*Optimal Insurance with Adverse Selection**

Hector Chade† Edward Schlee‡

Arizona State University, Department of Economics

Arizona State University, Department of Economics

This version: May 2008

Abstract

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

We derive several properties of optimal menus for an *arbitrary* type distribution: the highest type gets full coverage (efficiency at the top), all other types get less than full coverage (downward distortions elsewhere), the premium and indemnity are nonnegative and the principal makes positive expected profit. More importantly, we prove a novel comparative static result for *wealth effects*, showing that the principal always prefers an agent facing a larger loss, and a poorer one if the agent's risk aversion decreases with wealth.

We then specialize to the case with a continuum of types distributed according to an smooth density. We give sufficient conditions for complete sorting, exclusion, and quantity discounts. Our most surprising result is that, under two mild assumptions–the monotone likelihood ratio property for the density and decreasing absolute risk aversion for the agent–the optimal premium is *backwards-S shaped* in the amount of coverage, first concave, then convex.

We contrast our results with the standard monopoly model with private values and quasilinear preferences and with competitive insurance models. We calculate a closed form solution for the CARA case and use it to illustrate these differences.

Although we focus on the monopoly insurance problem, our proofs can be adapted to other screening problems with wealth effects and common values.

Keywords: Principal-Agent Model, Monopoly Insurance, Adverse Selection, Common Values, Wealth Effects, Quantity Discounts and Premia.

*We thank Lones Smith, Stephan Lauermann and audience members at the University of Montreal, CEA-Universidad de Chile, University of Miami, University of Southern California, Georgia, UC Santa Barbara, Ohio State, the 2007 Midwest Economic Theory Meetings at Michigan, the 2008 Winter Meetings of the Econometric Society in New Orleans, the 2008 NSF/NBER Decentralization Conference at Tulane and the 2008 SWET conference at UCSB for comments.

†Email: hector.chade@asu.edu

‡Email: edward.schlee@asu.edu

1 Introduction

Moral hazard and adverse selection are fundamental problems in insurance. A large literature has explored how each affects insurance contracts in competitive or monopoly markets. The usual approach for moral hazard has been the principal-agent model (which includes the case of monopoly).[1] The usual approach for adverse selection has been competitive models, either the Rothschild and Stiglitz (1976) model or one of its variants. An exception is 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 continuum-type case with a continuous density.

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. The monopoly insurance problem does not fit the standard principal-agent model of a monopolist selling to a privately informed consumer (e.g., Maskin and Riley (1984)) for three reasons: 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*. To the best of our knowledge, we are the first to provide a complete solution to a principal-agent problem with *all* these properties.

We divide the paper into two parts. In the first we allow 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: 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).

As just mentioned, one crucial difference with the standard monopoly model is that the agent's wealth matters. An important question is how the agent's initial wealth and

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

the loss size affect the principal's profit. Using monotone methods, we prove a novel 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.

In the second part of the paper we specialize to the case of a continuum of types distributed according to a smooth density. We derive conditions for complete sorting of types, exclusion (or inclusion) of types, and quantity discounts. 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 optimal premium as a function of the coverage amount: under two mild assumptions–the density satisfies the monotone likelihood ratio property and the agent's risk aversion decreases with wealth–the premium is *backwards S-shaped*, first concave, then convex. The curvature property sharply distinguishes a monopoly insurance policy from one offered by a 'standard' (i.e., Maskin-Riley) monopolist and from competitive insurance policies. In particular, monopoly insurers do not offer global quantity premia, an implication of many competitive insurance models (e.g., Rothschild and Stiglitz (1976)).

We calculate a closed-form solution in the CARA case, and use it to illustrate some properties peculiar to the insurance problem. We show that there is *partial* insurance at the top if the highest type suffers a loss for sure; that monotonicity of the hazard rate does *not* suffice for complete sorting; and that common values *precludes* a globally concave premium schedule (unless the type space is severely restricted).

Our emphasis is almost exclusively on the monopoly insurance problem. We point out, however, that several of the proofs can be adapted to other principal-agent problems with adverse selection, wealth effects, and common values. We hope that our arguments will be helpful in more general principal-agent problems with these features.

Related Literature. The paper is closely related to three literatures. First, it is related to the insurance with adverse selection literature started by Rothschild and Stiglitz (1976) for competition, and Stiglitz (1977) for monopoly; each focuses on the two-type case.[2] We completely solve a more general problem than does Stiglitz (1977), and we

[2]The renegotiation stage in Fudenberg and Tirole (1990) can be interpreted as a monopoly insurance problem, in which the (random) effort chosen in the first stage is the agent's type. With a continuum of efforts, 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.

compare the predictions of monopoly and competition. A large literature tests implications of the *joint* hypothesis of adverse selection and (some version of) competition (e.g., Chiappori, Jullien, Salanié, and Salanié (2006) and Cawley and Philipson (1999)); our results help separate implications of adverse selection from competition.[3] Second, it is related to the literature on principal-agent models with adverse selection, illustrated by Spence (1977), Mussa and Rosen (1978), Maskin and Riley (1984), Matthews and Moore (1987), Jullien (2000), Hellwig (2006), and Nöldeke and Samuelson (2007). The complications of the insurance problem–wealth effects, common values, and type-dependent reservation utilities–are absent in Mussa and Rosen (1978), Maskin and Riley (1984), and Matthews and Moore (1987).[4] Jullien (2000) emphasizes 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 (2006) derives the standard 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), who determine how an agent's wealth affects the principal's profit in the moral hazard case. In our adverse selection problem we are able to do so with weaker assumptions.

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 of $w > 0$, faces a potential loss of $\ell \in (0, w)$ 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 $\mathbb{R}_+$.

[3] 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 describes their data better than a competitive one does.

[4] Biais, Martimort, and Rochet (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 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.

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 $0 < \underline{\theta} < \overline{\theta} < 1$. This formulation includes the discrete case with a finite number of types, the continuum of types case with an atomless density function, and mixtures of both.

The principal chooses, for each $\theta \in \Theta$, a contract $(x,t) \in \mathbb{R}$ 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 type-θ agent for a contract (x,t) is θ is $U(x,t,\theta) = \theta u(w - \ell + x - t) + (1-\theta)u(w-t)$. The function U satisfies the following *single crossing property*: 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, this is equivalent to $-U_x(x,t,\theta)/U_t(x,t,\theta)$ strictly increasing in θ; i.e., indifference curves cross once, with higher types being willing to pay more for a marginal increase in insurance.

By the Revelation Principle, we restrict attention to (measurable) menus $(x(\theta),t(\theta))_{\theta\in\Theta}$ for which the agent chooses to participate and to announce its true type θ. Formally, the principal solves the following problem:

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

subject to

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

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

It is instructive to compare this problem with the standard monopoly problem (e.g., Maskin and Riley (1984)). In the standard problem, the agent has quasilinear utility, her type does not directly affect the principal's profit, and it does not enter her reservation

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

utility. Here, the agent is risk averse, implying nontrivial wealth effects (except for the case of CARA preferences); the agent's reservation utility depends on the type; and the agent's type affects the principal's profit directly, i.e., it is a common values model.

The common values aspect leads to an important distinction. In the standard monopoly model, profit is increasing in type both under complete (first-best) and incomplete information. In our model, however, 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 that type's risk premium, which is concave in the agent's type, first increasing, then decreasing. Moreover, under 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.

3 The General Case: Arbitrary Type Distribution

We begin with some general properties of optimal menus, and determine how the agent's wealth affects the principal's profit. We emphasize that these results hold for an *arbitrary* type set $\Theta \subset (0,1)$ and a *general* cumulative distribution function on it.

3.1 A Useful Lemma

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 given 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 $t'' - t' > \theta(x'' - x')$; for 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.

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* $(x(\theta), t(\theta))_{\theta\in\Theta}$ *to the principal's problem satisfies the following properties:*

(i) (Monotonicity) $x(\theta)$, $t(\theta)$, *and* $x(\theta) - t(\theta)$ *are increasing in* θ*;*

(ii) (No Overinsurance) $x(\theta) \leq \ell$ *for almost all* θ*;*

(iii) (Nonnegativity) $x(\theta)$, $t(\theta)$, *and* $x(\theta) - t(\theta)$ *are nonnegative for almost all* θ*;*

(iv) (Participation) (P) *is binding for the lowest type* $\underline{\theta}$ *;*

(v) (Efficiency at the Top) *Without loss of generality,* $x(\overline{\theta}) = \ell$*;*

(vi) (No Pooling at the Top) *If* u *is differentiable, then* $x(\theta) < \ell$ *for almost all* $\theta < \overline{\theta}$*;*

(vii) (Profitability) *The principal's expected profit is positive.*

Proof. Appendix. ■

The proof has many steps, but except for a few measure-theoretic details, each is elementary. Result (i) follows immediately from incentive compatibility and the single-crossing property. We prove results (ii) and (iii) by contraposition: we use Lemma 1 to show that if a menu does not satisfy either property, then there is another feasible menu that increases profit for a positive mass of agents. We prove (v) and (vi) similarly. For (iv), we show that if this property fails, the principal can feasibly reduce the utility of each type in each state by the same amount and hence increase expected profit. Finally, we prove (vii) 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 derives properties (i) and (iv)-(vi) for the two-type case. In independent work, Hellwig (2006) proves properties (v) and (vi) for an optimal tax problem with adverse selection with a general type distribution. His proofs use first-order conditions from an optimal control problem (where much of the contribution is to extend the first order conditions to allow for discontinuous densities). Theorem 1 also shows that an optimal menu is nonnegative (property (*iii*)), and that there are always gains to trade (property (*vii*)). We prove these properties for a *general* type distribution using fairly *elementary* arguments, repeatedly invoking Lemma 1 to find profitable deviations from a menu that fails one of these properties. We argue in Section 5 that the arguments can be adapted to the standard monopoly pricing model with quasilinear utility (with or without common values), and doubtless similar arguments can be used in a more general principal-agent problem.

3.3 The Principal Prefers a Poorer Agent

The agent's risk aversion introduces wealth effects absent in the standard screening model with quasilinear utility: 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; one facing a larger or smaller potential loss? In the first-best case (observable types), the answer is immediate: the demand for insurance is higher if the loss amount is higher, or the agent is poorer and risk aversion decreases with wealth (DARA); in either case profit increases. The first-best argument fails with adverse selection, since the incentive compatibility and participation constraints change with wealth. And unfortunately the constraint sets cannot be ordered by inclusion as wealth changes. Despite this complication, the principal still prefers a poorer agent (under DARA) and a larger loss size.[6]

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

(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 each change–an increase in ℓ, risk aversion, or w (under DARA)–the menu continues to satisfy the *downward* incentive and participation constraints (though some upward incentive constraints may

[6]Part (*iii*) says that the principal prefers a poorer agent for a *given* loss amount. Clearly, the principal might prefer a richer agent if the potential loss rises with wealth.

fail). Now if we simply let each type choose its best contract from the original menu, then the principal's profit does not fall. Since the resulting menu satisfies (IC) and (P), the principal's maximum expected profit cannot fall.

Proof: Since (*iii*) follows from (*ii*), we just prove (*i*) and (*ii*). Fix a menu $(x(\theta), t(\theta))_{\theta\in\Theta}$ satisfying (IC) and (P), with $0 \leq x(\theta) \leq \ell$ for all $\theta \in \Theta$. (Any optimal menu satisfies these conditions almost everywhere.) Recall that by Lemma 1, $t(\theta) - \theta' x(\theta) \geq t(\theta') - \theta' x(\theta')$ 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 $\ell < \tilde{\ell} < w$, let $\tilde{U}(x, t, \theta)$ be the expected utility of a type-θ agent for contract (x, t) when the loss is $\tilde{\ell}$ (and $U(x, t, \theta)$ 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:

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

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

These inequalities of course 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 $(0, 0)$.

We now show that $\tilde{U}(x(\theta'), t(\theta'), \theta') \geq \tilde{U}(\chi, \tau, \theta')$. 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 $u_\ell(\theta') = u(w - \ell + x(\theta') - t(\theta'))$, $u_n(\theta') = u(w - t(\theta'))$, $u_n = u(w - \tau)$, $u_\ell = u(w - \ell + \chi - \tau)$, $\tilde{u}_\ell(\theta') = u(w - \tilde{\ell} + x(\theta') - t(\theta'))$, and $\tilde{u}_\ell = (w - \tilde{\ell} + \chi - \tau)$. Rewrite the inequality $U(x(\theta'), t(\theta'), \theta') \geq U(\chi, \tau, \theta')$ as

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

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

$$\tilde{U}(x(\theta'), t(\theta'), \theta') > \tilde{U}(\chi, \tau, \theta'). \tag{2}$$

Since, for each $\theta' \in \Theta$, (2) holds for any $(\chi, \tau) \neq (x(\theta'), t(\theta'))$ satisfying (a) and (b), it follows that $(x(\theta), t(\theta))_{\theta \in \Theta}$ continues to satisfy all the *downward* incentive and participation constraints when the loss equals $\tilde{\ell}$.[7]

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 $\tilde{U}(\cdot, \theta)$. (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 $\tilde{U}(\cdot, \cdot, \theta)$ 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 (2), $(x, t) \geq (x(\theta), t(\theta))$.[8] Hence, by Lemma 1, profit from the new contract does not fall. (By Theorem 1 coverage levels in C are wlog bounded above by ℓ.) Consider a menu defined by choosing for each $\theta \in \Theta$ any maximizer of $\tilde{U}(\cdot, \cdot, \theta)$. It satisfies (IC) and (P) when the loss is $\tilde{\ell}$ and is at least as profitable as the original menu.

(*ii*) The Principal Prefers a More Risk Averse Agent. Let $v(\cdot)$ be more risk averse than $u(\cdot)$: i.e., $v(\cdot) = T(u(\cdot))$, for some strictly increasing and strictly concave function $T(\cdot)$. Denote by $V(x, t, \theta)$ the expected utility of a type-θ agent with von Neumann-Morgenstern utility function $v(\cdot)$.

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 $V(x(\theta'), t(\theta'), \theta') \geq V(\chi, \tau, \theta')$.[9] 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 $\Delta T_i = T(u_i(\theta')) - T(u_i)$ and $\Delta u_i = u_i(\theta') - u_i$, for $i = \ell, n$. Then

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

[7] If there were only a finite number of types, the proof would be complete at this point, since we can replace (IC) by the downward incentive compatibility constraints and monotonicity in that case.

[8] As the loss increases, the maximizers on C *strongly increase* in the sense of Shannon (1995), pp. 215-16. Equation (2) 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.

[9] This inequality follows from Theorem 1 in Jewitt (1987). We nonetheless include its simple proof.

where the first inequality follows from the strict concavity of $T(\cdot)$, and the second from the monotonicity of $T(\cdot)$ 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. ■

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 agent's wealth makes the agent less lazy (loosens the incentive constraints), but stronger conditions are needed to conclude that the principal prefers a poorer agent. Under 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.

If profit and wealth or loss size were observable, then Theorem 2 would be *testable*: profit is higher on menus offered to agents who are poorer or exposed to higher losses.

3.4 A Reformulation

We have restricted contracts to be deterministic: each type is offered a single premium-indemnity pair. The simplest way to justify this restriction is to reformulate the problem by a change of variables: instead of choosing a menu of contracts, the principal chooses a menu of state-contingent utilities. We use this formulation for the rest of the paper.

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

$$u(\theta) = u(w - t(\theta)) \quad (3)$$

$$\Delta(\theta) = u(w - t(\theta)) - u(w - \ell + x(\theta) - t(\theta)). \quad (4)$$

A menu $(x(\theta), t(\theta))_{\theta \in \Theta}$ uniquely defines a menu $(u(\theta), \Delta(\theta))_{\theta \in \Theta}$. Conversely, given

$(u(\theta), \Delta(\theta))_{\theta\in\Theta}$, we can recover $(x(\theta), t(\theta))_{\theta\in\Theta}$ by

$$t(\theta) = w - h(u(\theta)) \tag{5}$$
$$x(\theta) = l - (h(u(\theta)) - h(u(\theta) - \Delta(\theta))), \tag{6}$$

where $h = u^{-1}$. Theorem 1 (i), (ii), and (iii) imply that, for any optimal menu, $u(\cdot)$ and $\Delta(\cdot)$ are decreasing in θ, and that, for almost all $\theta \in \Theta$, $0 \leq \Delta(\theta) \leq \Delta_0 = u(w) - u(w - \ell)$.

Thus an equivalent formulation of the principal's optimal contracting problem is

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

subject to

$$\begin{aligned} u(\theta) - \theta\Delta(\theta) &\geq u(\theta') - \theta\Delta(\theta') & \forall\theta, \theta' \in \Theta \\ u(\theta) - \theta\Delta(\theta) &\geq U(0, 0, \theta) & \forall\theta \in \Theta \\ \Delta(\theta) &\leq \Delta_0 & \\ \Delta(\theta) &\geq 0 & . \end{aligned}$$

In other words, we can think of a menu of contracts as specifying, for each type θ, a utility $u(\theta)$ in the no loss state, and a decrease in utility $\Delta(\theta)$ in case of a loss. In this formulation, the constraints are *linear* in the screening variables, and the objective function is strictly concave in them. This formulation makes it clear that stochastic menus cannot improve upon deterministic ones.[10]

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 $(\tilde{u}(\theta), \tilde{\Delta}(\theta))$. Since the constraints are linear in these variables, any type-θ

[10] Arnott and Stiglitz (1988), Proposition 10, proved a similar result for the two-type case. Our reformulation allows us to extend this result to any number of types arbitrarily distributed.

agent's constraints are satisfied if $\tilde{u}(\theta)$ and $\tilde{\Delta}(\theta)$ are replaced by their expected values. But profit increases from this change since the objective function is strictly concave. ■

4 Continuum of Types: The Density Case

To obtain further properties, from now on we specialize to the case in which $u(\cdot)$ 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 $(\underline{\theta}, \overline{\theta})$. Except for the possibility of a zero density at the endpoints, this distribution assumption is the most common one used in contracting problems with adverse selection. Yet, except for Stiglitz (1977), this case has been neglected by the literature on insurance with adverse selection. Under this assumption we can provide strong results for complete sorting of types, exclusion, and for the existence of 'quantity discounts.'

Abusing notation, now let $U(\theta)$ be the expected utility of type $\theta \in \Theta$ for a menu satisfying (IC) and (P). We prove in the Appendix that if $f(\cdot)$ is continuous and positive in the interior of Θ, then $U(\cdot)$ is continuously differentiable. (This fact follows from showing that any type's best choice from an optimal menu is unique.) By the Envelope Theorem (Milgrom and Segal (2002), Theorem 3), we have $U'(\theta) = -\Delta(\theta)$. We can use these facts to formulate the principal's problem as an optimal control problem.

4.1 The Optimal Control Problem

Let $\dot{U}(\theta)$ denote the derivative of U. By standard arguments, $U(\cdot)$ is convex, and (IC) holds if and only if $\dot{U}(\theta) = -\Delta(\theta)$ almost everywhere and $\Delta(\cdot)$ is nonincreasing, so there is no loss of generality in replacing (IC) by these two conditions. Moreover, by Theorem 1 *(iv)*, $U(\underline{\theta}) = U(0, 0, \underline{\theta})$. Since $U(\cdot)$ is C^1, we can write the principal's problem as an optimal control problem with a continuous control variable $\Delta(\cdot)$, a C^1 state variable $U(\cdot)$, and a free endpoint at $U(\overline{\theta})$:

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

subject to

$$\begin{aligned}
&\Delta(\cdot) \quad \text{nonincreasing} && (7)\\
\Delta(\theta) &\geq 0 \qquad \forall\theta && (8)\\
\Delta(\theta) &\leq \Delta_0 \qquad \forall\theta && (9)\\
\dot{U}(\theta) &= -\Delta(\theta) \qquad \forall\theta && (10)\\
U(\underline{\theta}) &= U(0,0,\underline{\theta}) && (11)\\
&U(\overline{\theta}) \quad \text{free.} && (12)
\end{aligned}$$

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 optimize pointwise with respect to the remaining variable, a great simplification. Since our objective is not linear in the indirect utility U, we are forced to proceed with optimal control arguments.

Consider the 'relaxed problem' that ignores (7)-(9), 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

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

and any solution to the relaxed problem satisfies

$$\begin{aligned}
-\lambda(\theta) &= f(\theta)\theta(1-\theta)[h'(U(\theta)+\theta\Delta(\theta)) - h'(U(\theta)-(1-\theta)\Delta(\theta))] && (13)\\
\dot{\lambda}(\theta) &= f(\theta)[(1-\theta)h'(U(\theta)+\theta\Delta(\theta)) + \theta h'(U(\theta)-(1-\theta)\Delta(\theta))] && (14)\\
\lambda(\overline{\theta}) &= 0, && (15)
\end{aligned}$$

as well as (10) and (11).[11]

Note that $\lambda(\theta) \leq 0$, and $\lambda(\theta) < 0$ if $\Delta(\theta) > 0$; also $\dot{\lambda}(\theta) > 0$ for all θ. Integrate (14)

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

with respect to θ, use (15), and replace the resulting expression in (13) to find

$$f(\theta)\theta(1-\theta)[h'(U(\theta)+\theta\Delta(\theta))-h'(U(\theta)-(1-\theta)\Delta(\theta))]=\int_{\theta}^{\overline{\theta}} a(s)f(s)ds, \tag{16}$$

where $a(s)=(1-s)h'(U(s)+s\Delta(s))+sh'(U(s)-(1-s)\Delta(s))>0$. It follows from (16) that the omitted constraint (8) is satisfied for *all* types. Note that, in line with Theorem 1 (*ii*) and (*iv*), we have $\Delta(\overline{\theta})=0$ and $\Delta(\theta)>0$ for all $\theta<\overline{\theta}$: type $\overline{\theta}$ gets full coverage and, to ensure incentive compatibility, all other types get partial coverage.[12]

Equation (16) illustrates the standard *efficiency vs. 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 increases also the utility of all higher types by one unit, and thus the cost to the principal is given by $\int_{\theta}^{\overline{\theta}} a(s)f(s)ds$.

4.2 Complete Sorting

By the Implicit Function Theorem and (16), $\Delta(\cdot)$ is C^1 on Θ (except possibly at the endpoints when f is zero there). Hence we can replace (7) with $\dot{\Delta}(\theta)\leq 0$ for all $\theta\in\Theta$. We next determine when the solution to the relaxed problem satisfies this constraint.

Differentiate (13) with respect to θ and use (10) to obtain, after some algebra,

$$\dot{\Delta}(\theta)=\frac{\lambda(\theta)[f'(\theta)\theta(1-\theta)+f(\theta)(1-2\theta)]-f(\theta)\theta(1-\theta)\dot{\lambda}(\theta)}{f(\theta)^2\theta^2(1-\theta)^2[\theta h''(U(\theta)+\theta\Delta(\theta))+(1-\theta)h''(U(\theta)-(1-\theta)\Delta(\theta))]}. \tag{17}$$

Since $h''(\cdot)>0$, the denominator of (17) is positive, and the sign of $\dot{\Delta}(\theta)$ depends on the sign of the numerator. We now derive conditions for *complete sorting* of types at

[12] This result is clear 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 (13) by $f(\theta_n)$ and use the Mean Value Theorem to write the right side as $\psi(\theta)(1-F(\theta_n))/f(\theta_n)$. The conclusion now follows since $\lim_{\theta\to\overline{\theta}} f(\theta)/(1-F(\theta))=\infty$ (Barlow, Marshall, and Proschan (1963), pp. 377-378). In Section 6 we show by example that this conclusion fails if $\overline{\theta}=1$.

the optimal contract; i.e., for $\dot{\Delta}(\cdot) < 0$ everywhere. As already proved in Theorem 1, equation (17) implies that there is *no pooling of types at the top*; i.e., $\dot{\Delta}(\overline{\theta}) < 0$.[13]

Lemma 2 (Complete Sorting) *The optimal menu sorts all types who obtain some insurance completely if and only if $f(\cdot)$ satisfies, for every θ,*

$$\frac{f'(\theta)}{f(\theta)} \geq \frac{3\theta - 2 - b(\theta)}{\theta(1-\theta)}, \tag{18}$$

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

Proof. Appendix. ■

An obvious problem with condition (18) is that $b(\cdot)$ is endogenous. But it immediately gives $f'(\theta)/f(\theta) > (3\theta - 2)/\theta(1-\theta)$ as a sufficient condition for complete sorting, a fact pointed out by Stiglitz (1977). To improve upon this condition, note that $b(\theta) \geq b^l$ for every θ, where $b^l = h'(u(w-\ell) - (\overline{\theta} - \underline{\theta})\Delta_0)/[h'(u(w) + (\overline{\theta} - \underline{\theta})\Delta_0) - h'(u(w-\ell) - (\overline{\theta} - \underline{\theta})\Delta_0)] > 0$. By Lemma 2, complete sorting follows if $f'(\theta)/f(\theta) > (3\theta - 2 - b^\ell)/\theta(1-\theta)$. Although it depends only on primitives, this condition is hard to verify and does not even imply what we know from Theorem 1, that there is no pooling at the top.

The next result addresses these problems. Let $\rho(\theta) = \frac{f(\theta)}{1-F(\theta)}$ denote the *hazard rate* of the distribution. We say that $f(\cdot)$ satisfies the monotone hazard rate condition (MHRC) if $\rho(\cdot)$ is increasing in θ; it satisfies the monotone likelihood ratio property (MLRP) if $f'(\cdot)/f(\cdot)$ is decreasing in θ. As is well-known and easy to check, MLRP implies MHRC.

Theorem 3 (Complete Sorting: Sufficient Conditions) *The optimal menu completely sorts all types who get some insurance if*

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

(ii) $f(\cdot)$ *satisfies MLRP and either* $\overline{\theta} \leq 1/2$ *or* $f'(\cdot) \geq 0$*; or*

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

Proof. The proofs of parts *(i)* and *(ii)* are in the Appendix. To prove part *(iii)*, note that $\lim_{\ell \to 0} u(w-\ell) = u(w)$ and thus $\lim_{\ell \to 0} \Delta_0 = 0$. Consequently, $\lim_{\ell \to 0} b^l = \infty$.

[13] If $f(\overline{\theta}) > 0$, then (14) and (15) imply that $\dot{\Delta}(\overline{\theta}) < 0$. If $f(\overline{\theta}) = 0$, then $f'(\theta)/f(\theta) \to -\infty$ as $\theta \to \overline{\theta}$, so in either case $\limsup_{\theta \to \overline{\theta}} \dot{\Delta}(\theta) < 0$, implying $\dot{\Delta}(\theta) < 0$ for types near $\overline{\theta}$.

Since the ratio $f'(\cdot)/f(\cdot)$ is bounded, there exists a threshold for the loss, $\hat{\ell} > 0$, such that (18) is satisfied for all types if $\ell \in (0, \hat{\ell})$. ■

In the standard contracting model with quasilinear utility and private values, the MHRC implies complete sorting. Part (*i*) modifies that familiar condition: it is weaker than the MHRC for $\theta < 1/3$, stronger otherwise. In particular, if $\bar{\theta} \leq 1/3$, then the MHRC also implies complete sorting in our model. Common values and wealth effects make it hard to confirm whether the MHRC implies complete sorting when $\bar{\theta} > 1/3$.[14] In Section 6 we show that it does not: we calculate a closed-form solution for the CARA case, and find densities satisfying the MHRC for which some bunching is optimal.

Part (*ii*) shows that sorting is complete if the MLRP holds and either the highest type below 1/2 or the density is nondecreasing. It is easy to check that (*ii*) is satisfied 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. Part (iii) shows that if $f(\cdot)$ is C^1 and the likelihood ratio is bounded, then there is a region of losses for which sorting is complete.

In short, if either (*i*), (*ii*), or (*iii*) hold, then the omitted constraint (7) is satisfied.

4.3 Exclusion

It follows from (7) and (9) that the set of types that receive some insurance at the optimum is $[\theta_0, \overline{\theta}]$, 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) *Suppose that either Theorem 3 (i) or (ii) holds. If $f(\underline{\theta})$ is sufficiently large, then no type is excluded in an optimal menu.*

Proof. Appendix. ■

Intuitively, Proposition 2 shows that no type is excluded (and thus constraint (9) does *not* bind) if the presence of low types in the population is significant enough.

Proposition 3 (Exclusion) *Suppose that either Theorem 3 (i) or (ii) holds. There exists a $k > 0$ (depending on primitives) such that, if $f(\tilde{\theta})/(1-F(\tilde{\theta})) < k/\tilde{\theta}(1-\tilde{\theta})$, then all types $\underline{\theta} \leq \theta \leq \tilde{\theta}$ are excluded in an optimal menu.*

[14] One can show that if $-u'''(\cdot)/u''(\cdot) \leq -3u''(\cdot)/u'(\cdot)$ (a class which includes several commonly used functional forms), then we can replace 1/3 by 1/2.

Proof. Appendix. ■

The sufficient condition for exclusion in Proposition 3 depends *only* on the primitives of the problem, i.e., $u(\cdot)$, w, ℓ, and $f(\cdot)$. In particular, it suggests that a type is excluded if it is close to zero and if its presence in the population (density) is insignificant.

4.4 Curvature

We now turn to curvature of the premium as a function of coverage, including quantity discounts. The question is important for several reasons. First, firms commonly offer quantity discounts in practice, so it is natural to ask whether a monopolist insurer would use them. Second, in competitive insurance models, quantity *premia* rather of discounts are the rule since equilibrium prices equal marginal cost (in many competitive models); we would like to know if this implication holds for a monopolist.

Let $(x(\theta), t(\theta))_{\theta\in\Theta}$ be an optimal menu. Since the coverage $x(\cdot)$ cannot increase unless the premium $t(\cdot)$ increases, there is an increasing function $T(\cdot)$ on $[x(\underline{\theta}), \ell]$ such that $t(\theta) = T(x(\theta))$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$.[15] We want to know when $T(x)/x$ is nonincreasing. A simpler question is to determine when $T(\cdot)$ is concave and we focus on this property.

To begin, it is easy to show that $T(\cdot)$ cannot be concave if a positive measure of agents pool at any (x, t) with $x > 0$.[16] So we assume for the rest of this section that the optimal menu sorts types completely.

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 equation (42) in Section A.7):

$$\dot{T}(x(\theta)) = -U_x(x(\theta), t(\theta), \theta)/U_t(x(\theta), t(\theta), \theta). \tag{19}$$

We now determine the sign of the second derivative of $T(\cdot)$, which we denote by $\ddot{T}(\cdot)$.

[15] If $x(\underline{\theta}) > 0$, we can extend $T(\cdot)$ to all of $[0, \ell]$ by setting $T(x) = \tau(x)$, where $\tau(x)$ is defined by $U(x, \tau(x), \underline{\theta}) = U(0, 0, \underline{\theta})$. Note that if $x(\underline{\theta}) > 0$, then $T(\cdot)$ is concave on $[0, x(\underline{\theta})]$ and differentiable at $x(\underline{\theta})$.

[16] Suppose that $x(\theta_0) = x(\theta_1) = \tilde{x} > 0$ with $\theta_1 > \theta_0$. Setting $u_\ell = u(w - \ell + x(\theta) - t(\theta))$ and $u_n = u(w - t(\theta))$, we have $\dot{T}(\tilde{x}^-) \le \theta_0 u'_\ell/(\theta_0 u'_\ell + (1-\theta_0)u'_n) < \theta_1 u'_\ell/(\theta_1 u'_\ell + (1-\theta_1)u'_n) \le \dot{T}(\tilde{x}^+)$, so $T(\cdot)$ cannot be concave.

Lemma 3 (Curvature) *Let $T(\cdot)$ be an optimal nonlinear premium schedule that completely sorts types. We have $\ddot{T}(x(\theta)) < 0$ if and only if*

$$\frac{f'(\theta)}{f(\theta)} > \frac{3\theta - 2 + c(\theta)}{\theta(1-\theta)}, \tag{20}$$

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

Proof. Appendix. ■

Note that $c(\theta) < 1$ for all θ. Thus, (20) implies that the premium schedule is concave if $f'(\theta)/f(\theta) > (3\theta - 1)/\theta(1-\theta)$, which holds for example if $f(\cdot)$ is uniform with $\overline{\theta} < 1/3$.

As with our sorting lemma, an objection to (20) is that $c(\cdot)$ is endogenous; but in some cases, it gives us a complete description of the curvature of $T(\cdot)$.

Example 1 (Uniform distribution, log utility) *Let $f(\cdot)$ be uniform on $[0, \overline{\theta}]$ with $\overline{\theta} > \frac{1}{2}$ and $u(\cdot) = \log(\cdot)$. By Theorem 3, sorting is complete. Moreover $c(\theta) = \theta$ in this case. Since $f'(\theta) = 0$ for all θ, we have by (20) that $T(\cdot)$ is 'backwards-S shaped,' concave on $[0, x(\frac{1}{2})]$ and convex on $[x(\frac{1}{2}), \ell]$. Thus it exhibits quantity discounts, at least for small coverage levels. If $f(\cdot)$ is uniform on just $[0, \frac{1}{2}]$, then $T(\cdot)$ is globally concave, and exhibits quantity discounts globally.*

The curvature property in this example holds far more generally. Under the MLRP, the left side is decreasing. If the right side were increasing, then the backwards-S shaped property would hold. The endogenous $c(\theta)$ makes monotonicity of the right side hard to check, but if preferences satisfy DARA, then we can show that the right side of (20) crosses the left side at most once and from below.

Theorem 4 (Backwards S-shaped Premium) *Let $T(\cdot)$ be an optimal schedule that completely sorts types, and suppose that both the MLRP and DARA hold. Then*

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

(ii) if $f'(\cdot)$ takes positive and negative values, $T(\cdot)$ is strictly concave on an interval of positive length if $\underline{\theta} < 1/3$ and strictly convex on an interval of positive length if $\overline{\theta} > 2/3$.

Proof. (*i*) Denote the right side of (20) by $g(\theta)$. We first show that $c'(\theta) \geq 0$ implies that $g'(\theta) > 0$. We have

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

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 $\dot{T}(x) = \theta u'_\ell / [(1-\theta)u'_n + \theta u'_\ell]$ (equation (19)); rearrange to find $(1-\theta)u'_n/\theta u'_\ell = (1/\dot{T}) - 1 > 0$ and use the equality to rewrite $c(\theta)$ as

$$c(\theta) = \frac{1}{1 + \frac{r_\ell}{r_n}(\frac{1}{\dot{T}} - 1)}, \tag{21}$$

where r_i is the Arrow-Pratt risk aversion measure in state $i = \ell, n$. Hence

$$c'(\theta) = -\Omega \left[\frac{\partial \frac{r_\ell}{r_n}}{\partial \theta} \left(\frac{1}{\dot{T}} - 1 \right) - \frac{r_\ell}{r_n} \frac{\ddot{T}\dot{x}}{\dot{T}^2} \right], \tag{22}$$

where $\Omega = (1 + \frac{r_\ell}{r_n}(\frac{1}{\dot{T}} - 1))^{-2}$. Since the menu is increasing in θ, DARA implies that $\partial \frac{r_\ell}{r_n} / \partial \theta \leq 0$. By (22), if $\ddot{T}(x(\theta_0)) \geq 0$, 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 $\{x(\theta) | \theta \in (\hat{\theta}, \overline{\theta}]\}$ and concave otherwise. Setting $\hat{x} = x(\hat{\theta})$ completes the proof that $T(\cdot)$ is backwards-S shaped.

(*ii*) This result follows from (20) 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 the MLRP on the density and DARA on preferences. We are not aware of such a curvature result in other monopoly pricing models.[17]

For some intuition 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

[17] Spence (1977) considers a nonlinear pricing problem of allocating a good in fixed supply, and finds that the a tariff chosen to maximize a weighted sum of utilities can be backwards 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.

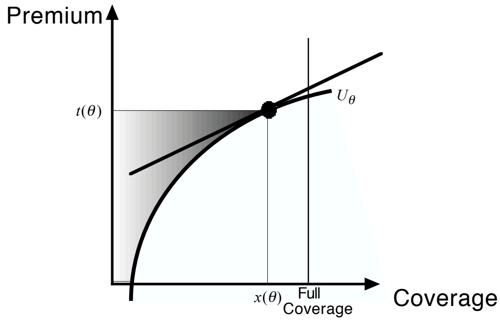

Figure 1: If an interior type gets 'close enough' to its first-best contract, then the premium cannot be convex below, or concave above that type's coverage

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 $[0, x(\theta)]$.[18] By equation (19), $\dot{T}(x(\theta))$ equals the slope of the indifference curve of θ at $(x(\theta), t(\theta))$. If $T(\cdot)$ were concave on $[x(\theta), \ell]$, then $\dot{T}(x) \leq \dot{T}(x(\theta))$ for all $x \in [x(\theta), \ell]$. But if $x(\theta)$ is 'close' to ℓ, then (see equation (19)) $\dot{T}(x(\theta)) \approx \theta < \overline{\theta} = \dot{T}(x(\overline{\theta}))$, which is inconsistent with profit maximization. In short, if some interior type's contract is close enough to its first-best contract, then $T(\cdot)$ at least *cannot be S-shaped* (first convex, then concave). Why should an interior type have a contract 'close' to its first best? First, interior types are relatively more likely under the MLRP (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, optimal menus push contracts for some interior types 'close' to their first-best.

Quantity discounts over some range *distinguishes* the monopoly case from the competitive model of Rothschild and Stiglitz (1976). In that model, 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.[19] Theorem 4 implies that this is *not* an implication of adverse selection as such, but from the *joint* imposition of adverse selection and (some form of) perfect competition.

[18] Graphically, the indifference curve of this type passes 'close' to the origin, since this type gets little surplus. Strict concavity of indifference curves then rules out convexity of $T(\cdot)$ on $[0, x(\theta)]$.

[19] See Wilson (1993), pp. 382-84, for another insurance example with quantity premia.

Theorem 4 also has implications for empirical work on insurance. For instance, Cawley and Philipson (1999) use the quantity premium implication from Rothschild and Stiglitz (1976) to test for the presence of adverse selection in (term) life insurance. 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. They conclude that the estimated affine function–which implies quantity discounts–is evidence against adverse selection in life insurance. But by Theorem 4, their empirical finding *could* be consistent with a monopolist facing adverse selection.

5 Comparison with the Standard Monopoly Model

In their seminal monopoly pricing paper, Maskin and Riley (1984) 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 (p. 172). The cost of selling x units to any type is cx, where c is a positive constant. Thus theirs is a private-values model with no wealth effects on demand, in contrast to our insurance model.

Maskin and Riley (1984) restrict attention either to a finite number of types or a continuum of types distributed according to an atomless density. A careful inspection of the proofs of Lemma 1 and Theorem 1, however, reveals that they can be adapted to their model. That is, each property has an analogue in the standard monopoly model, and can be derived for a *general* type distribution. For example, in that setting Lemma 1 says that a change in the quantity in the direction of a type's first-best quantity increases the monopolist's profit if the payment rises enough to make the type worse off. (One can even introduce common values in the monopoly pricing model and adapt these results, so long as the first best quantity is increasing in types.)

Arguably, the most important difference between the two models is curvature. For example, in Maskin and Riley (1984) there are always quantity discounts for the highest types; not so for a monopoly insurer. To see why, note that

$$\frac{d(t(\theta)/x(\theta))}{d\theta} = \frac{\dot{x}(\theta)}{x(\theta)} \left(\frac{\dot{t}(\theta)}{\dot{x}(\theta)} - \frac{t(\theta)}{x(\theta)} \right) \geq \frac{\dot{x}(\theta)}{x(\theta)} \left(\theta - \frac{t(\theta)}{x(\theta)} \right) = -\frac{\dot{x}(\theta)}{x(\theta)^2} \pi(\theta),$$

with equality if and only if $\theta = \overline{\theta}$. The inequality comes from $\dot{t}(\theta)/\dot{x}(\theta) = \theta u'_\ell/(\theta u'_\ell + (1-\theta)u'_n) \geq \theta$, with equality if and only if $\theta = \overline{\theta}$ (see equation (42) in Section A.7). If $\dot{x}(\theta)\pi(\theta) < 0$, 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. In Maskin and Riley (1984), profit from every type is nonnegative, and is *always* 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.

Remark 1 (Profit monotonicity) *Chiappori, Jullien, Salanié, and Salanié (2006) also use a 'profit monotonicity' condition–profit does not increase from contracts with higher coverage–to test for adverse selection and competition. They point out that this assumption need not hold in Stiglitz (1977). To understand the issue more clearly, note that* $\dot{\pi}(\theta) = \dot{t}(\theta) - \theta\dot{x}(\theta) - x(\theta)$. *Since* $x(\bar{\theta}) = \ell$ *and* $\dot{t}(\bar{\theta}) - \bar{\theta}\dot{x}(\bar{\theta}) = 0$ *(by (40)-(41) in the Appendix), we have* $\dot{\pi}(\bar{\theta}) = -\ell < 0$*: profit monotonicity holds near the highest type. But if a type gets small enough coverage, then profit monotonicity fails, since* $\dot{t}(\theta) - \theta\dot{x}(\theta) > 0$ *for* $\theta < \bar{\theta}$ *(again by (40)-(41)). In particular, it fails if any type is excluded. But if the lowest type is offered high enough coverage, profit monotonicity could hold for a monopoly as well. We leave as an open question the conditions under which it fails or holds.*

6 The CARA Case

Wealth effects make it hard to find closed-form solutions. Under CARA, however, we can calculate the optimal menu, and we use it to illustrate properties of the insurance problem absent in the standard monopoly pricing problem.

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 $w - t - v(x,\theta)$, which represents the same preferences over contracts as $U(x,t,\theta)$. Since the certainty equivalent is linear in t, it is simplest to proceed is as in the standard monopoly model, solving directly for $x(\cdot)$ without transforming the variables.[20]

Solution to the Relaxed Problem. 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,

[20] More precisely, replace $t(\cdot)$ from the objective function, integrate by parts, and maximize pointwise with respect to $x(\cdot)$ ignoring the monotonicity condition.

which simplifies to (setting $\xi = e^{r(\ell - x)}$ and recalling that ρ is the hazard rate)

$$\theta(1-\theta)[(1-\theta)+\theta\xi](\xi-1) - \frac{\xi}{\rho(\theta)} = 0. \tag{23}$$

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

$$\xi(\theta) = \frac{\frac{1}{\rho(\theta)} - \theta(1-\theta)(1-2\theta) + \sqrt{\left(\theta(1-\theta)(1-2\theta) - \frac{1}{\rho(\theta)}\right)^2 + 4\theta^3(1-\theta)^3}}{2\theta^2(1-\theta)}. \tag{24}$$

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.

Losses with Certainty and Partial Insurance at the Top. The proof in Theorem 1 that, wlog, the highest type gets full coverage uses the assumption that $\overline{\theta} < 1$. The highest type need *not* get full coverage when $\overline{\theta} = 1$, i.e., when the highest type suffers a loss with certainty. To illustrate, let the density be uniform on $[\underline{\theta}, 1]$, so that $\rho(\theta) = 1/(1-\theta)$. Simplify (24) to find

$$\xi(\theta) = \frac{1 - \theta + 2\theta^2 + \sqrt{1 - 2\theta + 5\theta^2}}{2\theta^2},$$

and evaluate at $\theta = 1$ to find $\xi(1) = 2$ and so $x(1) = \ell - (\log 2/r) < \ell$. In this case, the optimal menu is *uniformly* bounded away from full insurance.

MHRC Does Not Imply Complete Sorting. As noted, the MHRC implies complete sorting in many screening models. Since none of the conditions in Theorem 3 follow from the MHRC, it leaves open whether the MHRC implies complete sorting in our insurance model. Equation (24) suggests that it does not. Let $\overline{\theta} > 2/3$ and let $f(\cdot)$ be any density such that $\rho(\cdot)$ is globally increasing in θ but constant at a type $\hat{\theta} > 2/3$ (i.e., $\rho'(\hat{\theta}) = 0$). Then the left side of (23) is decreasing in θ at $\hat{\theta}$, and $\xi(\cdot)$ fails to be decreasing at that point.[21]

For an 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

[21] Consider the density $f(\theta) = ke^{-k\theta}$ if $\theta \in [\underline{\theta}, \tilde{\theta}]$ and $f(\theta) = ke^{-k\hat{\theta}}$ otherwise, where $k = 1/(\overline{\theta} - \tilde{\theta})$. Then $x(\cdot)$ is not increasing in a neighborhood of $\hat{\theta}$.

expression into (24) and set $\eta = 30$ and $\overline{\theta} = 0.9$ to find that $\xi(0.5) = 1.14 < 1.2 = \xi(0.8)$, so that $x(\cdot)$ is not increasing everywhere: Complete sorting fails even under the MHRC.

The example provides a clear intuition for why the MHRC does not imply complete sorting. Suppressing for a moment the common values feature of the model, suppose that the insurer's unit cost of coverage is constant and equal to $\underline{\theta}$. Then the first order condition for the relaxed problem is $v_x(x,\theta) - \underline{\theta} = v_{x\theta}(x,\theta)/\rho(\theta)$ for each θ. The marginal gain from an increase in coverage (the left side) is increasing in θ, while the marginal cost per unit of type θ of the additional information rent to all higher types (right side) is decreasing in θ if the MHRC holds. So without common values or wealth effects, the MHRC implies that the solution to the relaxed problem is increasing. With common values, the left side is no longer increasing in θ (i.e., $v_{x\theta}(x,\theta)$ is not always greater than 1), and a strengthening of the MHRC is needed.[22] With wealth effects this problem is compounded, since the MHRC does not even imply that $\int_{\theta}^{\overline{\theta}} a(s)f(s)ds/f(\theta)$ is decreasing in θ (see (16)), explaining why we impose the stronger conditions of Theorem 3.

Common Values and Curvature. From Theorem 4, if $f(\cdot)$ satisfies the MLRP and $\overline{\theta} > 2/3$, the premium is convex in coverage for $\theta \in (2/3, \overline{\theta}]$. The CARA case illustrates the role of common values in curvature of the premium. Suppress once again common values, so we are back in a standard monopoly pricing model with quasilinear utility and no common values. Following Maskin and Riley (1984), the necessary and sufficient condition for the premium to be strictly concave in coverage is that, for all θ,

$$\frac{\rho'(\theta)}{\rho(\theta)} + \left[\frac{v_{xx\theta}(x(\theta),\theta)}{v_{xx}(x(\theta),\theta)} - \frac{v_{x\theta\theta}(x(\theta),\theta)}{v_{x\theta}(x(\theta),\theta)}\right] > 0. \tag{25}$$

After tedious algebra, (25) becomes $\rho'(\theta)/\rho(\theta) > (2\theta - 1)/(\theta(1-\theta))$ for all θ: any density satisfying the MLRP and this inequality implies a concave premium *globally* (and so also quantity discounts) in the CARA case. For a simple example, consider the uniform on $[0, \overline{\theta}]$. Then $\rho'(\theta)/\rho(\theta) = 1/(\overline{\theta} - \theta) > (2\theta - 1)/(\theta(1-\theta))$ for all θ.

[22]Indeed, that the left side is not increasing is why there can be pooling even in the first-best case (observable types): all types get full coverage, and two distinct types can pay the same premium. In Maskin and Riley, the first-best menu completely sorts types.

7 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. But despite the importance of adverse selection in insurance and well-known problems with its competitive provision, the monopoly case has received surprisingly little attention.

Insurance markets surely lie somewhere in between competition and monopoly. Arguably, monopoly is the right place to start thinking about noncompetitive insurance markets: for insurance with adverse selection, there is no agreement on what a good model even of competition is, let alone oligopoly. Indeed, Cohen and Einav (2007) argue that a monopoly model describes their data for an entrant into the Israeli auto insurance market better than a competitive one.

We have stressed the differences between monopoly and competitive insurance with adverse selection. One similarity is monotonicity: riskier types buy more coverage; equivalently, those who buy more coverage experience higher losses on average. Many tests of adverse selection focus on this prediction. Monotonicity follows from incentive compatibility and the single crossing property, which in turn follows from the definition of a type (the loss chance) and the expected utility hypothesis. Relaxing either one can overturn this property–and likely monotonicity as well. For example, single crossing can fail if the type includes the agent's risk attitudes as well as riskiness; and it can fail if the agent's preferences violate expected utility.[23] Both extensions are worth pursuing.

Two other extensions are worth mentioning: allowing more than one loss amount; and more than one period. If the loss takes on more than one value, but the private information is still purely about likelihood of a loss, not its magnitude, then the principal will offer a menu of deductible insurance contracts. Since a deductible contract is still two-dimensional, many of our proofs can be adapted to this case. The multi-period case raises several interesting issues such as renegotiation and experience rating, and should reveal further implications of adverse selection for insurance.

[23] Ormiston and Schlee (2001) show it fails for mean-variance preferences. Since some nonexpected utility representations satisfy the single crossing property, we conjecture that adding that assumption will preserve the conclusion of Theorem 1, but not that of Theorems 2-4.

A 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 bounded and satisfy (IC), with $x(\theta) \leq \ell$ for all $\theta \in \Theta$. Then $U(x(\theta), t(\theta), \theta)$ is decreasing and continuous in θ.*

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

Monotonicity implies that the left and right limits exist at any $\theta \in \Theta$. Let $\theta' \in \Theta$. We will show that the left and right limits of $U(x(\theta), t(\theta), \theta)$ 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

$$0 \geq U(x(\theta'), t(\theta'), \theta') - U(x_n, t_n, \theta_n) \geq U(x_n, t_n, \theta') - U(x_n, t_n, \theta_n), \tag{26}$$

where the first inequality follows from monotonicity of $U(x(\theta), t(\theta), \theta)$ in θ, and the second from (IC). But $U(x_n, t_n, \theta') - U(x_n, t_n, \theta_n) = (\theta' - \theta_n)(u(w - \ell - t_n + x_n) - u(w - t_n))$. Since $(x(\cdot), t(\cdot))$ is bounded and $u(\cdot)$ continuous, $U(x_n, t_n, \theta') - U(x_n, t_n, \theta_n)$ tends to 0 as θ_n tends to θ', so by (26), $U(x(\theta), t(\theta), \theta)$ is left-continuous at $\theta = \theta'$.

Now consider any sequence θ_n approaching θ' from above. For every n,

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

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

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

(i) Premium, Indemnity, and Net Indemnity are Co-Monotone. Wlog, let $\theta' > \theta$. From (IC), $U(x(\theta), t(\theta), \theta) \geq U(x(\theta'), t(\theta'), \theta)$ and $U(x(\theta'), t(\theta'), \theta') \geq U(x(\theta), t(\theta), \theta')$, so that either (a) $x(\theta') - t(\theta') \geq x(\theta) - t(\theta)$ and $t(\theta') \geq t(\theta)$, or (b) $x(\theta') - t(\theta') \leq x(\theta) - t(\theta)$ and $t(\theta') \leq t(\theta)$ (the other remaining cases are easily dismissed by (IC)). Unless the contracts are the same, case (b) is ruled out by the single crossing property (if the low type prefers a higher contract, so does the high type). Thus, $x(\theta') - t(\theta') \geq x(\theta) - t(\theta)$ and $t(\theta') \geq t(\theta)$, which together imply that $x(\theta') \geq x(\theta)$.

(ii) No Overinsurance. Let $(x(\cdot), t(\cdot))$ be a feasible menu with $x(\theta) > \ell$ on a positive measure set of types. Let $\hat{\theta}$ be the infimum of types with coverage greater than the loss ℓ. By monotonicity, *every* type $\theta > \hat{\theta}$ must have $x(\theta) > \ell$. There are two cases.

(a) $\hat{\theta} \in \{\theta \in \Theta | x(\theta) > \ell\}$. In this case $x(\hat{\theta}) > \ell$ and $x(\theta) \leq \ell$ for every type $\theta < \hat{\theta}$ in Θ. For each $\theta \in [\underline{\theta}, \hat{\theta}] \cap \Theta$, let $(\ell, \tilde{t}(\theta))$ be the contract that leaves type θ indifferent between his contract in the menu and this one; formally, $\tilde{t}(\theta))$ solves

$$U(\ell, \tilde{t}(\theta), \theta) \quad = \quad U(x(\theta), t(\theta), \theta).$$

Define $\hat{t} = \sup \tilde{t}(\theta)$, where the supremum is taken over $\{\theta \in \Theta | \theta \leq \hat{\theta}\}$. Pool all types $\theta \geq \hat{\theta}$ at $(\ell, \hat{t})$, and leave the contract for every other type unchanged (see Figure 2(a)). The new menu satisfies (IC) and (P) and increases profit by Lemma 1.

(b) $\hat{\theta} \notin \{\theta \in \Theta | x(\theta) > \ell\}$. Then $x(\hat{\theta}) \leq \ell$. For each type $\theta \geq \hat{\theta}$, let $\tau(\theta)$ solve

$$U(\ell, \tau(\theta), \theta) \quad = \quad U(x(\theta), t(\theta), \theta).$$

Define $\tau^* = \sup \tau(\theta)$, where the supremum is taken over $\{\theta \in \Theta | \theta \geq \hat{\theta}\}$. Pool all types $\theta > \hat{\theta}$ at (ℓ, τ^*), and leave the contract for every other type the same (see Figure 2(b)). The new menu satisfies (IC) and (P) and profit increases by Lemma 1.

(iii) 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 positive measure set of types. By monotonicity, that set contains all types below $\theta^s = \sup\{\theta \in \Theta | t(\theta) < 0\}$. There are two cases.

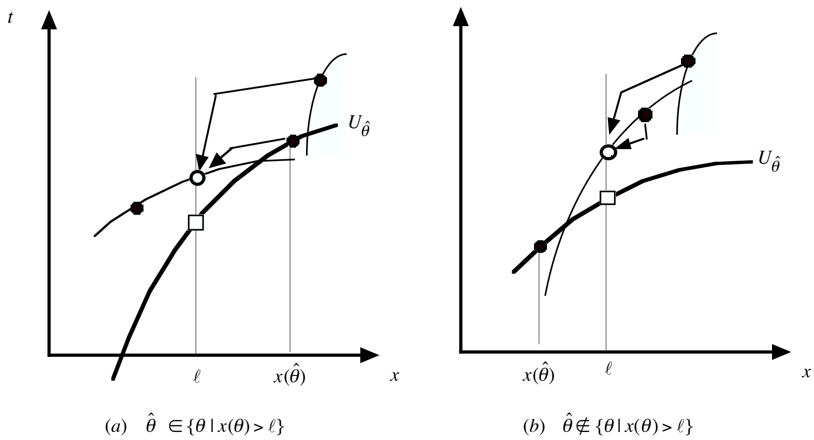

Figure 2: (ii) No overinsurance

(a) $\theta^s \in \{\theta \in \Theta | t(\theta) < 0\}$. In this case $t(\theta^s) < 0$. For each type $\theta \geq \theta^s$, find the contract $(\hat{x}(\theta), 0)$ that leaves type θ indifferent between his contract in the menu and this one; formally, $\hat{x}(\theta)$ solves

$$U(\hat{x}(\theta), 0, \theta) \quad = \quad U(x(\theta), t(\theta), \theta).$$

Since the menu satisfies (P), $\hat{x}(\theta) \geq 0$ for every θ. Let $x^m = \inf x(\theta)$, where the infimum is taken over all types $\theta \geq \theta^s$. (Note that $x^m \geq x(\theta)$ for all types $\theta \leq \theta^s$.) Pool all types $\theta \leq \theta^s$ at $(x^m, 0)$, and leave the contract the same for all other types (see Figure 3(a)). The new menu satisfies (IC) and (P) and profit increases by Lemma 1.

(b) $\theta^s \notin \{\theta \in \Theta | t(\theta) < 0\}$. In this case $t(\theta^s) \geq 0$. For each type $\theta \leq \theta^s$, let $\hat{x}(\theta)$ solve

$$U(\hat{x}(\theta), 0, \theta) = U(x(\theta), t(\theta), \theta),$$

and let $x^s = \inf \hat{x}(\theta)$, where the infimum is taken over $\{\theta \in \Theta | \theta \leq \theta^s\}$. Note that $x^s \geq x(\theta)$ for all types $\theta \leq \theta^s$. Pool all types with $t(\theta) < 0$ at $(x^s, 0)$, and leave the contract for every other type the same (see Figure 3(b)). The new menu satisfies (IC)

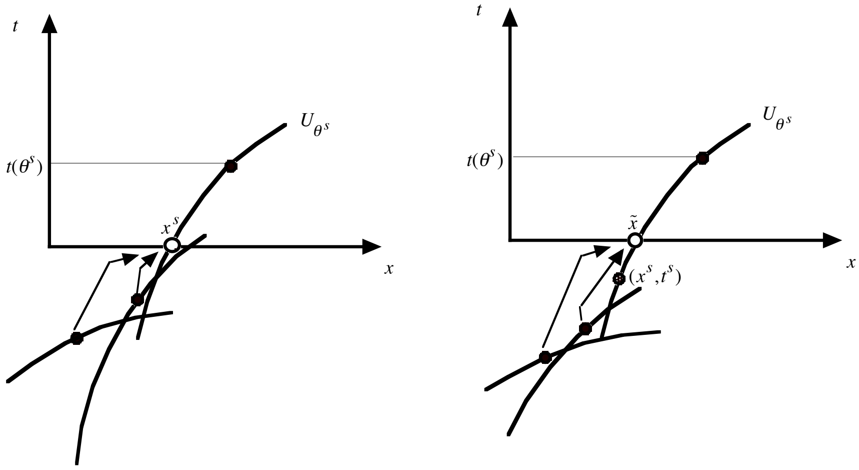

Figure 3: (iii) Nonnegativity

and (P) and profit increases by Lemma 1 and (*ii*).

(*iv*) Participation Binds at the Bottom. Let $(x(\cdot), t(\cdot))$ be a feasible menu with $U(x(\underline{\theta}), t(\underline{\theta}), \underline{\theta}) - U(0, 0, \underline{\theta}) = K > 0$. Since expected utility is continuous in θ, there is a nonnegative number ε such that $[\underline{\theta}, \underline{\theta}+\varepsilon]$ has positive measure and, for all $\theta \in [\underline{\theta}, \underline{\theta}+\varepsilon] \cap \Theta$,

$$U(x(\underline{\theta}), t(\underline{\theta}), \theta) - U(0, 0, \theta) > 0.$$

(If $\underline{\theta}$ itself has positive measure, then ε can be zero.) By (IC),

$$U(x(\theta), t(\theta), \theta) - U(0, 0, \theta) \geq U(x(\underline{\theta}), t(\underline{\theta}), \theta) - U(0, 0, \theta)$$

for all $\theta \in [\underline{\theta}, \underline{\theta} + \varepsilon] \cap \Theta$. But since the difference on the right side of this last inequality tends to K as θ tends to $\underline{\theta}$, there is a nonnegative real number η such that $[\underline{\theta}, \underline{\theta} + \eta]$ has positive measure and, for all $\theta \in [\underline{\theta}, \underline{\theta} + \eta] \cap \Theta$,

$$U(x(\theta), t(\theta), \theta) - U(0, 0, \theta) \geq K/2 > 0.$$

By (iii), there is a $\theta' \in [\underline{\theta}, \underline{\theta}+\eta] \cap \Theta$ with $(x(\theta') - t(\theta'), t(\theta')) \geq 0$, so $U(x(\theta'), t(\theta'), \theta) - U(0,0,\theta)$ is increasing in θ. Combine this fact with (IC) to conclude that

$$U(x(\theta), t(\theta), \theta) - U(0,0,\theta) \geq U(x(\theta'), t(\theta'), \theta) - U(0,0,\theta) \geq K/2 > 0$$

for all $\theta \in (\underline{\theta}+\eta, \bar{\theta}] \cap \Theta$, so $U(x(\theta), t(\theta), \theta) - U(0,0,\theta) \geq K/2 > 0$ for all $\theta \in \Theta$. Now change the menu to reduce the utility of each type by $K/2$ in each state. Both (P) and (IC) continue to hold, but profit increases.

(v) Full Insurance at the Top. We show that the essential supremum of $x(\cdot)$ is ℓ. Suppose that it is not ℓ for some feasible menu $(x(\theta), t(\theta))_{\theta\in\Theta}$. By (ii), wlog, it must be less than ℓ. We show that profit rises if we pool all 'sufficiently' high types at full insurance.

For each $\varepsilon \in (0, \bar{\theta} - \underline{\theta})$, let θ_ε be the smallest number in (the positive measure set) $[\bar{\theta} - \varepsilon, \bar{\theta}] \cap \Theta$. Let τ_ε be the premium for full insurance that leaves type-θ_ε indifferent between (ℓ, τ_ε) and $(x(\theta_\varepsilon), t(\theta_\varepsilon))$:

$$U(\ell, \tau_\varepsilon, \theta_\varepsilon) = U(x(\theta_\varepsilon), t(\theta_\varepsilon), \theta_\varepsilon).$$

By Lemma 4 we have $\lim_{\varepsilon\to 0} \tau_\varepsilon = \tau_0$, and τ_0 solves $U(\ell, \tau_0, \bar{\theta}) = U(x(\bar{\theta}), t(\bar{\theta}), \bar{\theta})$.

We now show that there is an $\varepsilon \in (0, \bar{\theta} - \underline{\theta})$ with $\tau_\varepsilon - \theta\ell > t(\theta) - \theta x(\theta)$ for every type $\theta \geq \bar{\theta} - \varepsilon$. Suppose to the contrary that, for all $\varepsilon \in (0, \bar{\theta} - \underline{\theta})$, there is a type $\theta(\varepsilon) \geq \bar{\theta} - \varepsilon$ such that

$$t(\theta(\varepsilon)) - \theta(\varepsilon)x(\theta(\varepsilon)) \geq \tau_\varepsilon - \theta(\varepsilon)\ell. \tag{27}$$

By Lemma 1, monotonicity, and $x < \ell$, we have that

$$\tau_0 - \theta(\varepsilon)\ell > t(\bar{\theta}) - \theta(\varepsilon)x(\bar{\theta}) \geq t(\theta(\varepsilon)) - \theta(\varepsilon)x(\theta(\varepsilon)), \tag{28}$$

since profit rises if $\theta(\varepsilon)$ gets the contract offered to type $\bar{\theta}$, and rises even more if $\theta(\varepsilon)$ gets (τ_0, ℓ). Letting ε go to zero in (27) and (28) yields $\tau_0 - \bar{\theta}\ell > \tau_0 - \bar{\theta}\ell$.[24] So for some $\varepsilon \in (0, \bar{\theta} - \underline{\theta})$, $\tau_\varepsilon - \theta\ell > t(\theta) - \theta x(\theta)$ for every type $\theta \geq \bar{\theta}$.

[24] Note that $\bar{\theta} < 1$ implies that the first inequality in (28) remains strict in the limit as ε goes to zero.

Fix such an ε and consider the following menu $(\hat{x}(\theta), \hat{t}(\theta))_{\theta\in\Theta}$: for $\theta \in [\bar{\theta} - \varepsilon, \bar{\theta}] \cap \Theta$, set $\hat{x}(\theta) = \ell$ and $\hat{t}(\theta) = \tau_\varepsilon$; otherwise set $(\hat{x}(\theta), \hat{t}(\theta)) = (x(\theta), t(\theta))$. This menu satisfies (IC) and (P) and has higher profit than $(x(\theta), t(\theta))_{\theta\in\Theta}$.

(vi) No Pooling at the Top. Suppose that the set of types *below* $\bar{\theta}$ receiving full coverage is of positive measure for some feasible menu $(x(\theta), t(\theta))_{\theta\in\Theta}$. Let $\hat{\theta}$ be the infimum of this set. We shall show that there is an alternative menu with higher profit.

By (IC), any type with full insurance is charged the same premium, call it τ. Since the menu is monotone and $x \leq \ell$, we have $x(\theta) < \ell$ if $\theta < \hat{\theta}$ and $x(\theta) = \ell$ if $\theta > \hat{\theta}$. By Lemma 4, $\lim_{\theta\to\hat{\theta}} U(x(\theta), t(\theta), \theta) = U(\ell, \tau, \hat{\theta})$. Wlog, we can set $x(\hat{\theta}) = \ell$ and $t(\hat{\theta}) = \tau$, since expected profit does not fall and (IC) and (P) still hold.

Fix $\varepsilon \in (0, \bar{\theta} - \hat{\theta})$. Let $\tau(\varepsilon)$ be the largest premium satisfying (P) for loss chance $\hat{\theta} + \varepsilon$ at full insurance: $U(\ell, \tau(\varepsilon), \hat{\theta} + \varepsilon) = U(0, 0, \hat{\theta} + \varepsilon)$.[25] For each $\delta \in (0, \tau(\varepsilon))$ let (x_δ, t_δ) be the contract that leaves type $\hat{\theta}$ indifferent between (ℓ, τ) and (x_δ, t_δ), and type $\hat{\theta} + \varepsilon$ indifferent between $(\ell, \tau + \delta)$ and (x_δ, t_δ). Formally, (x_δ, t_δ) solves (see Figure 4)

$$\begin{aligned} U(\ell, \tau, \hat{\theta}) &= U(x_\delta, t_\delta, \hat{\theta}) \\ U(\ell, \tau + \delta, \hat{\theta} + \varepsilon) &= U(x_\delta, t_\delta, \hat{\theta} + \varepsilon) \end{aligned}$$

Define a menu of contracts $(\hat{x}(\theta), \hat{t}(\theta))_{\theta\in\Theta}$ as follows: $(\hat{x}(\theta), \hat{t}(\theta)) = (\ell, \tau + \delta)$ if $\theta \in (\hat{\theta} + \varepsilon, \overline{\theta}] \cap \Theta$; $(\hat{x}(\theta), \hat{t}(\theta)) = (x_\delta, t_\delta)$ if $\theta \in [\hat{\theta}, \hat{\theta} + \varepsilon] \cap \Theta$; and $(\hat{x}(\theta), \hat{t}(\theta)) = (\tilde{x}_\delta(\theta), \tilde{t}_\delta(\theta))$ if $\theta \in [\underline{\theta}, \hat{\theta}) \cap \Theta$, where $(\tilde{x}_\delta(\theta), \tilde{t}_\delta(\theta))$ is equal to the *best* of the two contracts $(x(\theta), t(\theta))$ or (x_δ, t_δ) for type θ. The menu $(\hat{x}(\theta), \hat{t}(\theta))_{\theta\in\Theta}$ satisfies (IC) and (P) for every $\delta \in (0, \tau(\varepsilon))$, and its expected profit is

$$\int_{(\hat{\theta}+\varepsilon, \overline{\theta}]} [\tau + \delta - \theta\ell] dF(\theta) + \int_{[\hat{\theta}, \hat{\theta}+\varepsilon]} [t_\delta - \theta x_\delta] dF(\theta) + \int_{[\underline{\theta}, \hat{\theta})} [\tilde{t}_\delta(\theta) - \theta \tilde{x}_\delta(\theta)] dF(\theta). \quad (29)$$

We now show that each of the three expressions is 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 term at $\delta = 0$ is $\int_{(\hat{\theta}+\varepsilon, \overline{\theta}]} dF(\theta) > 0$ for every $\varepsilon \in [0, \bar{\theta} - \hat{\theta})$.

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

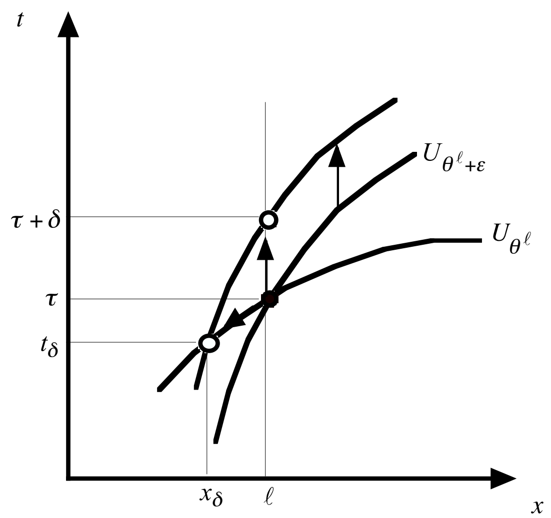

Figure 4: No pooling at the top

Since u is C^1, x_δ and t_δ are differentiable in δ for every $\delta \in [0, \tau)$, $t'_\delta - \theta x'_\delta$ is bounded on $[0, \tau) \times \Theta$, and tedious algebra reveals that $t'_0 - \theta x'_0 = (\hat{\theta} - \theta)/\varepsilon \leq 0$ for $\theta \in [\hat{\theta}, \hat{\theta} + \varepsilon]$. So the derivative of the second integral in (29) with respect to δ exists at $\delta = 0$ and equals $\int_{[\hat{\theta},\hat{\theta}+\varepsilon]} \frac{\hat{\theta}-\theta}{\varepsilon} dF(\theta) \leq 0$. By the mean-value theorem, for every $\varepsilon \in (0, \bar{\theta} - \hat{\theta})$, there is a number $\psi(\varepsilon) \in [\hat{\theta}, \hat{\theta} + \varepsilon]$ such that

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

Since $F(\cdot)$ is right-continuous, the last expression tends to zero as ε tends to zero.

Finally, consider the third integral in (29). The integrand is differentiable in δ at $\delta = 0$ and it is easy to show that the value of the derivative is zero for every $\theta \in \Theta \cap [\underline{\theta}, \hat{\theta})$. Moreover, since x_δ and t_δ are decreasing in δ, it follows that, for each $(\delta, \theta) \in (0, \tau(\varepsilon)) \times \left(\Theta \cap [\underline{\theta}, \hat{\theta})\right)$, $|\frac{\tilde{x}_\delta(\theta) - x(\theta)}{\delta}| \leq |\frac{x_\delta - x_0}{\delta}|$ and $|\frac{\tilde{t}_\delta(\theta) - t(\theta)}{\delta}| \leq |\frac{t_\delta - t_0}{\delta}|$, which shows that the derivatives of x_δ and t_δ are bounded on $[0, w - \tau]$. Hence, the third integral is differentiable at $\delta = 0$ and the derivative equals zero for every $\varepsilon \in [0, \bar{\theta} - \hat{\theta})$. Since the derivative of (29) with respect to δ is positive for some $\varepsilon > 0$, the original menu is not optimal.

(*vii*) The Principal Makes Positive Profit. Let $\varepsilon > 0$, and consider a menu in which each type chooses either $(0,0)$ 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 $\bar{\theta} < 1$, for $\varepsilon > 0$ small enough, we have that $\overline{\theta}\ell + \varepsilon < \ell$, so a positive measure of types in $[\underline{\theta}, \bar{\theta}]$ choose $(\ell, \overline{\theta}\ell+\varepsilon)$ and expected profit from this menu is positive. ■

A.2 Proof of Continuous Differentiability of $U(\theta)$

If $f(\cdot)$ is continuous on $\Theta = [\underline{\theta}, \overline{\theta}]$ and positive on $(\underline{\theta}, \overline{\theta})$, then $U(\cdot)$ is C^1 on $(\underline{\theta}, \overline{\theta})$. We prove this fact in three steps.

No Indifference at an Optimal Menu. We first show that if $(x(\theta), t(\theta))_{\theta\in\Theta}$ is optimal, then each type strictly prefers its contract to any other one in the menu. Consider a feasible menu $(x(\theta), t(\theta))_{\theta\in\Theta}$ violating this property: for two types θ', θ'' in Θ we have $(x(\theta''), t(\theta'')) \neq (x(\theta'), t(\theta'))$ and $U(\theta'') = U(x(\theta'), t(\theta'), \theta'')$. We will show that there is a feasible menu with higher expected profit.

There are two cases to consider: (*i*) $\theta'' > \theta'$; and (*ii*) $\theta' > \theta''$.

Consider case (*i*). The single-crossing property and (IC) imply that types in (θ', θ'') are pooled at $(x(\theta'), t(\theta'))$ (each type prefers $(x(\theta'), t(\theta'))$ to $(x(\theta''), t(\theta''))$, and any other distinct contract that a type $\theta \in (\theta', \theta'')$ likes as least as much as $(x(\theta'), t(\theta'))$ will be envied by either θ' or θ''). For any $\varepsilon \in (0, \theta'' - \theta')$, give all types in $[\theta' + \varepsilon, \theta'')$ a contract $(x(\theta') + \delta, \tau_\delta)$ satisfying $\delta \in (0, x(\theta'') - x(\theta'))$ and $U(x(\theta') + \delta, \tau_\delta, \theta'') = U(\theta'')$; and give types in $[\theta', \theta' + \varepsilon)$ a contract $(x(\theta') + d_\delta, p_\delta)$ satisfying the following equations:

$$\begin{aligned} U(x(\theta') + d_\delta, p_\delta, \theta' + \varepsilon) &= U(x(\theta') + \delta, \tau_\delta, \theta' + \varepsilon) \\ U(x(\theta') + d_\delta, p_\delta, \theta') &= U(\theta'). \end{aligned}$$

Leave the contract for every other type unchanged. The new menu satisfies (IC) and (P) for each $\varepsilon \in (0, \theta'' - \theta')$ and each δ satisfying the given equalities. The change in expected profit is

$$\int_{[\theta'+\varepsilon,\theta'')} [\tau_\delta - t(\theta') - \theta\delta] dF(\theta) + \int_{(\theta',\theta'+\varepsilon)} [p_\delta - t(\theta') - \theta d_\delta] dF(\theta).$$

Each of the functions τ_δ, p_δ, and d_δ are continuously differentiable in δ with derivatives uniformly bounded on $(0, x(\theta'') - x(\theta')) \times [\theta', \theta'']$. Hence the change in expected profit is differentiable (for each $\varepsilon \in (0, \theta'' - \theta')$) and equals

$$\int_{[\theta'+\varepsilon,\theta'')} [\tau_0' - \theta] dF(\theta) + \int_{(\theta',\theta'+\varepsilon)} [p_0' - \theta d_0'] dF(\theta).$$

As $\varepsilon \to 0$, the second term vanishes, while the first tends to $\int_{[\theta',\theta'')} [\tau_0' - \theta] dF(\theta)$, which is positive by Lemma 1, so the original menu $(x(\theta), t(\theta))_{\theta\in\Theta}$ cannot be optimal.

Now consider case (ii). As in case (i) any types in (θ'', θ') are pooled at $(x(\theta'), t(\theta'))$. Let $\varepsilon \in (0, \theta' - \theta'')$. For each type in $[\theta'' + \varepsilon, \overline{\theta}]$, consider a new contract that lowers the utility in each state by $\delta > 0$ satisfying $U(\theta'' + \varepsilon) - \delta > U(x(\theta'), t(\theta'), \theta'' + \varepsilon)$. Give types in $(\theta', \theta' + \varepsilon)$ the contract at the intersection of the indifference set of type θ' (through $(x(\theta'), t(\theta'))$) and type $\theta'' + \varepsilon$ (through the new contract for it). The contract remains the same for all other types. By the single crossing property and feasibility of $(x(\theta), t(\theta))_{\theta\in\Theta}$, the new menu satisfies (P) and (IC). Moreover, on the set $(\theta'' + \varepsilon, \overline{\theta}]$, expected profit rises and has a positive derivative with respect to δ at $\delta = 0$ for every $\varepsilon \in [0, \theta' - \theta'')$. As in case (i), the derivative of expected profit conditional on the complement of types wrt δ at $\delta = 0$ tends to zero as ε tends to zero. Since $(\theta'', \overline{\theta}] = \cup_{\varepsilon>0}[\theta''+\varepsilon, \overline{\theta}]$, $(\theta'', \theta') \subset (\theta'', \overline{\theta}]$, and by hypothesis (θ'', θ') is of positive measure, we have that for some $\varepsilon \in (0, \theta' - \theta'')$, expected profit from the new (feasible) menu is greater than expected profit from $(x(\theta), t(\theta))_{\theta\in\Theta}$.

$U(\cdot)$ is Differentiable. Let $(x(\theta), t(\theta))_{\theta\in\Theta}$ be an optimal menu. Since for any (χ, τ), $U(\chi, \tau, \cdot)$ is differentiable on Θ, and since, for each $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, the contract in $\{x(\theta'), t(\theta') | \theta' \in \Theta\}$ which maximizes $U(\cdot, \cdot, \theta)$ is unique, it follows from the Envelope Theorem (Milgrom and Segal (2002), Theorem 3) that $U'(\theta)$ exists and equals $-\Delta(\theta)$.

$U(\cdot)$ is Continuously Differentiable. Since $U(\chi, \tau, \cdot)$ is affine in θ, the indirect utility $U(\cdot)$ is convex. Since $U(\cdot)$ is also differentiable it follows from Rockafellar (1970) (Corollary 25.5.1) that it is C^1 on $(\underline{\theta}, \overline{\theta})$. ■

A.3 Proof of Lemma 2 (Complete Sorting)

Using (13)-(14) to eliminate $\dot{\lambda}(\theta)$, rewrite the numerator of (17) as $\lambda(\theta)B(\theta)$, where

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

Since

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

and $\lambda(\theta) \leq 0$, it follows that $\dot{\Delta}(\theta) \leq 0$ *if and only if*

$$\frac{f'(\theta)}{f(\theta)}\theta(1-\theta) + (2-3\theta) + \frac{h'(U(\theta)-(1-\theta)\Delta(\theta))}{h'(U(\theta)+\theta\Delta(\theta)) - h'(U(\theta)-(1-\theta)\Delta(\theta))} \geq 0,$$

or, equivalently,

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

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

A.4 Proof of Theorem 3 (Complete Sorting: Sufficiency)

(i) We show that the following condition is sufficient for complete sorting:

$$\frac{f'(\theta)}{f(\theta)} \quad > \quad \frac{3\theta-1}{\theta(1-\theta)} - \frac{f(\theta)}{1-F(\theta)}. \tag{30}$$

Fix $\hat{\theta} \in [\underline{\theta}, \bar{\theta})$. We first claim that if $\dot{\Delta}(\tau) < 0$ for all $\tau \in (\hat{\theta}, \bar{\theta})$, and condition (30) holds, then $\dot{\Delta}(\hat{\theta}) < 0$. To establish this claim, we show that

$$b(\hat{\theta}) > -1 + \frac{f(\hat{\theta})}{1-F(\hat{\theta})}\hat{\theta}(1-\hat{\theta}), \tag{31}$$

implying that the sufficient (and necessary) condition (18) holds at $\hat{\theta}$.

Let $h_n'(\theta) = h'(U(\theta) - \theta\Delta(\theta))$. Since $\dot{\Delta}(\tau) < 0$ and $f(\tau) > 0$ for all $\tau \in (\hat{\theta}, \bar{\theta})$,

$$f(\tau)h_n'(\hat{\theta}) > f(\tau)h_n'(\tau) > f(\tau)a(\tau), \tag{32}$$

for all $\tau \in (\hat{\theta}, \bar{\theta})$, with equalities at $\bar{\theta}$, where $a(\cdot)$ is defined in equation (16).

Integrate both sides of (32) from $\hat{\theta}$ to $\bar{\theta}$ and divide by $1 - F(\hat{\theta})$ to obtain

$$h'_n(\hat{\theta}) > \frac{1}{1 - F(\hat{\theta})} \int_{\hat{\theta}}^{\bar{\theta}} a(\tau) f(\tau) d\tau = (\Delta h)'(\hat{\theta}) \frac{\hat{\theta}(1 - \hat{\theta}) f(\hat{\theta}),}{1 - F(\hat{\theta})} \tag{33}$$

where $(\Delta h)'(\theta) = h'(U(\theta) + \theta\Delta(\theta)) - h'(U(\theta) - (1 - \theta)\Delta(\theta))$ and we have used (16). Add $-\hat{\theta}(\Delta h)'(\hat{\theta})$ to both sides of (33) and rearrange to get (31), so that the sufficient condition (18) holds at $\hat{\theta}$.

It now follows that, under condition (30), $\dot{\Delta}(\theta) < 0$ for all $\theta \in \Theta$: if $\dot{\Delta} \geq 0$ somewhere, then there would be a largest $\theta \in [\underline{\theta}, \bar{\theta})$ with $\dot{\Delta}(\theta) \geq 0$ (since $\dot{\Delta}(\cdot)$ is continuous and $\limsup_{\theta \to \bar{\theta}} \dot{\Delta}(\theta) < 0$). By the claim, condition (30) would fail.

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

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

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

$$g(\theta) = f'(\theta)/f(\theta) - (3\theta - 2 - b(\theta))/\theta(1 - \theta). \tag{34}$$

Moreover,

$$b'(\theta) = -\dot{\Delta}(\theta) \left[\frac{h'_n h''_\ell (1 - \theta) + h''_n h'_\ell \theta}{(h'_n - h'_\ell)^2} \right], \tag{35}$$

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

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

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

To show that $f'(\cdot)$ must sometimes be negative, rewrite (34) as $\tilde{g}(\theta) = (1 - \theta) g(\theta) = (1 - \theta)(f'(\theta)/f(\theta)) - (3\theta - 2 - b(\theta))/\theta$. Since $\dot{\Delta}(\theta) < 0$ for all $\theta > \hat{\theta}$, and $\tilde{g}(\hat{\theta}) = 0$, we must have $\tilde{g}'(\hat{\theta}) \geq 0$. But since $b'(\hat{\theta}) = 0$, the fraction $(3\theta - 2 - b(\theta))/\theta$ is strictly increasing in a neighborhood of $\hat{\theta}$, so $(1 - \theta) f'(\theta)/f(\theta)$ must be increasing in a neighborhood of $\hat{\theta}$, so $f'(\cdot)$ is negative on a neighborhood of $\hat{\theta}$. ■

A.5 Proof of Proposition 2 (No Exclusion)

Type $\underline{\theta}$ is not excluded from the optimal menu of contracts if $\Delta(\tilde{\theta}) > \Delta_0$. From equation (16) and the concavity of the optimal control problem, this is tantamount to showing that the marginal benefit of providing insurance to $\tilde{\theta}$ starting from no insurance is bigger than the marginal cost of doing so. Formally,

$$f(\underline{\theta})\underline{\theta}(1-\underline{\theta})[h'(u(w)) - h'(u(w-\ell))] > \int_{\underline{\theta}}^{\overline{\theta}} a(s)f(s)ds. \tag{36}$$

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 (36) 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

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

which is equivalent to

$$f(\underline{\theta}) > \frac{(1-\underline{\theta})h'(u(w)) + \underline{\theta}h'(u(w-\ell))}{\underline{\theta}(1-\underline{\theta})[h'(u(w)) - h'(u(w-\ell))]}. \tag{37}$$

Thus, type $\underline{\theta}$ is not excluded from the optimal menu if $f(\underline{\theta})$ is larger than the right side of (37). But Theorem 3 (*i*) or (*ii*) imply that $\theta(1-\theta)f(\theta)/(1-F(\theta))$ is increasing in θ, showing that no $\theta \geq \underline{\theta}$ will be excluded from the optimal menu of contracts.

Without imposing $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$, a stronger sufficient condition holds, with the numerator on the right side of (37) replaced by $h'(u(w))$. ■

A.6 Proof of Proposition 3 (Exclusion)

Type $\tilde{\theta}$ is excluded from the optimal menu of contracts if $\Delta(\tilde{\theta}) = \Delta_0$. From equation (16), we must show that the marginal benefit of providing insurance to $\tilde{\theta}$ starting from no insurance is less than the marginal cost of doing so. Formally,

$$f(\tilde{\theta})\tilde{\theta}(1-\tilde{\theta})[h'(u(w))-h'(u(w-\ell))] < \int_{\tilde{\theta}}^{\overline{\theta}} a(s)f(s)ds. \tag{38}$$

Assume first that $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$. In this case, $\int_{\tilde{\theta}}^{\overline{\theta}} a(s)f(s)ds > a(\overline{\theta})(1-F(\tilde{\theta}))$, and (38) holds if $f(\tilde{\theta})\tilde{\theta}(1-\tilde{\theta})[h'(u(w))-h'(u(w-\ell))] < a(\overline{\theta})(1-F(\tilde{\theta}))$. Since $a(\overline{\theta}) = h'(U(\overline{\theta})) \geq h'((1-\overline{\theta})u(w)+\overline{\theta}u(w-\ell))$, type $\tilde{\theta}$ is excluded from the optimal menu if

$$\frac{f(\tilde{\theta})}{(1-F(\tilde{\theta}))} < \frac{h'((1-\overline{\theta})u(w)+\overline{\theta}u(w-\ell))}{\tilde{\theta}(1-\tilde{\theta})[h'(u(w))-h'(u(w-\ell))]}. \tag{39}$$

But Theorem 3 (i) or (ii) implies that $\theta(1-\theta)f(\theta)/(1-F(\theta))$ is increasing in θ, which shows that any $\theta \leq \tilde{\theta}$ will be excluded from the optimal menu of contracts as well.

Without imposing $-u'''(\cdot)/u''(\cdot) < -3u''(\cdot)/u'(\cdot)$, a similar, but stronger, sufficient condition for exclusion holds, with the numerator of (39) replaced by $h'(u(w-\ell))$. ■

A.7 Proof of Lemma 3 (Curvature)

Let $(U(\theta), \Delta(\theta))$ solve the optimal control problem with $\dot{\Delta}(\cdot) < 0$ everywhere. Since $u(\theta) = U(\theta)+\theta\Delta(\theta)$ for all θ, we can use (5)-(6) to recover the optimal menu $(x(\theta), t(\theta))_{\theta\in\Theta}$.

Recall that $\dot{U}(\theta) = -\Delta(\theta)$, so $\dot{u}(\theta) = \theta\dot{\Delta}(\theta)$. Differentiate (5)-(6) to get

$$\dot{t}(\theta) = -\frac{\theta\dot{\Delta}(\theta)}{u'(w-t(\theta))} \tag{40}$$

$$\dot{x}(\theta) = -\frac{\dot{\Delta}(\theta)[(1-\theta)u'(w-t(\theta))+\theta u'(w-\ell+x(\theta)-t(\theta))]}{u'(w-\ell+x(\theta)-t(\theta))u'(w-t(\theta))}, \tag{41}$$

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

Since (by assumption) sorting is complete, we have $\dot{x}(\cdot) > 0$, 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, (40)-(41) and $\theta = z(x)$ give

$$\dot{T}(x) = \dot{t}(z(x))\dot{z}(x) = \frac{\dot{t}(z(x))}{\dot{x}(z(x))} = \frac{\theta u'(w-\ell+x(\theta)-t(\theta))}{(1-\theta)u'(w-t(\theta))+\theta u'(w-\ell+x(\theta)-t(\theta))}. \tag{42}$$

Differentiate $\dot{T}(\cdot)$ and use $\theta = z(x)$ to find, after some algebra,

$$\ddot{T}(x) = \frac{1}{\theta u'_\ell + (1-\theta)u'_n}\left\{\frac{u'_n u'_\ell}{\dot{x}(\theta)} + \theta(1-\theta)\left[u''_n u'_\ell \frac{\dot{t}(\theta)}{\dot{x}(\theta)} + u''_\ell u'_n\left(1 - \frac{\dot{t}(\theta)}{\dot{x}(\theta)}\right)\right]\right\}. \quad (43)$$

Insert (40)-(41) into (43) and manipulate the resulting expression to reveal that $\ddot{T}(x) < 0$ if and only if $\dot{\Delta}(\theta) < \left(\theta(1-\theta)\left[\theta\frac{u''_n}{u'^2_n} + (1-\theta)\frac{u''_\ell}{u'^2_\ell}\right]\right)^{-1}$. Use $h' = 1/u'$, $h'' = u''/u'^3$, and (13)-(14) to rewrite (17) as

$$\dot{\Delta}(\theta) = \frac{(\frac{1}{u'_n} - \frac{1}{u'_\ell})\Omega + E[\frac{1}{u'}]}{\theta(1-\theta)\left[\theta\frac{u''_n}{u'^3_n} + (1-\theta)\frac{u''_\ell}{u'^3_\ell}\right]}, \quad (44)$$

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 $\ddot{T}(x) < 0$ if and only if

$$\theta(1-\theta)\frac{f'}{f} + 1 - 2\theta > \frac{\left[\theta\frac{u''_n}{u'^3_n} + (1-\theta)\frac{u''_\ell}{u'^3_\ell}\right] - E[\frac{1}{u'}]\left[\theta\frac{u''_n}{u'^2_n} + (1-\theta)\frac{u''_\ell}{u'^2_\ell}\right]}{\left[\theta\frac{u''_n}{u'^2_n} + (1-\theta)\frac{u''_\ell}{u'^2_\ell}\right](\frac{1}{u'_n} - \frac{1}{u'_\ell})}. \quad (45)$$

The numerator on the right side of (45) simplifies to $\left(\frac{1}{u'_n} - \frac{1}{u'_\ell}\right)\left[\theta^2\frac{u''_n}{u'^2_n} - (1-\theta)^2\frac{u''_\ell}{u'^2_\ell}\right]$, and the entire right side simplifies to $\theta - 1 + c(\theta)$. Rearrange to get the result. ■

References

Arnott, R., and J. Stiglitz (1988): "Randomization with Asymmetric Information," *Rand Journal of Economics*, 19, 344–362.

Barlow, R., A. Marshall, and F. Proschan (1963): "Properties of Probability Distributions with Monotone Hazard Rate," *The Annals of Mathematical Statistics*, 34, 375–389.

Biais, B., D. Martimort, and J.-C. Rochet (2000): "Competing Mechanisms in a Common Value Environment," *Econometrica*, 68, 799–37.

Cawley, J., and T. Philipson (1999): "An Empirical Examination of Information Barriers to Trade in Insurance," *American Economic Review*, 89, 827–846.

CHIAPPORI, P.-A., AND A. BERNARDO (2003): "Bertrand and Walras Equilibria under Moral Hazard," *Journal of Political Economy*, 111, 785–817.

CHIAPPORI, P.-A., B. JULLIEN, B. SALANIÉ, AND F. SALANIÉ (2006): "Asymmetric Information in Insurance," *RAND Journal of Economics*, 37, 783–798.

COHEN, A., AND L. EINAV (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97, 745–788.

FUDENBERG, D., AND J. TIROLE (1990): "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, 58, 1279–1319.

HELLWIG, M. (2006): "Incentive Provision with Hidden Characteristics: A Unified Approach," SSRN WP.

JEWITT, I. (1987): "Risk Aversion and the Choice between Risky Prospects: The Preservation of Comparative Statics," *Review of Economic Studies*, 54, 73–85.

JULLIEN, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93, 1–47.

MASKIN, E., AND J. RILEY (1984): "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15, 171–196.

MATTHEWS, S., AND J. MOORE (1987): "Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening," *Econometrica*, 55, 441–67.

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–17.

NÖLDEKE, G., AND L. SAMUELSON (2007): "Optimal bunching without optimal control," *Journal of Economic Theory*, 134, 405–20.

Ormiston, M., and E. E. Schlee (2001): "Mean-Variance Preferences and Investor Behavior," *Economic Journal*, 111, 849–861.

Prescott, E., and R. Townsend (1984): "Pareto Optima and Competitive Equilibrium with Adverse Selection and Moral Hazard," *Econometrica*, 52, 21–46.

Rockafellar, T. (1970): *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 629–650.

Shannon, C. (1995): "Weak and Strong Monotone Comparative Statics," *Economic Theory*, 5, 209–227.

Spence, M. (1977): "Nonlinear Prices and Welfare," *Journal of Public Economics*, 8, 1–18.

Stiglitz, J. (1977): "Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market," *Review of Economic Studies*, 44, 407–430.

Thiele, H., and A. Wambach (1999): "Wealth Effects in the Principal Agent Model," *Journal of Economic Theory*, 89, 247–260.

Wilson, R. (1993): *Nonlinear Pricing*. Oxford University Press, New York.