Health Insurance without Single Crossing

Boone, J.; Schottmuller, C.

Publication date:
2011

Discussion paper

HEALTH INSURANCE WITHOUT SINGLE CROSSING: WHY HEALTHY PEOPLE HAVE HIGH COVERAGE

By
Jan Boone, Christoph Schottmüller

July 7, 2011

TILEC Discussion Paper No. 2011-037

CentER Discussion Paper No. 2011-095

ISSN 0924-7815
http://ssrn.com/abstract=1914473

Tilburg University

Understanding Society

Health Insurance without Single Crossing: why healthy people have high coverage*

Jan Boone and Christoph Schottmüller†

July 7, 2011

Abstract

Standard insurance models predict that people with high (health) risks have high insurance coverage. It is empirically documented that people with high income have lower health risks and are better insured. We show that income differences between risk types lead to a violation of single crossing in the standard insurance model. If insurers have some market power, this can explain the empirically observed outcome. This observation has also policy implications: While risk adjustment is traditionally viewed as an intervention which increases efficiency and raises the utility of low health agents, we show that with a violation of single crossing a trade off between efficiency and solidarity emerges.

Keywords: health insurance, single crossing, risk adjustment

JEL classification: D82, I11

*We gratefully acknowledge financial support from the Dutch National Science Foundation (VICI 453.07.003)

†Boone: Department of Economics, University of Tilburg, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands; Tilec, CentER and CEPR, email: j.boone@uvt.nl. Schottmüller: Department of Economics, Tilec and CentER, Tilburg University, email: christophschottmueller@googlemail.com.

1. Introduction

A well documented problem in health insurance markets with voluntary insurance (like the US) is that people either have no insurance at all or are underinsured.[1] Standard insurance models (inspired by the seminal work of Rothschild and Stiglitz (1976) (RS) and Stiglitz (1977)) predict that *healthy* people have less than perfect insurance or –in the extreme– no insurance at all. However, both popular accounts like Cohn (2007) and academic work like Schoen, Collins, Kriss, and Doty (2008) show that people with low health status are overrepresented in the group of uninsured and underinsured.[2] We develop a model to explain why sick people end up with little or no insurance. We do this by adding two empirical observations (discussed below) to the RS model: (i) richer people tend to be healthier and (ii) health is a normal good. Technically speaking, introducing the latter two effects can lead to a violation of single crossing in the model.

Another indication that the standard RS framework (with single crossing) does not capture reality well is the following. The empirical literature that is based on RS does not unambiguously show that asymmetric information plays a role in health insurance markets. One would expect people to be better informed about their health risks than their insurers –think for example of preconditions, medical history of parents and other family members or life style. However, some papers, like for example Cardon and Hendel (2001) or Dowd, Feldman, Cassou, and Finch (1991), do not find evidence of asymmetric information while others do, e.g. Bajari, Hong, and Khwaja (2006) or Munkin and Trivedi (2010). The test for asymmetric information employed in these papers is the so called "positive correlation test," i.e. testing whether riskier types buy more coverage.

We show that the RS model with a violation of single crossing is capable of explaining why healthy people have better insurance (in equilibrium) than people with a low health status. In particular the positive correlation property no longer holds if single crossing is violated. Consequently, testing for this positive correlation can no longer be viewed as a test for asymmetric information. As mentioned, we use two well documented stylized facts to motivate this violation of single crossing in the market for health insurance.

First, richer people face lower health risks, see for example Frijters, Haisken-DeNew, and Shields (2005), Gravelle and Sutton (2009) or Munkin and Trivedi (2010). Potential explanations for this

[1] In empirical studies, underinsurance is defined using indicators of financial risk. To illustrate, one definition of underinsurance used by Schoen, Collins, Kriss, and Doty (2008) is "out-of-pocket medical expenses for care amounted to 10 percent of income or more". In our theoretical model, underinsurance refers to less than socially optimal/efficient insurance.

[2] In the words of Schoen, Collins, Kriss, and Doty (2008, pp. w303): "underinsurance rates were higher among adults with health problems than among healthier adults".

correlation between income and health include the following. High income people are better educated and hence know the importance of healthy food, exercise etc. Healthy food options tend to be more expensive and therefore better affordable to high income people. Or (with causality running in the other direction) healthy people are more productive and therefore earn higher incomes. According to standard insurance models, this would imply that rich people buy less generous health insurance (compared to poor people). However, richer people have more generous health insurance, for example, in the form of lower deductibles or higher coverage. Evidence for this can be found in Munkin and Trivedi (2010), Finkelstein and McGarry (2006), Kuttner (1999) or DeNavas-Walt, Proctor, and Smith (2008) where the extreme form of underinsurance is emphasized: Low income citizens are more likely to have no health insurance at all.

The second stylized fact is that health is a normal good. The effect of income on treatment choice is well documented in the medical literature. The main emphasis in this literature is that patients with low income cannot afford treatment even if they have a prescription by their doctor. Piette, Heisler, and Wagner (2004b, p. 384) for instance report that from a sample of chronically ill diabetes patients "A total of 19% of respondents reported cutting back on medication use in the prior year due to cost [...]. Moreover, 28% reported forgoing food or other essentials to pay medication costs." By extrapolating from their sample to the US population Piette, Heisler, and Wagner (2004a, p. 1786) conclude that "2.9 million of the 14.1 million American adults with asthma (20%) may be cutting back on their asthma medication because of cost pressures." They also document for a number of chronical conditions that people from low income groups are much more likely to report foregoing prescribed treatment due to costs.[3] Further examples can, for instance, be found in Goldman, Joyce, and Zheng (2007).

For the effect of insurance coverage on treatment choice, Schoen, Collins, Kriss, and Doty (2008, pp. w305) report that "[b]ased on a composite access indicator that included going without at least one of four needed medical care services, more than half of the underinsured and two-thirds of the uninsured reported cost-related access problems". A similar picture emerges in the international comparison by Schoen, Osborn, Squires, Doty, Pierson, and Applebaum (2010).[4] Hence, low income people with high copayments will tend to forego treatment or choose cheaper treatment options.

Single crossing means that people with higher health risks have a higher willingness to pay for

[3]For most chronic diseases people with income less than \$ 20000 are roughly 2 (5) times more likely to forego prescribed treatment due to costs than people with an income between \$ 20000 and \$ 40000 (more than \$ 60000); see table 3 in Piette, Heisler, and Wagner (2004a) for details.

[4]Li and Trivedi (2010) also show that marginal utility of health insurance coverage is influenced by both income and risk factors.

marginally increasing coverage, e.g. reducing copayments. If this property holds for all possible coverage levels, a given indifference curve of a high risk type can cross a given indifference curve of a low risk type at most once. To see why the stylized facts above can lead to a violation of single crossing, consider the following. At full coverage (indemnity insurance that pays for all medical costs) high risk (low health) types will tend to spend more on treatments than low risk types. Hence a small reduction in coverage, leads to a bigger loss in utility for high risk types. Now consider health insurance with low coverage where the insured faces substantial copayments. Because health is a normal good, it is possible that the rich-healthy type spends more on treatment than the low income, low health type. In that case, a small change in coverage has a bigger effect on the utility of the healthy type than of the low health agent. The healthy type will therefore have a higher willingness to pay for a marginal increase in coverage than the low health type. This violates single crossing.

We show that in insurance models without single crossing higher health risks are not necessarily associated with more coverage while this prediction is inevitable with single crossing. Not only leads this to predictions that are closer to empirical observations (as documented above), it also has clear policy implications. We illustrate this with risk adjustment.

Risk adjustment is used by the sponsor (government or employer) of a health insurance scheme to reduce expected cost differences between types. Based on observable characteristics (like age, gender etc.)[5] the health insurer is subsidized (taxed/subsidized less) for customers with high (low) expected costs. This is used in a number of countries like Australia, Canada, Germany, Ireland (until 2008), The Netherlands, South Africa and the USA (see Ellis (2007) and Armstrong, Paolucci, McLeod, and Van de Ven (2010)) in both mandatory state insurance and voluntary private insurance. The two main goals of risk adjustment are efficiency and fairness (or solidarity); see Van de Ven and Ellis (2000) for an overview of the literature on risk adjustment and the way it is used in practice. Indeed, in the standard RS model, these two goals go in hand in hand. Starting from zero risk adjustment, increasing the subsidy to the high risk consumer increases both efficiency and the utility of people with low health status. Hence the sponsor of the health insurance scheme does not need to choose between these two objectives. As we show below, this is no longer the case with a violation of single crossing.

In particular, when single crossing is violated, risk adjustment will either increase efficiency (by reducing co-payments for high risk types) or increase the utility of high risk types; but not both. Hence when designing the risk adjustment scheme, the sponsor needs to be explicit whether the goal is efficiency or solidarity. In other words, with a violation of single crossing there is a trade off between efficiency and equity.

[5] Clearly, observable characteristics are not a perfect predictor of risk type. Glazer and McGuire (2000) show how the imperfection of the signals should be taken into account when designing the risk adjustment scheme.

The literature on violations of single crossing is relatively scarce. There are three papers analyzing perfectly competitive insurance markets with 2×2 types: People differ in two dimensions and both dimensions can either take a high or a low value. In Smart (2000) the two dimensions are risk and risk aversion. Netzer and Scheuer (2010) model an additional labor supply decision and the two dimensions are productivity and risk. Wambach (2000) models both wealth and risk. All papers have a pooling result, i.e. if single crossing does not hold two of the four types can be pooled. Our paper contributes by deviating from the perfect competition assumption. We show that under imperfect competition types cannot only be pooled but high risk types might get even less coverage in equilibrium than low risk types.

Jullien, Salanie, and Salanie (2007) take a different approach to answer the question why high risk types might have lower coverage in general insurance markets. They use a model where types differ in risk aversion and single crossing is satisfied. Hence, types with higher risk aversion will have more coverage in equilibrium. At the same time more risk averse agents might engage more in preventive behavior. If types are still separated in equilibrium and risk aversion differences remain the driving force, high risk aversion types will exhibit less risk (due to prevention) and higher coverage. Similar lines of reasoning can be found in Hemenway (1990) and De Meza and Webb (2001).

Since risk in the health sector is exogenously different for different persons (e.g. due to genetics), we follow RS and take a different starting point than Jullien, Salanie, and Salanie (2007). We assume risk differences instead of risk aversion differences. The result that high risk people have low coverage is in our paper not the result of low risk aversion. The driving force is the violation of single crossing caused by empirically documented income differences between high risks and low risks.

Finally, our paper is related to the industrial economics literature on non-linear pricing in oligopoly (see, for instance, Armstrong and Vickers (2001), Stole (1995) and Stole (2007)). This literature focuses on the welfare effects of non-linear pricing in imperfectly competitive markets. Our contribution to this literature is to do comparative static analysis with respect to risk adjustment. In other words, our paper has positive (testable) implications for non-linear pricing in oligopoly. Moreover, whereas the insurance papers mentioned above focus on either perfect competition or monopoly, we actually analyze all three cases: perfect competition, monopoly and oligopoly. As one would expect, the assumptions on the competitive situation matter for the results.

In the following section, the model is introduced and illustrated with an example. Then equilibrium in monopoly and oligopoly is derived. Thereafter we introduce risk adjustment and show the trade off between efficiency and solidarity in case single crossing is violated. We conclude with the implications of our model for so called advantageous selection. Proofs are relegated to the appendix.

2. Insurance model

This section introduces a general model of health insurance that allows us to consider both the case where single crossing (SC) is satisfied and the case where it is not satisfied (NSC). The section concludes with a model where SC is not satisfied because of income differences between consumer types. Whereas the RS model predicts that people with high expected health care expenditures have generous coverage in their insurance, we allow for the case that these people cannot afford such generous insurance.

Following RS, we consider an agent with utility function $u(q, p, \theta)$ where $q \in [0, 1]$ denotes coverage or generosity of her insurance contract,[6] $p \geq 0$ denotes the price of insurance (insurance premium) and $\theta \in \{\theta^l, \theta^h\}$ with $\theta^h > \theta^l > 0$ denotes the type of consumer.[7] Higher θ denotes a higher risk in the sense of higher expected costs (in case $q^h = q^l = 1$; see below). This could, for instance, be the case due to chronic illness or higher risk due to a genetic precondition. We make the following assumptions on the utility function.

Assumption 1 *The utility function $u(q, p, \theta)$ is continuous and differentiable. It satisfies $u_q > 0, u_p < 0$. We define the indifference curve $p(q, u, \theta)$ as follows:*

$$u(q, p(q, u, \theta), \theta) \equiv u \tag{1}$$

We assume that these indifference curves $p(q, u, \theta)$ are differentiable in q and u with $p_q = -u_q/u_p > 0, p_u = 1/u_p < 0$.
Further, the crossing at $q = 1$ satisfies:

$$p_q(1, u^h, \theta^h) > p_q(1, u^l, \theta^l) \tag{C1}$$

for all $u^l \geq \bar{u}^l = u(0, 0, \theta^l), u^h \geq \bar{u}^h = u(0, 0, \theta^h)$.

In words, utility u is increasing in coverage q and decreasing in the premium p paid for insurance. For given type θ and utility level u, the indifference curve $p(q, u, \theta)$ maps out combinations (q, p) that yield the same utility. Because higher q leads to higher utility, p increases to keep utility constant. Hence, indifference curves are upward sloping in (q, p) space ($p_q > 0$). Increasing u (for given q) requires a lower price. Thus, raising u shifts an indifference curve downwards ($p_u < 0$).

[6] Apart from literal coverage –where $1 - q$ denotes the agent's copayments– q could, for example, be interpreted as $1/(1 + \text{deductible})$. Note that in models without moral hazard both parameters are similar in the sense that high risk types dislike co-payments and deductibles more relative to low risk types.

[7] We follow RS and much of the risk adjustment literature in assuming that there are only two types. For an analysis of a violation of single crossing with a continuum of types θ, see Araujo and Moreira (2010) and Schottmüller (2011).

Type $k \in \{h, l\}$ buys insurance if it leads to a higher utility than her outside option $\bar{u}^k$. This outside option is given by the "empty insurance contract": $q = p = 0$.

At $q = 1$ a marginal reduction in q should be compensated by a bigger decrease in p for θ^h compared to θ^l. This reflects the fact that the θ^h type faces higher expected health care expenditures. At $q = 1$, i.e. at full coverage, other factors like willingness to pay for treatment (which could be different for different types) play no role. In this sense, this assumption "defines" what higher θ means: at $q = 1$, higher θ types face higher expected costs. With the same idea we assume that expected costs for the insurer of a contract with $q = 1$ is higher for the θ^h than for the θ^l type: $c(1, u^h, \theta^h) > c(1, u^l, \theta^l)$ for all $u^h \geq \bar{u}^h, u^l \geq \bar{u}^l$. Intuitively, u should not matter for health care consumption at full coverage and the high risk type will use the insurance more.

To allow for income effects (for instance, in treatment choice; see below) the cost function depends on u. However, we assume two regularity conditions.

Assumption 2 *For each type $k \in \{h, l\}$ and $q \in [0, 1]$ we assume that*

- $c_u(q, u^k, \theta^k) \geq 0$ *for* $u^k \geq \bar{u}^k$,
- $c(1, u^k, \theta^k) = c(1, \tilde{u}^k, \theta^k)$, *for* $u^k, \tilde{u}^k \geq \bar{u}^k$.

In words, as the income of the agent increases (which *ceteris paribus* leads to higher utility), the agent has more money to spend on treatment. As the insurer pays a fraction $q \geq 0$ of these treatments, this leads to (weakly) higher costs for the insurer. Second, costs at full coverage ($q = 1$) do not vary in utility. Intuitively, if $q = 1$ treatments are for free for the agent and there is no reason to forgo treatments, irrespective of the level of $u^k \geq \bar{u}^k$.[8]

Because of (C1), the single crossing condition reads[9]

$$p_q(q, u^h, \theta^h) > p_q(q, u^l, \theta^l) > 0 \text{ for all } q \in [0, 1] \tag{SC}$$

and $u^h \geq \bar{u}^h$, $u^l \geq \bar{u}^l$ such that $p(q, u^h, \theta^h) = p(q, u^l, \theta^l)$. The intuition is the following. Suppose an indifference curve of type θ^h intersects with an indifference curve of type θ^l in some point (p, q). Then (SC) implies that the slope of the θ^h indifference curve will be higher. It follows that these two indifference curves can intersect only once.

We consider both the case where (SC) is satisfied and the case where it is violated (denoted by NSC). In both the SC and NSC cases, we maintain the assumption that $q = 1$ is the efficient insurance level (EI) for each type $\theta \in \{\theta^l, \theta^h\}$.

[8] By assumption 3 full coverage is socially desirable. Hence we do not consider the case where insurance leads to inefficiency by inducing over-consumption of treatments.

[9] This is also called sorting, constant sign or Spence-Mirrlees condition (Fudenberg and Tirole (1991, pp. 259)).

Assumption 3 *For a given utility level u^k, welfare (and therefore profits) are maximized at full coverage, i.e.*

$$\max_{q\in[0,1]} p(q,u^k,\theta^k) - c(q,u^k,\theta^k) \tag{EI}$$

is maximized by $q = 1$ for each $k \in \{h,l\}$ and $u^k \geq \bar{u}^k$.

This basically means that the insurance motive, i.e. transferring risk from a risk averse agent to a risk neutral insurer, is not overruled by other considerations. To illustrate, we do not assume that the low income agent's preference for health/treatment is so low that foregoing insurance would be socially optimal. Put differently, we assume that full insurance is socially desirable. Underinsurance–with no insurance as extreme case–results therefore not from first best but from informational distortions and price discrimination motives.

Our motivation for making this assumption is twofold. First, this assumption simply normalizes the socially efficient insurance level in the same way as in RS. Hence, we only deviate from the RS set up by allowing for both SC and NSC. Second, we want to argue that under realistic assumptions, θ^h types have less than full insurance. If the optimal insurance level is actually below one, than this result would follow rather trivially.

The literature on insurance models considers mostly perfect competition.[10] We show that with the assumptions made so far, perfect competition implies $q^h = 1$ (even if (SC) is not satisfied). Hence, market power on the insurance side is needed to get $q^h < 1$. Following the RS definition of the perfect competition equilibrium, we require that (i) each offered contract makes nonnegative profits and (ii) given the equilibrium contracts there is no other contract yielding positive profits.

Proposition 1 *If an equilibrium exists under perfect competition then $q^h = 1$.*

As is well known, existence of equilibrium in the RS framework is not guaranteed. Equilibrium does not exist if the only possible (separating) equilibrium is broken by a pooling contract. If the fraction of θ^h type agents in the population is high enough, then such a deviation to a pooling contract is not profitable and an equilibrium exists. If an equilibrium exists, it has $q^h = 1$.

The proposition shows that even with violations of single crossing, high risk types will get (weakly) higher coverage than low risk types. Hence we need to deviate from perfect competition to get $q^h < q^l$. This proposition is in some sense reminiscent of Wambach (2000), Smart (2000) and Netzer and Scheuer (2010): these papers analyze perfectly competitive insurance markets and a reverse order, i.e. riskier types have less coverage, is impossible in these papers.

[10] See Jack (2006) and Olivella and Vera-Hernández (2007) for exceptions using a Hotelling model to formalize market power on the insurer side of the market. These papers assume that (SC) is satisfied and hence find efficient insurance for the θ^h type.

Corollary 1 *Whenever $q^h < q^l$ is observed, insurers have market power.*

Although previous models of insurance markets assume perfect competition, recent research for the US (see Dafny (2010)) shows that health insurers do have market power. More generally, in most countries where health insurance is provided by private companies, these firms tend to be big (due to economies of scale in risk diversification). Hence one would expect them to have some market power.

In order to analyse the contracts offered by insurers with market power, we explicitly introduce the incentive compatibility (IC) constraints for each type

$$p(q^l, u^l, \theta^l) \geq p(q^l, u^h, \theta^h) \tag{IC_h}$$

$$p(q^h, u^h, \theta^h) \geq p(q^h, u^l, \theta^l) \tag{IC_l}$$

The first constraint implies that the contract intended for θ^h (i.e. $(q^h, p(q^h, u^h, \theta^h)))$ lies on a (weakly) lower indifference curve for θ^h than the contract that is meant for the θ^l type $(q^l, p(q^l, u^l, \theta^l))$. That is, the inequality implies $u(q^h, p^h, \theta^h) \geq u(q^l, p^l, \theta^h)$ where $p^i = p(q^i, u^i, \theta^i)$ with $i \in \{h, l\}$. Similarly, the second inequality implies that $u(q^l, p^l, \theta^l) \geq u(q^h, p^h, \theta^l)$.

Irrespective of the mode of competition and whether (SC) holds, we have the following result that we use below.

Lemma 1 *At least one type has full coverage. If the types are separated under the optimal contract scheme $(q^l, p^l), (q^h, p^h)$ with $q^l \neq q^h$, then at most one incentive constraint binds.*

We conclude this section with a model where SC is violated due to differences in income between types. The idea of the model is that for $q < 1$ people have to finance part of the costs of treatment out of their own pocket and low income agents may decide to choose cheaper treatment or forgo treatment altogether. This effect is documented in the medical literature, see for example Piette, Heisler, and Wagner (2004b), Piette, Heisler, and Wagner (2004a) or Goldman, Joyce, and Zheng (2007).

In particular, we assume that a type θ consumer faces health shock $s \in [0, 1]$ with distribution (density) function $F(s|\theta)(f(s|\theta))$. We take $s = 1$ as the state in which the agent is healthy and needs no treatment. Lower health states s correspond to worse health. The assumption that the θ^h type has worse health than the θ^l type can now be stated as $F(s|\theta^h) > F(s|\theta^l)$ for each $s \in \langle 0, 1 \rangle$. In words, low s states are more likely for the θ^h then for the θ^l type.

Once an agent receives a health shock $s < 1$ she can increase her health by treatment $h \in H(s)$ to health level $s + h$, where $H(s)$ denotes the set of possible treatments in state s. We assume that the set $H(s)$ is compact and $s + h \leq 1$ for each $h \in H(s)$ and each $s \in [0, 1]$. That is, treatment cannot lead to higher health states than not falling ill. If $H(s)$ is a singleton, the consumer has no treatment choice. If the set $H(s)$ has more than one element, low income consumers with partial insurance, i.e. $q < 1$, may decide to choose cheaper treatment than if they have full insurance, i.e. $q = 1$. We define $\bar{h}(s) = \max\{h \in H(s)\}$ as the best possible treatment and assume that $\bar{h}(s)$ is non-increasing in s. This means that a less afflicted agent (high $s < 1$) cannot increase his health by treatment more than an agent who is more seriously ill (low s). If $0 \in H(s)$, an agent can forgo treatment altogether.

Let $w(\theta)$ denote the wealth (or income) of a type θ agent. Then we write

$$\begin{aligned} u(q,p,\theta) &= \int_0^1 \{v(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta))\}dF(s|\theta) \\ &\text{where } h(s,q,\theta) \text{ is defined as:} \\ h(s,q,\theta) &= \arg\max_{h \in H(s)} v(w(\theta) - p - (1-q)h, s + h) \end{aligned} \tag{2}$$

where $v(y,h)$ is the utility function of an agent which depends on consumption of other goods (y) and health (h). We assume that $v(y,h)$ satisfies $v_y, v_h > 0, v_{yy}, v_{hh} < 0$ and that health is a normal good: $v_{hy} \geq 0$. That is, utility increases in both health and consumption of other goods at a decreasing rate. As income increases, people's preference for health increases as well. Further, we assume that income and health status are negatively correlated: $w(\theta^h) \leq w(\theta^l)$. This negative correlation is empirically documented, for example, in Frijters, Haisken-DeNew, and Shields (2005), Gravelle and Sutton (2009) or Munkin and Trivedi (2010).

Using this notation we can write

$$c(q,u,\theta) = q \int_0^1 h(s,q,\theta)dF(s|\theta) \tag{3}$$

The first order condition for an interior solution $h(s,q,\theta) \in H(s)$ can be written as

$$(1-q)v_y(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta)) = v_h(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta)) \tag{4}$$

To see the implications of this model for single crossing, consider the slope of the indifference curves in (q,p)-space:

$$p_q(q,u,\theta) = -\frac{u_q}{u_p} = \frac{\int_0^1 v_y(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta))h(s,q,\theta)dF(s|\theta)}{\int_0^1 v_y(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta))dF(s|\theta)} \tag{5}$$

In words, the slope p_q equals the weighted average of treatment $h(s,q,\theta)$ over the states s with weight

$$\frac{v_y(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta))f(s|\theta)}{\int_0^1 v_y(w(\theta) - p - (1-q)h(s,q,\theta), s + h(s,q,\theta))dF(s|\theta)} \tag{6}$$

on state s (where the weights integrate to 1).

To illustrate (C1) assume that $s+\bar{h}(s)=1$ (treatment makes a patient healthy again),[11] then it is routine to verify that

$$p_q(1,u,\theta)=\frac{\int_0^1 v_y(w(\theta)-p,1)\bar{h}(s)dF(s|\theta)}{\int_0^1 v_y(w(\theta)-p,1)dF(s|\theta)}=\int_0^1(1-s)dF(s|\theta)$$

where the last equality follows from the fact that $v_y(w(\theta)-p,1)$ is constant in s. Note that we use here that $h(s,1,\theta)=\bar{h}(s)$ for both types. If treatment is free ($q=1$) each agent uses the highest treatment ($\bar{h}(s)$). The stochastic dominance assumption implies that θ^h puts more weight on low s states (where $\bar{h}(s)=1-s$ is high) compared to θ^l. Hence under these assumptions, condition (C1) is satisfied.

(SC) is satisfied if there are no wealth differences between types, i.e. $w(\theta^h)=w(\theta^l)$, and $H(s)$ satisfies some regularity condition. The idea is that without wealth differences, (4) yields for both types the same optimal treatment. Put differently, $h(s,q,\theta)$ is independent of θ. If patients choose more treatment in worse health states, single crossing will be satisfied: due to stochastic dominance, θ^h types have higher weight (6) on low s states with high $h(s)$. Hence p_q in (5) is higher for θ^h than for θ^l types for all $q\in[0,1]$. Treatment $h(s,q,\theta)$ is indeed non-increasing in s if $H(s)$ is well behaved: $H(s)$ is convex for each s and non-increasing in s.[12] It then follows from equation (4) –using the implicit function theorem– that

$$(-(1-q)^2v_{yy}+2(1-q)v_{yh}-v_{hh})\frac{dh}{ds}=v_{hh}-(1-q)v_{yh} \tag{7}$$

Hence from the assumptions on v it follows that $h(s,q,\theta)$ is non-increasing in s. As $H(s)$ is non-increasing, this also holds true for boundary solutions where the implicit function theorem cannot be used.

However, if $w(\theta^h)<w(\theta^l)$ then $q<1$ can imply that $h(s,q,\theta^h)<h(s,q,\theta^l)$. This follows from equation (4), since

$$\frac{dh}{dw}=\frac{-(1-q)v_{yy}+v_{hy}}{-(1-q)^2v_{yy}+2(1-q)v_{yh}-v_{hh}}>0 \tag{8}$$

Hence, if $h(s,q,\theta^l)\in H(s)$ is an interior maximum, the θ^h type tends to choose lower treatment h. In words, since a fraction $1-q$ of the treatment cost has to be paid by the insured, a low income θ^h patient may choose cheaper treatment than the richer θ^l type (as health is a normal good). Since

[11] Alternatively, we can assume that $\bar{h}'(s)\in\langle-1,0\rangle$ such that $s+\bar{h}(s)$ is increasing in s. In words, if an agent falls ill, treatment does not bring back full health. Then a sufficient condition for (C1) is $v_{yyh}\geq 0$: Suppose it were the case that $\bar{h}'(s)=0$, then $p_q(1,\cdot)$ would be the same for both types. $\bar{h}'(s)<0$ will now put more weight on low states (as $\bar{h}(s_2)\leq\bar{h}(s_1)$ for $s_1\leq s_2$). $v_{yyh}\geq 0$ guarantees that v_y is increasing less in s for the high type. Hence, putting more weight on low states where v_y is low affects the θ^l type less than the θ^h type. Consequently, (C1) is satisfied.

[12] We say that the set $H(s)$ is non-increasing in s if for each s_1,s_2 with $s_1\leq s_2$ we have that for each $h\in H(s_2)$ there exists $h'\in H(s_1)$ such that $h'\geq h$. As a special case this includes the possibility that $H(s)=h(s)$ is a singleton, with $h'(s)<0$.

he does not utilize the insurance as much as the (rich) low risk type, type θ^h has a lower marginal willingness to pay for insurance coverage (for q close to zero). However, for high levels of coverage, i.e. q close to 1, wealth differences matter less in the treatment choice because the patient does not have to pay (much) for the treatment. Consequently, although (C1) is satisfied with $w(\theta^h) < w(\theta^l)$, (SC) can be violated.

Hence, this model –where agents differ in income and treatment choice $h \in H(s)$ is endogenous– can generate the violation of (SC) mentioned above. In the next section, we give a numerical example where (SC) is indeed violated.

3. Example

As an example of an utility function that satisfies the assumptions (C1) and (EI) above and violates (SC), consider the following mean-variance utility set up.[13]

There are two states of the world: An agent either falls ill or stays healthy. The probability of falling ill is denoted by F^h ($F^l < F^h$) for type θ^h (θ^l). In the numerical example, we choose $F^h = 0.07 > 0.05 = F^l$. Once an agent falls ill, the set of possible treatments is denoted by $H = \{\underline{h}, \bar{h}\}$. The utility of an agent of type $i = h, l$ with treatment choice $h \in \{\underline{h}, \bar{h}\}$ is written as:

$$\begin{aligned} u(q,p,\theta^i) &= F^i(v(h,\theta^i) - (1-q)h) + (1-F^i)v(1,\theta^i) - p \\ &\quad -\tfrac{1}{2}r^iF^i(1-F^i)(v(1,\theta^i) - v(h,\theta^i) + (1-q)h)^2 \end{aligned} \tag{9}$$

where $v(h, \theta^i)$ denotes the utility for type $i = h, l$ of having health h and $r^i > 0$ denotes the degree of risk aversion. Hence an agent's utility is given by the expected utility minus $\frac{1}{2}r^i$ times the variance in the agent's utility. This is a simple way to capture that the agent is risk averse.[14]

Along an indifference curve where u is fixed, we find the following slope:

$$\frac{dp}{dq} = F^ih(q,\theta^i) + r^iF^i(1-F^i)(v(1,\theta^i) - v(h(q,\theta^i),\theta^i) + (1-q)h(q,\theta^i))h(q,\theta^i) \tag{10}$$

[13] For the Python code used to generate this example, see:
http://sites.google.com/site/janboonehomepage/home/webappendices.

[14] When an agent of type i buys a product at price p that gives utility v, there are two ways to capture the marginal utility of income for agent i. First, overall utility can be written as $v - \alpha^i p$ where v is the same for each type i and α^i can differ. Low income types are then modelled to have high α^i; high marginal utility of income. Alternatively, one can write $v^i - \alpha p$ where α is the same for all types. Then low income types have low v^i. We have chosen the latter formalization with $\alpha = 1$. The assumption that treatment is a normal good is then implemented by assuming that

$$v(\bar{h}, \theta^h) - v(\underline{h}, \theta^h) < v(\bar{h}, \theta^l) - v(\underline{h}, \theta^l).$$

where $h(q,\theta^i)$ is the solution for h solving

$$\max_{h\in\{\underline{h},\bar{h}\}} v(h,\theta^i)-(1-q)h$$

In words, once an agent falls ill, she decides which treatment to choose based on the benefit $v(h,\theta^i)$ and the out-of-pocket expenses $(1-q)h$.

With the parameter values that we consider below, it is the case that

$$r^h F^h(1-F^h)(v(1,\theta^h)-v(\bar{h},\theta^h))\bar{h} = r^l F^l(1-F^l)(v(1,\theta^l)-v(\bar{h},\theta^l))\bar{h} \tag{11}$$

In words, at $q=1$ (where both types choose the highest treatment $\bar{h}$) the variance terms in the slope dp/dq (equation (10)) are equalized. Hence assumption (C1) is satisfied because $F^h\bar{h} > F^l\bar{h}$.

For the numerical example we assume $\bar{h}=0.6, \underline{h}=0.2$ and the associated utilities for the θ^h type equal $v(1,\theta^h)=0.9, v(\bar{h},\theta^h)=0.7, v(\underline{h},\theta^h)=0.45$ and similarly for the θ^l type: $v(1,\theta^l)=1.1, v(\bar{h},\theta^l)=0.9, v(\underline{h},\theta^l)=0.5$. Hence, having high health is more important for the θ^l type compared to the θ^h type. This implies that θ^l type is willing to spend more on treatment than the θ^h type. Since $0.9-0.6\geq 0.5-0.2$ the θ^l type chooses $\bar{h}$ even if $q=0$ (and the inequality is strict for $q>0$). This implies that condition (EI) is satisfied for the θ^l type as q does not affect treatment choice and higher q leads to more insurance (provided by a risk neutral insurer). The θ^h type chooses $\bar{h}$ if $q=1$ but prefers $\underline{h}$ for low values of q. In particular, for $q=0$ we have $v(\bar{h},\theta^h)-\bar{h} < v(\underline{h},\theta^h)-\underline{h}$. Let $\tilde{q}$ denote the value for q such that the θ^h type is indifferent between treatment $\bar{h}$ and $\underline{h}$:

$$v(\bar{h},\theta^h)-(1-\tilde{q})\bar{h} = v(\underline{h},\theta^h)-(1-\tilde{q})\underline{h} \tag{12}$$

To verify that (EI) is satisfied for the θ^h type, we proceed in two steps. First, consider $q>\tilde{q}$ such that $h(q,\theta^h)=\bar{h}$. Then increasing coverage q reduces the variance in utility for the risk averse θ^h type and hence (EI) is satisfied for $q>\tilde{q}$. Now consider $q<\tilde{q}$ such that $h(q,\theta^h)=\underline{h}$. In order to satisfy (EI), it must be the case that profits (price minus expected costs) when offering full coverage are higher than profits when offering a partial coverage contract yielding the same utility. This can be written as:

$$F^h(v(\bar{h},\theta^h)-\bar{h})+(1-F^h)v(1,\theta^h)-u^h-\tfrac{1}{2}r^hF^h(1-F^h)(v(1,\theta^h)-v(\bar{h},\theta^h)^2 \geq$$
$$F^h(v(\underline{h},\theta^h)-\underline{h})+(1-F^h)v(1,\theta^h)-u^h-\tfrac{1}{2}r^hF^h(1-F^h)(v(1,\theta^h)-v(\underline{h},\theta^h)+(1-q)\underline{h})^2$$

Note that the right hand side of this inequality increases in q and hence is highest at $\tilde{q}$. In our numerical example, we choose r^h such that the inequality holds with equality at $q=\tilde{q}$.[15] This implies that it is satisfied for all $q\leq\tilde{q}$ and hence (EI) is satisfied.

[15] Given this value of r^h, r^l is chosen to satisfy equation (11).

With the parameter values above, it is routine to verify that (SC) is violated. Figure 1 below shows two indifference curves for the θ^l type (in red) and one for the θ^h type (in blue). Indeed, for $q < \tilde{q}$ the indifference curve for the θ^l type is steeper than for the θ^h type. This is due to the fact that the θ^l type buys the expensive treatment $\bar{h}$ while the θ^h type buys $\underline{h}$. The kink in the indifference curve for the θ^h type happens at $\tilde{q}$ where the θ^h type switches from the cheap to the more expensive treatment. Hence small increases in q for $q > \tilde{q}$ are worth more to the θ^h type than small increases in $q < \tilde{q}$. In fact, the figure shows that for $q > \tilde{q}$, the indifference curve for the θ^h type is steeper than the one for the θ^l type. This is the violation in single crossing.

Hence in a simple mean-variance utility framework, it is straightforward and intuitive to generate a violation of (SC).

4. Insurance market monopoly

This section derives a simple result in a monopoly setting within the general reduced form framework of section 2. The motivation for analyzing the monopoly case is proposition 1: An equilibrium with $q^h < q^l$ can only exist if insurance companies have market power. Although the extreme case of a monopoly is not very realistic, it allows to derive a clear cut result and gives some intuition for the more realistic case of an oligopoly which will be analyzed in the following section.

Proposition 2 *The type with the highest willingness to pay for full coverage, i.e. the type θ^k with highest $p(1, \bar{u}^k, \theta^k)$, obtains a full coverage contract in an insurance monopoly. Either his incentive compatibility or his individual rationality constraint is binding (or both). The other type's individual rationality constraint is binding.*

Let θ^k denote the type with the highest willingness to pay for full coverage. It follows from the proposition that θ^k obtains a contract $(q, p) = (1, p^k)$ for some $p^k \leq p(1, \bar{u}^k, \theta^k)$. The monopoly outcome is now pinned down by the choice of p^k. If $p^k = p(1, \bar{u}^k, \theta^k)$, then both individual rationality constraints are binding. In this case, type θ^{-k} might be excluded, i.e. θ^{-k} gets the contract $(0, 0)$. If $p^k = p(1, \bar{u}^{-k}, \theta^{-k})$, both types are pooled. If $p^k \in \langle p(1, \bar{u}^{-k}, \theta^{-k}), p(1, \bar{u}^k, \theta^k)\rangle$, the equilibrium separates the types and θ^{-k} gets an insurance contract with partial coverage. The optimal level of p^k is determined by the share of θ^k types in the population.

A direct implication of proposition 2 is that high risk types will always have full coverage if single crossing is satisfied. To see this, note that the indifference curve corresponding to $\bar{u}^k$ (that is the individual rationality constraint) goes through the origin $(p, q) = (0, 0)$ for both types. With (SC) the indifference curve of the high risk type is steeper and lies therefore above the individual rationality constraint of the low risk type for all coverage levels.

Without single crossing this is no longer the case. Figure 1 shows indeed that in our numerical example the low risk type has a higher willingness to pay for full coverage than the high risk type. Therefore, a monopolist will give full coverage to low risk types in our example. If the types are separated, we find that $q^h < q^l = 1$.

5. Insurance market oligopoly

This section characterizes equilibrium in a duopoly insurance market. Again the general reduced form model introduced in section 2 is used. We illustrate with the example from section 3 that equilibria with $q^h < q^l$ exist if single crossing is violated.

There are three reasons why we choose to consider an oligopoly insurance market. First, proposition 1 shows that with perfect competition it is impossible to have $q^h < 1$. Second, as mentioned above, assuming a monopoly market is not very realistic. We are not aware of a country where health insurance is provided privately by a monopolist. Third, in countries like Germany or the Netherlands, insurance is mandatory and the government runs a risk adjustment scheme. The next section analyzes the effects of risk adjustment on efficiency and solidarity. If there is mandatory insurance, risk adjustment has no effect on the outcome at all in case of monopoly. The reason is that the monopolist serves all types and hence risk adjustment is a lump sum transfer for a monopolist insurer. As we show below, risk adjustment (even if the whole market is served, as with mandatory insurance) does affect the market outcome in case of oligopoly.

The demand share of insurer $j \in \{a, b\}$'s product on market $i \in \{h, l\}$ is written as $D(u^i_j, u^i_{-j}) \in [0, 1]$. This function is the same for both types. That is, agents only differ in expected health care costs and income, not in their preferences over insurers.[16] We make the following natural assumption on demand.

Assumption 4

$$D_1(u^i_j, u^i_{-j}) > 0, D_2(u^i_j, u^i_{-j}) < 0$$

An insurer gains market share if it offers consumers higher utility and loses market share if its opponent offers higher utility. Let $\phi \in [0, 1]$ denote the share of θ^h types in the population.

[16] See Bijlsma, Boone, and Zwart (2011) for an analysis where θ^h and θ^l have different demand elasticities.

When insurer b chooses $u_b^h, q_b^h, u_b^l, q_b^l$, the maximization problem for insurer a can be written as

$$\begin{aligned}\max_{u^h,q^h,u^l,q^l} & \phi D(u^h, u_b^h)(p(q^h, u^h, \theta^h) - c(q^h, u^h, \theta^h)) \\ & + (1-\phi)D(u^l, u_b^l)(p(q^l, u^l, \theta^l) - c(q^l, u^l, \theta^l)) \\ & + \mu_h(p(q^l, u^l, \theta^l) - p(q^l, u^h, \theta^h)) \\ & + \mu_l(p(q^h, u^h, \theta^h) - p(q^h, u^l, \theta^l))\end{aligned} \tag{$P_{u_b^h, u_b^l}$}$$

We focus on a symmetric pure strategy equilibrium of this problem where $u_a^h = u_b^h = u^h$, $u_a^l = u_b^l = u^l$, $q_a^h = q_b^h = q^h$ and $q_a^l = q_b^l = q^l$. That is, (u^h, q^h, u^l, q^l) is a solution to (P_{u^h,u^l}): problem $(P_{u_b^h,u_b^l})$ with $u_b^h = u^h$ and $u_b^l = u^l$. We assume that problem (P_{u^h,u^l}) is concave such that a solution is characterized by the first order conditions for q^h, q^l, u^h, u^l:[17]

$$\phi D(u^h, u^h)\frac{\partial(p(q^h, u^h, \theta^h) - c(q^h, u^h, \theta^h))}{\partial q^h} + \mu_l(\frac{\partial p(q^h, u^h, \theta^h)}{\partial q^h} - \frac{\partial p(q^h, u^l, \theta^l)}{\partial q^h}) = 0 \tag{13}$$

$$(1-\phi)D(u^l, u^l)\frac{\partial(p(q^l, u^l, \theta^l) - c(q^l, u^l, \theta^l))}{\partial q^l} + \mu_h(\frac{\partial p(q^l, u^l, \theta^l)}{\partial q^l} - \frac{\partial p(q^l, u^h, \theta^h)}{\partial q^l}) = 0 \tag{14}$$

$$\begin{aligned}\phi D_1(u^h, u^h)(p(q^h, u^h, \theta^h) - c(q^h, u^h, \theta^h)) + \phi D_1(u^h, u^h)\frac{\partial(p(q^h, u^h, \theta^h) - c(q^h, u^h, \theta^h))}{\partial u^h} \\ -\mu_h\frac{\partial p(q^l, u^h, \theta^h)}{\partial u^h} + \mu_l\frac{\partial p(q^h, u^h, \theta^h)}{\partial u^h} = 0\end{aligned} \tag{15}$$

$$\begin{aligned}(1-\phi)D_1(u^l, u^l)(p(q^l, u^l, \theta^l) - c(q^l, u^l, \theta^l)) + (1-\phi)D(u^l, u^l)\frac{\partial(p(q^l, u^l, \theta^l) - c(q^l, u^l, \theta^l))}{\partial u^l} \\ +\mu_h\frac{\partial p(q^l, u^l, \theta^l)}{\partial u^l} - \mu_l\frac{\partial p(q^l, u^h, \theta^l)}{\partial u^l} = 0\end{aligned} \tag{16}$$

It follows from lemma 1 together with the assumption that $(P_{u_b^h,u_b^l})$ is concave that in any separating equilibrium, exactly one IC constraint is binding. Hence either

$$\begin{cases}\mu_h > 0 & \text{and } p(q^l, u^l, \theta^l) = p(q^l, u^h, \theta^h) \\ \mu_l = 0 & \text{and } p(q^h, u^h, \theta^h) \geq p(q^h, u^l, \theta^l)\end{cases} \tag{17}$$

or

$$\begin{cases}\mu_h = 0 & \text{and } p(q^l, u^l, \theta^l) \geq p(q^l, u^h, \theta^h) \\ \mu_l > 0 & \text{and } p(q^h, u^h, \theta^h) = p(q^h, u^l, \theta^l)\end{cases} \tag{18}$$

We define the following three cases as solutions to equations (13)-(16):

(H) $(u_H^h, q_H^h, u_H^l, q_H^l)$ solves (13)-(16) with $\mu_h > 0, \mu_l = 0$,

(L) $(u_L^h, q_L^h, u_L^l, q_L^l)$ solves (13)-(16) with $\mu_h = 0, \mu_l > 0$ and

[17] The concavity assumption can be stated in terms of the bordered Hessian of problem $(P_{u_b^h,u_b^l})$. However, it is more conveniently stated after some more notation is introduced. This is done in equations (30) and (33) below.

(P) $(u^h_P, q^h_P, u^l_P, q^l_P)$ solves (13)-(16) with $\mu_h = \mu_l = 0$.

It is straightforward to verify that in case of pooling (P) we have $q^h_P = q^l_P = 1$. However, we are interested in the case where $q^h \neq q^l$ (as this is the relevant case in practice). Further, below we consider the role of risk adjustment to enhance efficiency. This is only relevant for the case where for one type we have $q < 1$. Hence we will ignore the pooling case from here onwards and focus on separating equilibria (H) and (L).

The following proposition derives sufficient conditions for an (H) and (L) equilibrium.

Proposition 3 *With single crossing, solution (H) is a symmetric equilibrium and we have $q^h = 1 > q^l$. If single crossing is not satisfied and*

- *the solution $(u^h_H, q^h_H, u^l_H, q^l_H)$ under (H) above satisfies*

$$p(1, u^h_H, \theta^h) - p(1, u^l_H, \theta^l) = \int_{q^l_H}^{1} (p_q(q, u^h_H, \theta^h) - p_q(q, u^l_H, \theta^l))dq \geq 0 \tag{19}$$

  then this solution is a symmetric equilibrium and we have $q^h = 1 > q^l$

- *the solution $(u^h_L, q^h_L, u^l_L, q^l_L)$ under (L) above satisfies*

$$p(1, u^l_L, \theta^l) - p(1, u^h_L, \theta^h) = \int_{q^h_L}^{1} (p_q(q, u^l_L, \theta^l) - p_q(q, u^h_L, \theta^h))dq \geq 0 \tag{20}$$

  then this solution is a symmetric equilibrium and we have $q^l = 1 > q^h$.

The condition (19) [(20)] in the (H) [(L)] outcome makes sure that (IC_l) [(IC_h)] is satisfied as well. If (SC) is satisfied, equation (IC_l) is automatically satisfied for the following reason. Because of incentive compatibility, the two equilibrium indifference curves have to cross somewhere. Under (SC) the high type's indifference curve is steeper there (and consequently above the low type's curve for slightly higher q). By (SC) there is no other intersection and therefore the high type's indifference curve is above the low type's also at $q = 1$. Hence condition (19) is satisfied. Without single crossing, the θ^h indifference curve is less steep than the θ^l indifference curve for low q and the other way around for q close to 1. In that case, we need to check explicitly in the solution whether the IC constraint that was ignored is indeed satisfied.

To prove existence of the (H) and (L) equilibria in an oligopoly framework, it is sufficient to give an example for each. The existence of (H) equilibria in oligopoly is already established in the literature through models satisfying the single crossing assumption, see for example Olivella and Vera-Hernández (2007). Here we give an example using the utility setup of section 3 to demonstrate existence of a (L) equilibrium with oligopoly. Recall from proposition 1 that there is no (L) equilibrium with perfect competition.

Example (cont.) *On the supply side we assume that there are two insurers located at the end points 0 and 1 of a Hotelling line. Agents are uniformly distributed over the* $[0, 1]$ *interval. An agent at position* $x \in [0, 1]$ *incurs transportation cost* xt $((1 - x)t)$ *when buying from insurer a (b). Each insurer offers a menu of contracts* $\{(q^h, p^h), (q^l, p^l), (0, 0)\}$ *where the first contract is intended for the* θ^h *type, the second for the* θ^l *type and the third "contract" denotes the agent's outside option of not buying insurance at all (which will not be used in equilibrium).*

We will show that it is straightforward to find examples with a (L) equilibrium. The easiest way to do this is to find parameter values such that the individual rationality (IR) curve (that is, the indifference curve $p(q, \bar{u}, \theta)$*) for the* θ^l *type lies everywhere above the IR curve for the* θ^h *type. As shown in figure 1 this is the case for the parameter values chosen in section 3. Clearly the Hotelling equilibrium contracts have to lie on or below the relevant IR curves.*

First, assume that $\phi = 0$*. In words, there are only* θ^l *types. Then it is routine to verify that* $q^l = 1$ *(because of assumption 3) and the Hotelling equilibrium price on the* θ^l*-market equals* $p^l = F^l \bar{h} + t$*.*[18] *This contract is denoted* $(1, p^l)$ *in figure 1 for the parameter values in section 3 and* $t = 0.018$*. As this contract lies below* θ^l*'s IR curve it is, indeed, the equilibrium outcome. Let* $u^l_{hotel.}$ *denote* θ^l*'s utility level associated with the* $(1, p^l)$ *contract:* $u^l_{hotel.} = u(1, p^l, \theta^l)$*. Contract* (q^h, p^h) *(although not bought by anyone as* $\phi = 0$*) is defined by the intersection of indifference curve* $p(q, u^l_{hotel.}, \theta^l)$ *(dashed curve in the figure) and* θ^h*'s IR curve. This is the best contract on* θ^h*'s IR curve that satisfies* θ^l*'s incentive compatibility constraint.*

Now increase ϕ *slightly to* $\phi > 0$ *(but small). We claim that this results in a (L) equilibrium with* $q^l = 1 > q^h$*. For this to be an equilibrium we need that the indifference curve for the* θ^l *type at* $q = 1$ *lies above the indifference curve for the* θ^h *type at* $q = 1$*. Note that the equilibrium indifference curve for the* θ^h *type* $(p(q, u^h_{hotel}, \theta^h))$ *cannot lie above* θ^h*'s IR curve. Hence a sufficient condition for a (L) equilibrium is that* θ^l*'s indifference curve* $p(q, u^l_{hotel.}, \theta^l)$ *at the new Hotelling equilibrium lies above* θ^h*'s* IR *curve at* $q = 1$*. We formally show that this is the case in lemma 3 in the appendix. Intuitively, small changes in* ϕ *will lead to small changes in the indifference curve* $p(q, u^l_{hotel.}, \theta^l)$*. As this curve is above* θ^h*'s* IR *curve at* $q = 1$ *in case* $\phi = 0$*, it will be above* θ^h*'s* IR *curve for small positive values of* ϕ*.*

Hence a straightforward way to generate (L) equilibria is to find examples where the IR constraint for the θ^l type lies above the IR constraint for the θ^h type for each $q \in \langle 0, 1]$. Then there exist $t > 0$ and $\phi > 0$ such that the example has an (L) equilibrium.

Having proved the existence of a (L) equilibrium, we move to the policy implications of such an

[18] Recall that in a Hotelling model with constant marginal costs c, the equilibrium price is given by $c + t$. See, for instance, Tirole (1988, pp. 280).

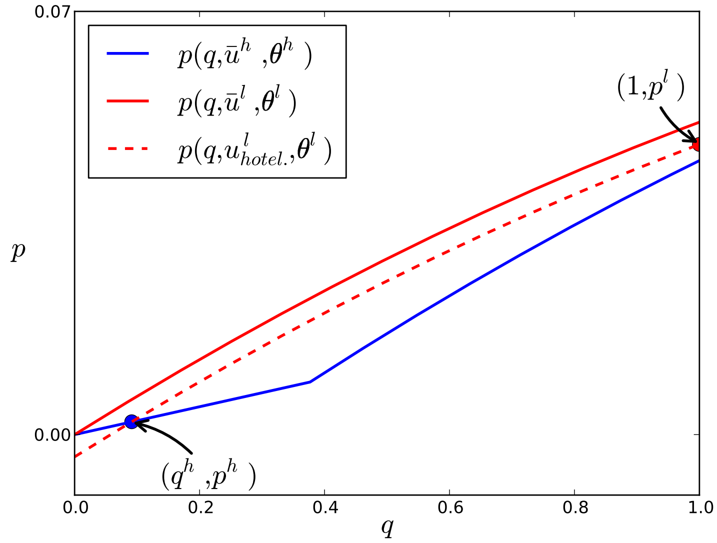

Figure 1: Example with parameter values in section 3 and $t = 0.018$

equilibrium.

6. Risk adjustment

Above we have shown that an equilibrium can exist where θ^h types have less than full insurance. This is in line with the empirical findings mentioned in the introduction where people with low income and low health status have less generous insurance than people with high income and good health. In this section, we show what the policy implications are of an equilibrium with $q^h < 1$. In particular, risk adjustment is often presented as a win-win policy: Starting from a separating equilibrium it increases both the utility of people with bad health (solidarity with those unlucky enough to be born this way)[19] and it increases equilibrium coverage (efficiency).

That is, when it comes to risk adjustment, there is no trade off between efficiency and distributional objectives for the government. The papers and country policies surveyed in Van de Ven and Ellis (2000) see risk adjustment as serving both efficiency and fairness or solidarity. We show below that under a

[19]The underlying assumption is that the θ^h type was born with, say a chronic disease like diabetes. Hence θ^h's high expected health care costs are exogenously given. Then fairness or solidarity considerations can lead the planner to give a higher weight than λ to the θ^h type in the objective function. Alternatively, high health care costs can be endogenous due to, for example, smoking behavior, food habits, drug use etc. See Van de Ven and Ellis (2000) for a discussion of the limits to solidarity due to such moral hazard effects.

relatively mild assumption, this result holds in an oligopoly model with SC.[20] However, in the NSC case, the same assumption implies that there is a trade off between efficiency and solidarity. In that case it is not possible for risk adjustment to increase both q^h and u^h. Hence in the empirically relevant case (with $q^h < 1$), the sponsor has to choose explicitly whether the goal is efficiency or solidarity. Risk adjustment cannot promote both goals.

We model risk adjustment as follows. The sponsor, say the government, of the health insurance scheme pays an insurer $\rho^h(\rho^l)$ for each of its customers that are of type $\theta^h(\theta^l)$. We assume that the government can perfectly observe each customer's type.[21]

In case (H), we write the profits of insurer a as follows:

$$\Pi(u_a^h, u_a^l, \rho, u_b^h, u_b^l) = \phi D(u_a^h, u_b^h)(\pi(u_a^h, \theta^h) + \rho^h) + (1-\phi)D(u_a^l, u_b^l)(\pi(u_a^l, u_a^h, \theta^l) + \rho^l) \tag{21}$$

where we use the following definition of the functions π (with a slight abuse of notation) capturing the profit margins on the θ^h and θ^l markets resp.:

$$\begin{aligned} \pi(u^h, \theta^h) &= p(1, u^h, \theta^h) - c(1, u^h, \theta^h) \\ \pi(u^l, u^h, \theta^l) &= p(q^l(u^l, u^h), u^l, \theta^l) - c(q^l(u^l, u^h), u^l, \theta^l) \end{aligned} \tag{22}$$

with $q^l(u^l, u^h)$ the value of q^l that solves equation (IC_h) with equality. If (SC) is satisfied, differentiating the binding (IC_h) yields

$$\left(p_q(q^l, u^l, \theta^l) - p_q(q^l, u^h, \theta^h)\right) \frac{dq^l}{du^l} = -p_u(q^l, u^l, \theta^l)$$

and hence

$$\frac{dq^l}{du^l} < 0 \tag{23}$$

since the left hand side is negative by (SC) and the right hand side is positive by assumption 1 ($p_u < 0$). Using a similar derivation, one can show in this case that

$$\frac{dq^l}{du^h} > 0 \tag{24}$$

In words, increasing u^h relaxes the (IC_h) constraint thereby allowing for higher q^l (for given u^l). Increasing u^l does the opposite and hence leads to lower q^l (for given u^h).

[20] Selden (1998) makes this point using a model with perfect competition where single crossing holds.

[21] This assumption is usually justified by assuming that the government has (ex post) more information than the insurer ex ante. Hence the insurer is not able to (explicitly) select risks ex ante but the government can perfectly risk adjust ex post. Alternatively, the insurer has the relevant information but is prevented by law to act upon this information. To illustrate, in the Netherlands an insurer cannot refuse a customer who wants to buy a certain insurance contract. As mentioned, Glazer and McGuire (2000) show how optimal risk adjustment can work with imperfect signals of customers' types. In the latter case, the government cannot perfectly observe each customer's type.

In case (L) we have:

$$\tilde{\Pi}(u_a^h, u_a^l, \rho, u_b^h, u_b^l) = \phi D(u_a^h, u_b^h)(\tilde{\pi}(u_a^h, u_a^l, \theta^h) + \rho^h) + (1-\phi)D(u_a^l, u_b^l)(\tilde{\pi}(u_a^l, \theta^l) + \rho^l) \tag{25}$$

where we use the following definition of the functions $\tilde{\pi}$:

$$\begin{aligned} \tilde{\pi}(u^l, \theta^l) &= p(1, u^l, \theta^l) - c(1, u^l, \theta^l) \\ \tilde{\pi}(u^h, u^l, \theta^h) &= p(q^h(u^h, u^l), u^h, \theta^h) - c(q^h(u^h, u^l), u^h, \theta^h) \end{aligned} \tag{26}$$

with $q^h(u^h, u^l)$ being the value of q^h that solves equation (IC_l) with equality. If the θ^l indifference curve is steeper than the θ^h indifference curve at q^h, we find –with a similar derivation as above– that

$$\begin{aligned} \frac{dq^h}{du^h} &< 0 \\ \frac{dq^h}{du^l} &> 0 \end{aligned} \tag{27}$$

In this case, raising u^l relaxes the (IC_l) constraint and hence allows for higher q^h (for given u^h).

Let Π_i denote the derivative of Π with respect to its i-th argument. Then for the (H) case, a symmetric equilibrium (focusing on an interior maximum) is characterized by $\Pi_1 = \Pi_2 = 0$ which we write as

$$\phi D_1(u^h, u^h)(\pi(u^h, \theta^h) + \rho^h) + \phi D(u^h, u^h)\pi_1(u^h, \theta^h) + (1-\phi)D(u^l, u^l)\pi_2(u^l, u^h, \theta^l) = 0 \tag{28}$$

$$(1-\phi)\Big(D_1(u^l, u^l)(\pi(u^l, u^h, \theta^l) + \rho^l) + D(u^l, u^l)\pi_1(u^l, u^h, \theta^l)\Big) = 0 \tag{29}$$

together with the second order conditions:

$$\begin{aligned} \Pi_{11} < 0, \Pi_{22} < 0 \\ \Pi_{11}\Pi_{22} - \Pi_{12}^2 > 0 \end{aligned} \tag{30}$$

The interpretation of the first order condition with respect to u_a^h (equation (28)) is as follows. By increasing u_a^h slightly, insurer a increases the demand for its product on the θ^h market on which it earns the margin $\pi^h + \rho^h$. On the inframargin, however, increasing u^h (or equivalently reducing p^h) leads to lower profits ($\pi_1 < 0$). Finally, there is the effect of u^h on the θ^l market. By increasing u^h, the firm can raise q^l and still satisfy (IC_h). Hence u^h affects the profits on the θ^l market.

The first order condition with respect to u^l (equation (29)) only features the marginal and inframarginal effects on the θ^l market.

In the (L) case, we have $\tilde{\Pi}_1 = \tilde{\Pi}_2 = 0$ which we write as

$$\phi D_1(u^h, u^h)(\tilde{\pi}(u^h, u^l, \theta^h) + \rho^h) + \phi D(u^h, u^h)\tilde{\pi}_1(u^h, u^l, \theta^h) = 0 \tag{31}$$

$$(1-\phi)\Big(D_1(u^l, u^l)(\tilde{\pi}(u^l, \theta^l) + \rho^l) + D(u^l, u^l)\tilde{\pi}_1(u^l, \theta^l)\Big) + \phi D(u^h, u^h)\tilde{\pi}_2(u_a^h, u_a^l, \theta^h) = 0 \tag{32}$$

together with the second order conditions:

$$\begin{aligned} \tilde{\Pi}_{11} < 0, \tilde{\Pi}_{22} < 0 \\ \tilde{\Pi}_{11}\tilde{\Pi}_{22} - \tilde{\Pi}_{12}^2 > 0 \end{aligned} \tag{33}$$

The interpretation of the first order conditions with respect to u^h and u^l is the same as in the (H) equilibrium. The only difference is that in this case u^l has the additional effect on the θ^h market via the (IC_l) constraint.

We use the following assumption that we discuss below.

Assumption 5 *Assume that both markets are fully covered:*

$$D(u_j^i, u_{-j}^i) + D(u_{-j}^i, u_j^i) = 1$$

for $i = h, l$. Consider changes in risk adjustment with a fixed budget: $d\rho^h = d\rho, d\rho^l = -\frac{\phi}{1-\phi}d\rho$ such that

$$\phi d\rho^h + (1-\phi)d\rho^l = 0 \tag{34}$$

Then we assume that in symmetric equilibrium $u^h(\rho), u^l(\rho)$ it is the case that

$$\frac{du^h(\rho)}{d\rho}\frac{du^l(\rho)}{d\rho} < 0 \tag{35}$$

Health insurance is a product of which one buys one unit or zero. In fact, it is often not allowed to buy more than one insurance contract due to moral hazard problems (see Pauly (1974)). We follow the literature on competition with price discrimination and assume that both markets are fully covered (this is called "full scale competition" by Schmidt-Mohr and Villas-Boas (1999) and "pure competition" by Stole (1995)). When risk adjustment is implemented at the country level, it usually goes hand in hand with mandatory insurance (e.g. as in the Netherlands). Indeed, if insurance would not be mandatory, a cross subsidy (due to risk adjustment) may induce the net payers to forgo insurance thereby reducing the efficiency of the insurance market. If risk adjustment is organized at the firm level (as, for instance, in the US with big employers) the scheme is usually so attractive (due to tax breaks) that the assumption of full scale competition is not an unreasonable one. The advantage of making this assumption is that the budget constraint for changes in risk adjustment can simply be written as (34).

Assumption (35) states that if an increase in ρ raises u^h, then it decreases u^l. This is a natural assumption: as ρ^h increases (ρ^l decreases) one expects insurers to compete more for θ^h types (compete less for θ^l types) and hence u^h increases (u^l decreases). We are especially interested here in the (L) equilibrium. Lemma 4 in the appendix verifies that for the example above this assumption is indeed satisfied.[22]

We further motivate this assumption in two ways. The first is to assume that the planner choosing ρ has an objective function $f(u^h(\rho), u^l(\rho))$ that is increasing in both u^h and u^l (i.e. $f_1, f_2 > 0$). A solution to this optimization problem with respect to ρ is characterized by

$$f_1 \frac{du^h(\rho)}{d\rho} + f_2 \frac{du^l(\rho)}{d\rho} = 0$$

hence $\frac{du^h(\rho)}{d\rho}$ and $\frac{du^l(\rho)}{d\rho}$ have opposite signs. In words, the planner chooses a value of ρ that places the insurance on the Pareto frontier in (u^h, u^l) space. Put differently, if a change in ρ can raise the utility of both types, the planner exhausts such possibilities. Then, at the margin, there is trade off between u^h and u^l.

The second way to motivate inequality (35) is in terms of equilibrium selection. In particular, assumption 5 rules out that there exists another equilibrium close by in which both u^h and u^l are lower.

Lemma 2 *Assume that*

$$D_1(u,u) \text{ is independent from the level of } u$$

Let u^h, u^l denote a symmetric equilibrium. Define the following set

$$B_\varepsilon(u^h, u^l) = \left\{ \tilde{u}^h < u^h, \tilde{u}^l < u^l | \sqrt{(u^h - \tilde{u}^h)^2 + (u^l - \tilde{u}^l)^2} < \varepsilon \right\} \tag{36}$$

If assumption 5 holds, then for $\varepsilon > 0$ small enough, the set $B_\varepsilon(u^h, u^l)$ does not contain another symmetric equilibrium.

The assumption implies that in symmetric equilibrium the slope of a firm's demand function is not affected by the level of u. This assumption is satisfied in the Hotelling example that we use in this paper.

The lemma excludes the possibility that close to the symmetric equilibrium u^h, u^l there would be another equilibrium with lower values for both u^h and u^l. Clearly, such an equilibrium would be more profitable for both insurers because a first order Taylor expansion implies that

$$\begin{aligned}
\Pi(\tilde{u}^h, \tilde{u}^l, \rho, \tilde{u}^h, \tilde{u}^l) &= \Pi(u^h, u^l, \rho, u^h, u^l) + (\Pi_1(u^h, u^l, \rho, u^h, u^l) + \Pi_4(u^h, u^l, \rho, u^h, u^l))(\tilde{u}^h - u^h) \\
&+ (\Pi_2(u^h, u^l, \rho, u^h, u^l) + \Pi_5(u^h, u^l, \rho, u^h, u^l))(\tilde{u}^l - u^l) \\
&= \Pi(u^h, u^l, \rho, u^h, u^l) - \Pi_4(u^h, u^l, \rho, u^h, u^l)(u^h - \tilde{u}^h) - \Pi_5(u^h, u^l, \rho, u^h, u^l)(u^l - \tilde{u}^l) \\
&> \Pi(u^h, u^l, \rho, u^h, u^l)
\end{aligned}$$

[22] The proof of lemma 4 is given after the proof of lemma 2 below to avoid duplication. That is, we first derive some general results in the context of lemma 2 and then apply these to the example in lemma 4.

for $\tilde{u}^{h,l}$ close enough to $u^{h,l}$; where we have used the stationarity conditions $\Pi_1 = \Pi_2 = 0$ and $\Pi_4, \Pi_5 < 0$. And similarly for profit function $\tilde{\Pi}$. In this sense, assumption 5 rules out a local multiplicity of equilibria leading to a coordination problem between insurance providers.

Now we can show the following.

Proposition 4 *Let u^h, u^l denote the utility levels for type θ^h, θ^l in a symmetric equilibrium. Assume that the indifference curves $p(q, u^h, \theta^h)$ and $p(q, u^l, \theta^l)$ have one point of intersection. Then the relation between u^h and efficiency is as follows:*

- *if $q^h = 1$, a change in ρ which raises q^l also increases u^h;*
- *if $q^l = 1$, a change in ρ which raises q^h reduces u^h.*

The assumption that the indifference curves corresponding to the equilibrium values of u^h, u^l cross once is by definition satisfied if (SC) holds. SC implies that any indifference curves cross at most once. In the NSC case, we need a bit more structure. In figure 1 the assumption is satisfied. In fact, the following reasoning shows that a model with more than one intersection point will not be very intuitive. In the (L) equilibrium, (20) implies that $p(1, u^l, \theta^l) > p(1, u^h, \theta^h)$. Further, (C1) implies that $p_q(1, u^l, \theta^l) < p_q(1, u^h, \theta^h)$. Hence, at the first intersection point (at q^h) we have $p_q(q^h, u^l, \theta^l) > p_q(q^h, u^h, \theta^h)$. To generate another intersection point, we need again a switch in slopes $p_q(q, u^l, \theta^l) < p_q(q, u^h, \theta^h)$ for $q < q^h$. As the slope is related to the treatment choice in our model with income differences, this would imply that for q close to zero, the low income θ^h type decides to spend more on treatment than the high income θ^l type. This does not seem very reasonable.[23]

Hence, in the case where $q^h = 1$ we find that (at the margin) solidarity (with the unlucky θ^h type) and efficiency (q^l) go hand in hand. In terms of the objective function $f(u^h, u^l)$ above: if the planner decides to put relatively more weight on the θ^h type (i.e. f_1 increases relatively to f_2), she will change ρ in such a way that both u^h and q^l increase.

However, if single crossing is violated and we have $q^h < 1$, then it is not possible to raise both u^h and q^h. In other words, in this case there is a trade off between solidarity and efficiency.

The violation of (SC) is vital to get this result: At (p^h, q^h) the slope of θ^l's indifference curve is steeper than the slope of the indifference curve of θ^h. Therefore, an increase in q^h (efficiency) is valued

[23] Another way to put more structure on outcomes in the NSC case is the following. One can allow for more than one intersection point and then strengthen assumption 3 to read that $p(q, u^k, \theta^k) - c(q, u^k, \theta^k)$ is increasing in q. Then firms will choose the intersection point with highest q. It is routine to verify that this gives $p_q(q^l, u^h, \theta^h) > p_q(q^l, u^l, \theta^l)$ in an (H) equilibrium and $p_q(q^h, u^l, \theta^l) > p_q(q^h, u^h, \theta^h)$ in an (L) equilibrium. Under this assumption, the results in proposition 4 hold as well.

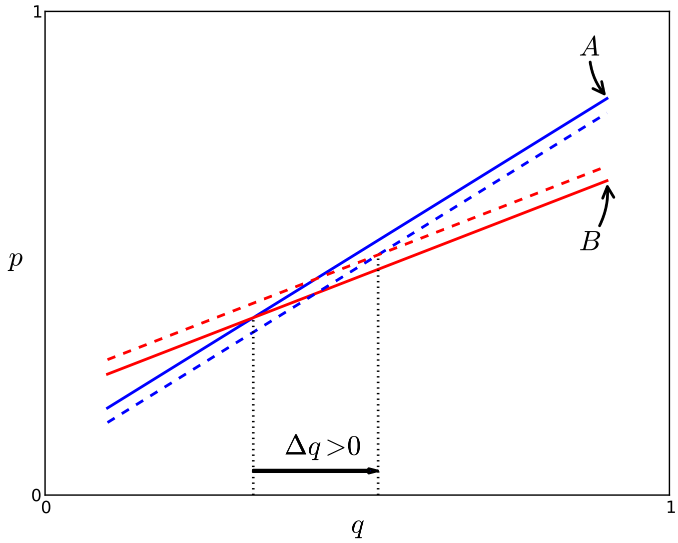

Figure 2: A change in risk adjustment which increases q.

higher by the θ^l type. Consequently, the utility of θ^h can only be increased if q^h is reduced as otherwise incentive compatibility fails.

Figure 2 illustrates this graphically. The lines A and B denote parts of indifference curves of a θ^l and θ^h type. The intersection of these lines gives the coverage q of the type that has less than full coverage. Given assumption (35), a change in risk adjustment that increases q has to shift the A curve downwards and the B curve upwards. In the (H) equilibrium, the (steeper) A curve is the indifference curve of the θ^h type. The B curve is the indifference curve of the θ^l type. The point of intersection determines q^l. Hence increasing q^l and u^h go hand in hand.

However, in the (L) equilibrium the steeper A curve is the indifference curve of the θ^l type. Increasing q^h then requires that the indifference curve B for the θ^h type shifts upwards. That is, the utility for the θ^h type falls.

7. Conclusion

Standard insurance models, e.g. Rothschild and Stiglitz (1976) or Stiglitz (1977), predict higher coverage for agents with higher risks. We show that this prediction no longer holds if single crossing is violated and firms have market power.

In the health care sector agents with higher income have lower risks and more insurance. Put differently, the predictions of the standard insurance model with single crossing are contradicted by the data. We show that the negative correlation between income and risk can cause a violation of single crossing. With a violation of single crossing, the empirical findings in the health literature can be reconciled with a standard insurance model.

We show the policy implication of such a violation of single crossing for risk adjustment. The traditional insurance model (with single crossing) views risk adjustment as a measure increasing efficiency as well as solidarity, i.e. coverage levels are closer to first best and low health agents are better off. Without single crossing there is a trade off between these two policy goals. This implies that the sponsor of the health insurance scheme has to be explicit about the goals of risk adjustment.

From an empirical point of view, our paper casts doubt on the positive correlation test: Given our result that separating equilibria exist in which agents with higher risk have less coverage (negative correlation), it is evident that the results of such a test have to be interpreted with care. In particular such a test cannot be used to test for the presence of asymmetric information when single crossing is violated.

We conclude with a discussion of advantageous selection. This can be modeled by assuming that people differ in their preferences for risks. If high risk individuals are less risk averse than low risk people, it can happen that consumers who are willing to pay the most for health insurance are people with low expected health care costs. Hence offering health insurance with high coverage is especially attractive for agents with low expected costs: advantageous selection. The implication of some advantageous selection models is that policies that stimulate insurance coverage are welfare reducing. In fact, there may be over-insurance in equilibrium. See Einav and Finkelstein (2011) for a recent review of advantageous selection and empirical papers documenting this in health care markets.

In our model (in the (L) equilibrium), we also see that at the margin low risk types are willing to pay more for insurance than high risk types. This is caused by the fact that at less than perfect coverage, low income, high risk types tend to reduce expenditure on treatments. Basically, they cannot afford the treatments that they need. Hence, although the equilibrium is an advantageous selection equilibrium, in our model stimulating insurance coverage (e.g. through mandatory insurance at full coverage) is efficient (because of assumption 3).

References

Araujo, A., and H. Moreira (2010): "Adverse Selection Problems without the Spence-Mirrlees Condition," *Journal of Economic Theory*, 145(5), 1113–1141.

Armstrong, J., F. Paolucci, H. McLeod, and W. Van de Ven (2010): "Risk equalisation in voluntary health insurance markets: a three country comparison," *Health Policy*, 98, 39–49.

Armstrong, M., and J. Vickers (2001): "Competitive price discrimination," *RAND Journal of Economics*, 32(4), 579–605.

Bajari, P., H. Hong, and A. Khwaja (2006): "Moral hazard, adverse selection and health expenditures: A semiparametric analysis," *NBER Working paper*.

Bijlsma, M., J. Boone, and G. Zwart (2011): "Competition leverage and risk adjustment in health insurance markets," CEPR DP No. 8461.

Cardon, J. H., and I. Hendel (2001): "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *The RAND Journal of Economics*, 32(3), 408–427.

Cohn, J. (2007): *Sick: the untold story of America's health care crisis–and the people who pay the price*. Harper Perennial.

Dafny, L. (2010): "Are health insurance markets competitive?," *American Economic Review*, 100(4), 1399–1431.

De Meza, D., and D. Webb (2001): "Advantageous selection in insurance markets," *RAND Journal of Economics*, 32(2), 249–262.

DeNavas-Walt, C., B. Proctor, and J. Smith (2008): "Income, poverty, and health insurance coverage in the United States: 2008," *US Census Bureau. Current population reports, series P*, 60.

Dowd, B., R. Feldman, S. Cassou, and M. Finch (1991): "Health plan choice and the utilization of health care services," *The Review of Economics and Statistics*, 73(1), 85–93.

Einav, L., and A. Finkelstein (2011): "Selection in Insurance Markets : Theory and Empirics in Pictures," *Journal of Economic Perspectives*, 25(1), 115–138.

Ellis, R. P. (2007): "Risk adjustment in health care markets: concepts and applications," Boston Univsersity.

FINKELSTEIN, A., AND K. MCGARRY (2006): "Multiple dimensions of private information: evidence from the long-term care insurance market," *The American Economic Review*, 96(4), 938–958.

FRIJTERS, P., J. HAISKEN-DENEW, AND M. SHIELDS (2005): "The Causal Effect of Income on Health: Evidence from German Reunification," *Journal of Health Economics*, 24(5), 997–1017.

FUDENBERG, D., AND J. TIROLE (1991): *Game theory.* MIT Press.

GLAZER, J., AND T. G. MCGUIRE (2000): "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care," *The American Economic Review*, 90(4), 1055–1071.

GLICKSBERG, I. L. (1952): "A Further Generalization of the Kakutani Fixed Point Theorem with Application to Nash Equilibrium Points," *Proceedings of the American Mathematical Society*, 38, 170–174.

GOLDMAN, D., G. JOYCE, AND Y. ZHENG (2007): "Prescription drug cost sharing: associations with medication and medical utilization and spending and health," *The Journal of the American Medical Association (JAMA)*, 298(1), 61–69.

GRAVELLE, H., AND M. SUTTON (2009): "Income, Relative Income, and Self-Reported Health in Britain 1979-2000," *Health Economics*, 18(2), 125–145.

HEMENWAY, D. (1990): "Propitious Selection," *The Quarterly Journal of Economics*, 105(4), 1063–1069.

JACK, W. (2006): "Optimal risk adjustment with adverse selection and spatial competition," *Journal of Health Economics*, 25(5), 908 – 926.

JULLIEN, B., B. SALANIE, AND F. SALANIE (2007): "Screening risk-averse agents under moral hazard: Single-crossing and the CARA case," *Economic Theory*, 30(1), 151–169.

KUTTNER, R. (1999): "Health Insurance Coverage," *The New England Journal of Medicine*, 340, 163–168.

LI, Q., AND P. TRIVEDI (2010): "Medicare Health Plan Choice of the Elderly - A Choice-with-Screening Model," mimeo.

MUNKIN, M., AND P. TRIVEDI (2010): "Disentangling incentives effects of insurance coverage from adverse selection in the case of drug expenditure: a finite mixture approach," *Health Economics*, 19(9), 1093–1108.

Netzer, N., and F. Scheuer (2010): "Competitive Screening in Insurance markets with endogenous Wealth heterogeneity," *Economic Theory*, 44(2), 187–211.

Olivella, P., and M. Vera-Hernández (2007): "Competition among differentiated health plans under adverse selection," *Journal of Health Economics*, 26(2), 233 – 250.

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

Piette, J., M. Heisler, and T. Wagner (2004a): "Cost-related medication underuse among chronically ill adults: the treatments people forgo, how often, and who is at risk," *American Journal of Public Health*, 94(10), 1782–1787.

——— (2004b): "Problems paying out-of-pocket medication costs among older adults with diabetes," *Diabetes Care*, 27(2), 384–391.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90(4), 629–649.

Schmidt-Mohr, U., and M. Villas-Boas (1999): "Oligopoly with Asymmetric Information: Differentiation in Credit Markets," *RAND Journal of Economics*, 30, 375–396.

Schoen, C., S. R. Collins, J. L. Kriss, and M. M. Doty (2008): "How many are underinsured? Trends among U.S. adults, 2003 and 2007.," *Health affairs (Project Hope)*, 27(4), 298–309.

Schoen, C., R. Osborn, D. Squires, M. M. Doty, R. Pierson, and S. Applebaum (2010): "How Health Insurance Design Affects Access To Care And Costs, By Income, In Eleven Countries," *Health Affairs*, 29(12), 1–12.

Schottmüller, C. (2011): "Adverse selection without single crossing: The monotone solution," mimeo.

Selden, T. (1998): "Risk adjustment for health insurance: Theory and implications," *Journal of Risk and Uncertainty*, 17(2), 167–182.

Smart, M. (2000): "Competitive Insurance Markets with two Unobservables," *International Economic Review*, 41(1), 153–170.

Stiglitz, J. (1977): "Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market," *The Review of Economic Studies*, 44(3), 407–430.

STOLE, L. (1995): "Nonlinear Pricing and Oligopoly," *Journal of Economics and Management Strategy*, 4, 529–562.

——— (2007): Price discrimination and competition in Armstrong, M. and Porter, R., Ed., *Handbook of industrial organization*pp. 2221–2299. North Holland.

TIROLE, J. (1988): *The theory of industrial organization.* MIT Press.

VAN DE VEN, W., AND R. ELLIS (2000): Risk adjustment in competitive health care markets, in Culyer, A.P. and Newhouse, J. P., Ed., *Handbook of health economics*pp. 754–845. North Holland.

WAMBACH, A. (2000): "Introducing Heterogeneity in the Rothschild-Stiglitz Model," *The Journal of Risk and Insurance*, 67(4), 579–591.

A. Proof of results

Proof of proposition 1 Suppose to the contrary that $q^h < 1$ in equilibrium. The contract (q^h, p^h) leads to nonnegative profits; otherwise it would not be offered in equilibrium. Denote by u^h the utility level θ^h derives from (q^h, p^h) and by $p(q, u^h, \theta^h)$ the indifference curve of θ^h associated with his contract. By assumption 3, the contract $(1, p(1, u^h, \theta^h))$ for type θ^h yields higher profits than (q^h, p^h). For $\varepsilon > 0$ small enough the contract $(1, p(1, u^h, \theta^h) - \varepsilon)$ is strictly preferred by θ^h to (q^h, p^h) and yields higher profits than (q^h, p^h). If the contract $(1, p(1, u^h, \theta^h) - \varepsilon)$ also attracts θ^l types, profits will remain positive as those are better risks. Therefore, $(1, p(1, u^h, \theta^h) - \varepsilon)$ is a profitable deviation, i.e. a contract with strictly positive profits and demand. Consequently, $q^h < 1$ cannot be an equilibrium. *Q.E.D.*

Proof of lemma 1 We start with the proof of the second statement. Suppose both incentive constraints were binding, i.e. θ^h and θ^l are both indifferent between the two contracts. First, look at the case where $q^h, q^l < 1$. Call the utility levels of the two types under the equilibrium contracts u^l and u^h. Now take the indifference curves corresponding to these utility levels and call them $p(q, u^l, \theta^l)$ and $p(q, u^h, \theta^h)$ and define $\iota = \arg\max_{k \in \{l,h\}} p(1, u^k, \theta^k)$. Changing θ^ι's menu point to $(1, p(1, u^\iota, \theta^\iota))$ will increase profits by assumption 3. By the definition of ι, this change is also incentive compatible.

Second, take the case where $q^k = 1$ and $q^{-k} < 1$ for some $k \in h, l$ and suppose again that both incentive constraints were binding. But according to assumption 3 pooling on the contract of θ^k would lead to higher profits. Hence, at most one incentive constraint is binding.

$q^\iota = 1$ follows from the argument in the first step and therefore at least one type has to have full coverage. *Q.E.D.*

Proof of proposition 2 Define $\iota = \arg\max_{k \in \{h,l\}} p(1, \bar{u}^k, \theta^k)$. By lemma 1, one type has full coverage. Suppose that $q^\iota < 1$ and therefore $q^\kappa = 1$ with $\kappa \in \{h, l\}$ and $\kappa \neq \iota$. Note that the individual rationality constraint of θ^ι cannot be binding as otherwise θ^ι would misrepresent as θ^κ by the definition of ι. But then the incentive compatibility constraint of θ^ι has to be binding as the monopolist could increase p^ι otherwise. By assumption 3, the monopolist could achieve a higher profit by pooling both types on θ^κ's contract. This contradicts the optimality of $q^\iota < 1$.

If both types are pooled, the optimal contract will be $(q, p) = (1, p(1, \bar{u}^\kappa, \theta^\kappa))$ and the individual rationality constraint of θ^κ will be binding. If the types are separated, the incentive compatibility constraint of θ^κ cannot bind: Since $q^\iota = 1$, pooling on θ^ι's contract would lead to higher profits by assumption 3 if the incentive constraint was binding. As increasing p^κ relaxes the incentive compatibility constraint of θ^ι, the individual rationality constraint of θ^κ has to bind: Otherwise, increasing p^κ would increase profits.

Last note that increasing p^ι would be feasible and increase profits if neither the incentive compatibility nor the individual rationality constraint of θ^l was binding. *Q.E.D.*

Proof of proposition 3 With single crossing, the solution (H) also satisfies (IC_l) because (SC) implies that

$$p(1, u_H^h, \theta^h) - p(1, u_H^l, \theta^l) = \int_{q_H^l}^{1} (p_q(q, u_H^h, \theta^h) - p_q(q, u_H^l, \theta^l) > 0 \qquad (37)$$

By the concavity assumption on problem $(P_{u_b^h, u_b^l})$, the solution (H) is a symmetric equilibrium.

If single crossing is not satisfied, we need to check that (IC_l) in case (H) ((IC_h) in case (L)) is satisfied. Using an equation similar to (37) this is the assumption given in each case. *Q.E.D.*

Lemma 3 *In the example depicted in figure 1 (L) equilibria exist for $\phi > 0$.*

Proof of lemma 3 We start with two straightforward observations. First, for $\phi = 0$ the game has the unique equilibrium illustrated in figure 1: By assumption 3 $q^l = 1$ is optimal and then the game is a standard Hotelling game with exogenous location. Second, an equilibrium will exist for each positive ϕ. This follows immediately from the existence theorem in Glicksberg (1952).[24]

Now take a sequence of $\phi_n > 0$ converging to 0, e.g. the sequence $\{1/n\}_{n=1,2,...}$. In the game where $\phi = \phi_n$, denote expected equilibrium utilities of type θ^l that are offered by firm $j \in \{a, b\}$ by u_n^{lj}.[25] Since u^l is chosen from a closed and bounded interval (see footnote 24), there is a converging subsequence of u_n^{la}. With a slight abuse of notation we denote the elements of this subsequence by u_n^{la} as well and continue to work with this subsequence only. To each u_n^{la} corresponds an equilibrium value u_n^{lb} (associated with the game where $\phi = \phi_n$). Again there will be a converging subsequence of u_n^{lb} because u^l is taken from a closed and bounded interval. Continuing in this way we can find a subsequence $(u_n^{ha}, u_n^{hb}, u_n^{la}, u_n^{lb}, q_n^{ha}, q_n^{hb}, q_n^{la}, q_n^{lb})$ of equilibria converging to some values $(\tilde{u}^{ha}, \tilde{u}^{hb}, \tilde{u}^{la}, \tilde{u}^{lb}, \tilde{q}^{ha}, \tilde{q}^{hb}, \tilde{q}^{la}, \tilde{q}^{lb})$.

To prove the lemma it is sufficient to show that $(\tilde{u}^{ha}, \tilde{u}^{hb}, \tilde{u}^{la}, \tilde{u}^{lb}, \tilde{q}^{ha}, \tilde{q}^{hb}, \tilde{q}^{la}, \tilde{q}^{lb})$ is an equilibrium for $\phi = 0$. From the uniqueness of the equilibrium for $\phi = 0$, it will then follow that there are equilibria for $\phi > 0$ where u^{lj} is arbitrarily close to u_0^l where u_0^l is the unique equilibrium utility of θ^l for $\phi = 0$.

But then q^h has to be strictly less than 1 (i.e. it is an (L) equilibrium): Type θ^l could achieve a discretely higher utility than u_0^l in a contract with full coverage where the individual rationality constraint of θ^h holds (see figure 1), i.e. the incentive compatibility constraint of θ^l is violated if $q^h = 1$ and $u^l \approx u_0^l$. It is evident from figure 1 that partial insurance is possible for θ^h types and it is

[24] The highest relevant utility level is the utility resulting if an agent gets full coverage for free. The individual rationality constraint gives a minimum utility level. Consequently, the action space is compact (this is obvious for $q \in [0, 1]$). As profit functions are continuous, the theorem applies.

[25] If there are several equilibria, one can choose an arbitrary one.

routine to check that these partial insurance contracts are profitable. Consequently, $0 < q_n^{hj} < 1$ for n high enough.

The last step –showing that $(\tilde{u}^{ha}, \tilde{u}^{hb}, \tilde{u}^{la}, \tilde{u}^{lb}, \tilde{q}^{ha}, \tilde{q}^{hb}, \tilde{q}^{la}, \tilde{q}^{lb})$ is an equilibrium for $\phi = 0$– follows Fudenberg and Tirole (1991, pp. 30) and is only sketched: Suppose, it was not an equilibrium. Then there is a deviation yielding a strictly higher profit for one firm. By the continuity of the profit function, however, this deviation would also increase profits for $(u_n^{ha}, u_n^{hb}, u_n^{la}, u_n^{lb}, q_n^{ha}, q_n^{hb}, q_n^{la}, q_n^{lb})$ close enough to $(\tilde{u}^{ha}, \tilde{u}^{hb}, \tilde{u}^{la}, \tilde{u}^{lb}, \tilde{q}^{ha}, \tilde{q}^{hb}, \tilde{q}^{la}, \tilde{q}^{lb})$ contradicting that $(u_n^{ha}, u_n^{hb}, u_n^{la}, u_n^{lb}, q_n^{ha}, q_n^{hb}, q_n^{la}, q_n^{lb})$ is an equilibrium under $\phi = \phi_n$. *Q.E.D.*

Proof of lemma 2 To see what assumption 5 implies in terms of derivatives of the profit function Π, we define the matrix A (and analogously $\tilde{A}$) as

$$A = \begin{pmatrix} \Pi_{11} + \Pi_{14} & \Pi_{12} + \Pi_{15} \\ \Pi_{12} + \Pi_{24} & \Pi_{22} + \Pi_{25} \end{pmatrix} \tag{38}$$

This allows us to linearize equations (28) and (29) as follows (and similarly for equations (31) and (32) with $\tilde{A}$)

$$A \begin{pmatrix} du^h \\ du^l \end{pmatrix} = 0$$

Then it is straightforward to verify that

$$\begin{aligned} \Pi_{24} &= 0 \\ \Pi_{15} &< 0 \\ \tilde{\Pi}_{15} &= 0 \\ \tilde{\Pi}_{24} &< 0 \end{aligned} \tag{39}$$

Inverting matrix A gives

$$A^{-1} = \frac{1}{\Delta} \begin{pmatrix} \Pi_{22} + \Pi_{25} & -(\Pi_{12} + \Pi_{15}) \\ -\Pi_{12} & \Pi_{11} + \Pi_{14} \end{pmatrix} \tag{40}$$

where

$$\Delta \equiv (\Pi_{11} + \Pi_{14})(\Pi_{22} + \Pi_{25}) - \Pi_{12}(\Pi_{12} + \Pi_{15}) \tag{41}$$

Hence, we find

$$\begin{pmatrix} \frac{du^h}{d\rho} \\ \frac{du^l}{d\rho} \end{pmatrix} = A^{-1} \begin{pmatrix} -\Pi_{13} \\ -\Pi_{23} \end{pmatrix} \tag{42}$$

with $\Pi_{13} = \phi D_1(u^h, u^h) > 0, \Pi_{23} = -\phi D_1(u^l, u^l) < 0$. Using the assumption that $D_1(u, u)$ is independent of u, we write

$$\frac{du^h}{d\rho} = \tfrac{1}{2}\phi \frac{-(\Pi_{22} + \Pi_{25}) - (\Pi_{12} + \Pi_{15})}{\Delta} D_1(u^h, u^h) \tag{43}$$

Similarly, we have

$$\frac{du^l}{d\rho} = \tfrac{1}{2}\phi \frac{\Pi_{12} + \Pi_{11} + \Pi_{14}}{\Delta} D_1(u^h, u^h) \tag{44}$$

Assumption 5 that $du^h/d\rho$ and $du^l/d\rho$ have opposite signs can be written as

$$sign(\Pi_{11} + \Pi_{14} + \Pi_{21} + \Pi_{24}) = sign(\Pi_{12} + \Pi_{15} + \Pi_{22} + \Pi_{25}) \tag{45}$$

Suppose –by contradiction– that such a second symmetric equilibrium would exist. We write $\delta^h = u^h - \tilde{u}^h, \delta^l = u^l - \tilde{u}^l$ where $\delta^h, \delta^l > 0$ are small since by assumption $(\tilde{u}^h, \tilde{u}^l) \in B_\varepsilon(u^h, u^l)$. Using a first order Taylor expansion of Π_1, Π_2 resp. we find

$$\begin{aligned}
\Pi_1(u^h - \delta^h, u^l - \delta^l, \rho, u^h - \delta^h, u^l - \delta^l) &= \Pi_1(u^h, u^l, \rho, u^h, u^l) + (\Pi_{11}(u^h, u^l, \rho, u^h, u^l) + \Pi_{14}(u^h, u^l, \rho, u^h, u^l))\delta^h \\
&\quad + (\Pi_{12}(u^h, u^l, \rho, u^h, u^l) + \Pi_{15}(u^h, u^l, \rho, u^h, u^l))\delta^l \\
\Pi_2(u^h - \delta^h, u^l - \delta^l, \rho, u^h - \delta^h, u^l - \delta^l) &= \Pi_2(u^h, u^l, \rho, u^h, u^l) + (\Pi_{21}(u^h, u^l, \rho, u^h, u^l) + \Pi_{24}(u^h, u^l, \rho, u^h, u^l))\delta^h \\
&\quad + (\Pi_{22}(u^h, u^l, \rho, u^h, u^l) + \Pi_{25}(u^h, u^l, \rho, u^h, u^l))\delta^l
\end{aligned}$$

Adding both equations and using the assumption that both (u^h, u^l) and $(\tilde{u}^h, \tilde{u}^l)$ are an equilibrium, we find that

$$\delta^h(\Pi_{11} + \Pi_{14} + \Pi_{21} + \Pi_{24}) + \delta^l(\Pi_{12} + \Pi_{15} + \Pi_{22} + \Pi_{25}) = 0 \tag{46}$$

However, it follows from equation (45) above that the expression $(\Pi_{11} + \Pi_{14} + \Pi_{21} + \Pi_{24})$ has the same sign as $(\Pi_{12} + \Pi_{15} + \Pi_{22} + \Pi_{25})$. Together with $\delta^h, \delta^l > 0$ this leads to a contradiction.

Similar argument can be given in case of $\tilde{\Pi}$. *Q.E.D.*

Lemma 4 *Consider the example with mean-variance utility and Hotelling competition. Then condition (35) is satisfied in (L) equilibrium.*

Proof of lemma 4 With Hotelling competition, $D_1(u, u) = 1$ as consumers are uniformly distributed on the line $[0, 1]$.[26] Hence $D_1(u, u)$ is independent of u and we can use equation (45). Following equation (45), we need to show that

$$sign(\tilde{\Pi}_{11} + \tilde{\Pi}_{14} + \tilde{\Pi}_{21} + \tilde{\Pi}_{24}) = sign(\tilde{\Pi}_{12} + \tilde{\Pi}_{22} + \tilde{\Pi}_{25}) \tag{47}$$

In our example, the equilibrium indifference curves at their point of intersection can be written as

$$\begin{aligned}
p(q^h, u^l, \theta^l) &= F^l(v(\bar{h}, \theta^l) - (1 - q^h)\bar{h}) + (1 - F^l)v(1, \theta^l) - u^l \\
&\quad - \tfrac{1}{2} r^l F^l (1 - F^l)(v(1, \theta^l) - v(\bar{h}, \theta^l + (1 - q^h)\bar{h})^2 \\
p(q^h, u^h, \theta^h) &= F^h(v(\underline{h}, \theta^h) - (1 - q^h)\underline{h}) + (1 - F^h)v(1, \theta^h) - u^h \\
&\quad - \tfrac{1}{2} r^h F^h (1 - F^h)(v(1, \theta^h) - v(\underline{h}, \theta^h) + (1 - q^h)\underline{h})^2
\end{aligned}$$

[26] More generally, if consumers are distributed with symmetric density function g on $[0, 1]$, it is also true that $D_1(u, u) = g(\frac{1}{2})$ is independent of u.

Further, $q^h(u^h, u^l)$ is defined as

$$p(q^h(u^h, u^l), u^h, \theta^h) \equiv p(q^h(u^h, u^l), u^l, \theta^l) \tag{48}$$

From these observations it follows that

$$\begin{aligned} p_u(q, u, \theta) &= -1 \\ p_{qu}(q, u, \theta) &= 0 \\ \frac{dq^h(u^h, u^l)}{du^h} &= -\frac{dq^h(u^h, u^l)}{du^l} < 0 \\ \frac{d^2q^h(u^h, u^l)}{(du^h)^2} &= -\frac{d^2q^h(u^h, u^l)}{du^h du^l} \end{aligned}$$

Using this we can write

$$\begin{aligned} \tilde{\Pi}_1 &= \frac{\phi}{2t}(p(q^h(u_a^h, u_a^l), u_a^h, \theta^h) - q^h(u_a^h, u_a^l)\underline{h}F^h + \rho^h) \\ &\quad + \phi\left(\tfrac{1}{2} + \frac{u_a^h - u_b^h}{2t}\right)\left((p_q(q^h(u_a^h, u_a^l), u_a^h, \theta^h) - \underline{h}F^h)\frac{dq^h(u_a^h, u_a^l)}{du_a^h} - 1\right) \\ \tilde{\Pi}_2 &= (1-\phi)\left(\frac{1}{2t}(p(1, u_a^l, \theta^l) - F^l\bar{h} + \rho^l) - \left(\tfrac{1}{2} + \frac{u_a^l - u_b^l}{2t}\right)\right) \\ &\quad + \phi\left(\tfrac{1}{2} + \frac{u_a^h - u_b^h}{2t}\right)(p_q(q^h(u_a^h, u_a^l), u_a^h, \theta^h) - \underline{h}F^h)\frac{dq^h(u_a^h, u_a^l)}{du_a^l} \end{aligned}$$

Now it is routine to verify that

$$\tilde{\Pi}_{11} + \tilde{\Pi}_{12} + \tilde{\Pi}_{14} = -\frac{\phi}{2t}\left(1 + (p_q(q^h, u^h, \theta^h) - \underline{h}F^h)\frac{dq^h(u^h, u^l)}{du^l}\right) < 0$$

with $\tilde{\Pi}_{24} < 0$ from equation (39). Moreover

$$\tilde{\Pi}_{12} + \tilde{\Pi}_{22} + \tilde{\Pi}_{25} = -\frac{\phi}{2t}(p_q(q^h, u^h, \theta^h) - \underline{h}F^h)\frac{dq^h(u^h, u^l)}{du^h} - \frac{1-\phi}{2t}$$

which we will shown to be negative in every (L) equilibrium. Suppose this were not the case, then

$$\phi(p_q(q^h, u^h, \theta^h) - \underline{h}F^h) \geq -(1-\phi)\frac{du^h}{dq^h(u^h, u^l)} \tag{49}$$

where $\frac{du^h}{dq^h(u^h,u^l)} < 0$ is the amount u^h changes if q^h is increased marginally while keeping u^l constant. The left hand side of (49) denotes the additional profits from θ^h types when marginally increasing q^h while keeping u^h fix. The right hand side of (49) denotes the marginal loss in profits if one reduces p^l such that a marginal increase of q^h with fixed u^h is incentive compatible for the θ^l types: As $p_u(q, u, \theta) = -1$ for both types, the necessary reduction in p^l is given by $-\frac{du^h}{dq^h(u^h,u^l)} > 0$. But then increasing q^h keeping u^h fixed and adjusting p^l to keep incentive compatibility is a profitable deviation whenever (49) holds. It is strictly profitable as the decrease in p^l will attract additional customers from the competing insurer. This contradicts that the original situation is an equilibrium and therefore (49) cannot hold.

Hence equation (47) has to be satisfied in every (L) equilibrium. *Q.E.D.*

Proof of proposition 4 If $q^h = 1$ and given that the equilibrium indifference curves have only one point of intersection (at q^l) and are continuous, equations (19) and (C1) imply that $p_q(q^l, u^h, \theta^h) > p_q(q^l, u^l, \theta^l)$. It follows then from equations (23) and (24) that a change in ρ which increases q^l raises u^h while reducing u^l. Given assumption 5, the other possibility is that u^l increases while u^h falls. But then equations (23) and (24) imply that q^l falls. Hence the latter case can be ruled out.

Similarly, if $q^l = 1$ and given that the equilibrium indifference curves are continuous and only have one point of intersection (at q^h), equations (20) and (C1) imply that $p_q(q^h, u^l, \theta^l) > p_q(q^h, u^h, \theta^h)$. It follows then from equation (27) that a change in ρ which raises q^h must be accompanied by an increase in u^l and a fall in u^h. *Q.E.D.*