [574]

Chiappori, Pierre-André and Alberto Bernardo (2003), "Bertrand and Walras equilibria under moral hazard." *Journal of Political Economy*, 111, 785–817. [571]

Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié, and François Salanié (2006), "Asymmetric information in insurance: General testable implications." *Rand Journal of Economics*, 37, 783–798. [573, 575, 591, 592]

Chiappori, Pierre-André and Bernard Salanié (2000), "Testing for asymmetric information in insurance market." *Journal of Political Economy*, 108, 56–78. [591]

Clarke, Frank H. (1976), "The maximum principle with minimal hypotheses." *SIAM Journal of Control and Optimization*, 14, 1078–1091. [583]

Cohen, Alma (2005), "Asymmetric information and learning: Evidence from the automobile insurance market." *Review of Economics and Statistics*, 87, 197–207. [591]

Cohen, Alma and Liran Einav (2007), "Estimating risk preferences from deductible choice." *American Economic Review*, 97, 745–788. [592]

Dafny, Leemore S. (2010), "Are health insurance markets competitive?" *American Economic Review*, 100, 1399–1431. [592]

Dionne, Georges, Christian Gouriéroux, and Charles Vanasse (2001), "Testing for evidence of adverse selection in the automobile insurance market: A comment." *Journal of Political Economy*, 109, 444–453. [591]

Finkelstein, Amy and James Poterba (2004), "Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market." *Journal of Political Economy*, 112, 183–208. [591]

Fudenberg, Drew and Jean Tirole (1990), "Moral hazard and renegotiation in agency contracts." *Econometrica*, 58, 1279–1319. [573]

Guesnerie, Roger and Jean-Jacques Laffont (1984), "A complete solution to a class of principal–agent problems with an application to the control of a self-managed firm." *Journal of Public Economics*, 25, 329–369. [573, 574, 582]

He, Daifeng (2009), "The life insurance market: Asymmetric information revisited." *Journal of Public Economics*, 93, 1090–1097. [591]

Hellwig, Martin (2010), "Incentive problems with unidimensional hidden characteristics: A unified approach." *Econometrica*, 78, 1201–1237. [573, 574, 575, 577]

Jewitt, Ian (1987), "Risk aversion and the choice between risky prospects: The preservation of comparative statics." *Review of Economic Studies*, 54, 73–85. [599]

Jullien, Bruno (2000), "Participation constraints in adverse selection models." *Journal of Economic Theory*, 93, 1–47. [573, 574]

Machina, Mark J. (1995), "Non-expected utility and the robustness of the classical insurance paradigm." *Geneva Papers on Risk and Insurance Theory*, 20, 9–50. [577, 579]

Maskin, Eric S. and John Riley (n.d.), "Monopoly with incomplete information." Unpublished paper. [581]

Maskin, Eric S. and John Riley (1984), "Monopoly with incomplete information." *Rand Journal of Economics*, 15, 171–196. [572, 573, 574, 575, 590, 598]

Matthews, Steven and John Moore (1987), "Monopoly provision of quality and warranties: An exploration in the theory of multidimensional screening." *Econometrica*, 55, 441–467. [573, 574]

Milgrom, Paul and Ilya Segal (2002), "Envelope theorems for arbitrary choice sets." *Econometrica*, 70, 583–601. [583]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [573, 574]

Nöldeke, Georg and Larry Samuelson (2007), "Optimal bunching without optimal control." *Journal of Economic Theory*, 134, 405–420. [573, 574]

Ormiston, Michael B. and Edward E. Schlee (2001), "Mean–variance preferences and investor behaviour." *Economic Journal*, 111, 849–861. [577]

Page, Frank H. (1992), "Mechanism design for general screening problems with moral hazard." *Economic Theory*, 2, 265–281. [573, 577]

Prescott, Edward C. and Robert M. Townsend (1984), "Pareto optima and competitive equilibria with adverse selection and moral hazard." *Econometrica*, 52, 21–46. [571]

Puelz, Robert and Arthur Snow (1994), "Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market." *Journal of Political Economy*, 102, 236–257. [591]

Rothschild, Michael and Joseph E. Stiglitz (1976), "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information." *Quarterly Journal of Economics*, 90, 629–649. [571, 572, 573, 591]

Schlesinger, Harris (1983), "Nonlinear pricing strategies for competitive and monopolistic insurance markets." *Journal of Risk and Insurance*, 50, 61–83. [573]

Shannon, Chris (1995), "Weak and strong monotone comparative statics." *Economic Theory*, 5, 209–227. [599]

Spence, A. Michael (1977), "Nonlinear prices and welfare." *Journal of Public Economics*, 8, 1–18. [573, 589]

Stiglitz, Joseph E. (1977), "Monopoly, non-linear pricing and imperfect information: The insurance market." *Review of Economic Studies*, 44, 407–430. [571, 572, 573, 577, 580, 585, 587, 592]

Strausz, Roland (2006), "Deterministic versus stochastic mechanisms in principal-agent models." *Journal of Economic Theory*, 128, 306–314. [581]

Szalay, Dezsö (2008), "Monopoly, non-linear pricing, and imperfect information: A reconsideration of the insurance market." Warwick Economic Research Papers 863, University of Warwick. [573]

Thiele, Henrik and Achim Wambach (1999), "Wealth effects in the principal agent model." *Journal of Economic Theory*, 89, 247–260. [574, 579]

Vinter, Richard (2000), *Optimal Control*. Birkhäuser, Boston. [583]

Wilson, Robert B. (1993), *Nonlinear Pricing*. Oxford University Press, Oxford. [591]

Submitted 2009-11-2. Final version accepted 2011-8-15. Available online 2011-8-16.