Universität St.Gallen

Hidden Insurance in a Moral Hazard Economy

Giuseppe Bertola, Winfried Koeniger

November 2013 Discussion Paper no. 2013-37

School of Economics and Political Science,
Department of Economics

University of St. Gallen

Editor: Martina Flockerzi
University of St.Gallen
School of Economics and Political Science
Department of Economics
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35

Electronic Publication: http://www.seps.unisg.ch

Hidden Insurance in a Moral Hazard Economy[1]

Giuseppe Bertola, Winfried Koeniger[2]

Author's address:	Prof. Dr. Winfried Koeniger HSG-SEW Varnbüelstr. 14 CH-9000 St. Gallen Phone +41 71 2242309 Fax +41 71 2242302 Email winfried.koeniger@unisg.ch Website www.sew.unisg.ch/en

[1] This draft benefits from helpful comments by colleagues and seminar participants. Part of this research has been conducted while Giuseppe Bertola was at the Università di Torino and while Winfried Koeniger was at Queen Mary, University of London.

[2] Giuseppe Bertola: Edhec Business School, CEPR, CESifo, CFS. Winfried Koeniger: University of St.Gallen (Swiss Institute for Empirical Economic Research), CEPR, CFS, IZA.

Abstract

We consider an economy where individuals privately choose effort and trade competitively priced securities that pay off with effort-determined probability. We show that if insurance against a negative shock is sufficiently incomplete, then standard functional form restrictions ensure that individual objective functions are optimized by an effort and insurance combination that is unique and satisfies first- and second-order conditions. Modeling insurance incompleteness in terms of costly production of private insurance services, we characterize the constrained inefficiency arising in general equilibrium from competitive pricing of non-exclusive financial contracts.

Keywords

Hidden action; Principal agent; First-order approach; Constrained efficiency.

JEL Classification

E21, D81, D82.

1 Introduction

A recent literature has studied the implications of hidden savings for constrained-efficient consumption profiles when the probability distribution of income realizations depends on privately chosen effort (see Ábrahám, Koehne and Pavoni, 2011, and their references). Not only non-contingent savings, but also insurance is at least partly excluded from the public information set in reality, and a less recent and very influential literature established that a competitive insurance equilibrium is inefficient and may fail to exist in the presence of moral hazard. Inefficiency arises from the fact that the premium charged by competitive financial intermediaries, rather than by a single "exclusive" insurer covering each individual's risks, neglects the effect of insurance on effort unless the full extent of coverage is publicly observed (Pauly, 1974). Insurance against specific accident or health events, like consumer credit (Bizer and DeMarzo, 1992) and other contracts where default is possible, tries to address this problem with provisions that prevent multiple coverage of specific risks and market-level information pooling arrangements. Moral hazard inefficiencies cannot be eliminated completely, however, if portfolios of competitively traded contracts are only partially observable.

When competitively priced contingent securities pay off with a probability that depends on hidden actions, the equilibrium may fail to exist because standard functional form assumptions do not suffice to ensure that expected utility is a concave function of individual choices and that first-order optimality conditions are sufficient for optimality (Helpman and Laffont, 1975). As it may be optimal for price-taking individuals to jointly change insurance and effort choices, it is then not possible to ensure that insurance be priced consistently with effort-determined payoff probabilities.

In this paper, we analyze constrained efficiency and equilibrium existence in an economy where a continuous unobservable effort choice influences the probability of two possible income realizations and non-exclusive unit insurance contracts are competitively traded. We establish that if insurance is actuarially unfair enough then an equilibrium with positive effort can exist and be characterized by first-order conditions. Under plausible functional form restrictions, local concavity can be guaranteed by unfair insurance at all allocations where the first-order conditions hold. Standard continuity and differentiability properties of the model's primitive functions ensure that only one such allocation may exist and that, even though the objective function may be non-concave at allocations for which the first-order conditions are *not* satisfied, joint deviations in effort and insurance cannot be individually optimal.

This novel formal result illustrates more general insights, and is related to previous contributions. Moral hazard is one of the forms of asymmetric-information-based discretion on contractual deliveries for which Bisin et al. (2011) provide formal restrictions that ensure existence of equilibria, which are generically constrained inefficient in the presence of adverse selection. Mirrlees (1971), Cole and Kocherlakota (2001), and others let observable outcomes be a deterministic function of effort choices based on privately observed ability realizations, a source of adverse-selection problems. In that setting, as shown by Golosov and Tsyvinski (2007, appendix A), there can be no trade in non-exclusive contingent securities. In our model, as in Pauly (1974), Ábrahám and Pavoni (2005), and other moral-hazard settings, privately chosen actions determine a non-degenerate probability distribution for outcomes. Then, a non-exclusive market for trade in securities contingent on idiosyncratic realizations is active unless it is ruled out by assuming exclusive insurance.

We establish our result, in a single-period economy, by a technique similar to that applied by Ábrahám, Koehne and Pavoni (2011) to a two-period moral-hazard economy where non-contingent savings are hidden, but insurance is assumed to be exclusive. The resulting insights are related to those obtained by Helpman and Laffont (1975), Greenwald and Stiglitz (1986), and others in models where insurance is actuarially fair. However, we offer a novel perspective on the source and implications of the insurance incompleteness which, in previous work, plays a role in ensuring equilibrium existence for competitive non-exclusive asset trade under asymmetric information. Asset position limits or bid-ask spreads prevent deviations in Bisin and Gottardi (1999), and latent contracts support a competitive high-effort equilibrium in Bisin and Guaitoli's (2004) model with two effort levels (see also Hellwig 1983a,b). Loss and Piaser (2013) extend this analysis allowing for continuous effort, and show that non-concavity of the objective function may let individuals be indifferent between exerting an interior amount of effort (and insuring the loss partially) and setting the effort at the lowest possible level (and overinsuring). In equilibrium, the insurance price is fair for the probabilities implied by the low effort choice, but unfair for the probabilities implied by the higher effort choice. This makes it optimal to insure only partially if high effort is chosen. Partial insurance equilibria with positive profits are theoretically possible in such settings, because if any insurer deviated to offer a lower price, then individuals would discontinuously shift to low effort and overinsurance. Unprofitability of such deviations can sustain an equilibrium with positive profits, but may appear not particularly plausible (Hellwig, 1983b, p.4).

We argue that costs for insurance production are a more plausible source of incomplete insurance, and we show that a standard representation of the fixed and variable costs of insurance production can support a unique competitive free-entry equilibrium with partial insurance. Our technical result makes it possible to use the first-order conditions to characterize the nature of the equilibrium inefficiency implied by non-exclusive trade of unit insurance contracts in the presence of moral hazard. Extending the pure-exchange structure of the economies studied by earlier contributions, we explicitly model insurance production costs.[1] A non-exclusive competitive equilibrium where contingent securities are traded at actuarially unfair prices can exist, and is characterized by first-order conditions, when processing unit insurance contracts is costly. Transaction costs do not by themselves imply inefficiency, and the inefficiency arising from moral hazard can be addressed by linear taxation of anonymous insurance transactions or by public transfer schemes that use a transaction technology different from the one that produces private insurance contracts. On such issues, our contribution is related to those of Gottardi and Pavoni (2011) and of papers where first-order conditions characterize interactions between exclusive insurance and public policies (Golosov and Tsyvinsky, 2007; Chetty and Saez, 2010).[2]

In Section 2 we set up the structure of the economy and discuss how first-order conditions may characterize its equilibrium. We show in Section 3 that a given insurance price can be sufficiently unfair to imply that individually optimal choices are unique and are characterized by necessary and sufficient first-order conditions. Section 4 models the insurance price in terms of the cost structure of insurance production, Section 5 characterizes the resulting general equilibrium with moral hazard and hidden insurance, and Section 6 discusses how taxes or public transfers may address its inefficiency. Section 7 concludes.

[1]Ales and Maziero (2009) also model insurance production costs, but assume that each exclusive individual insurance contract entails a fixed cost, so that endogenous subsets of the population are perfectly insured or completely uninsured. Our assumptions instead imply that partial insurance is available to all of the economy's (representative) individuals.

[2]Krueger and Perri (2011) study similar issues in a setting where financial market imperfections arise from limited enforcement rather than from asymmetric information.

2 The problem

We adapt to our purposes the structure of a standard hidden-action problem. Ex-ante identical, risk-averse individuals experience idiosyncratic income shocks that are observable and verifiable but occur with probabilities that depend on unobservable effort. Denoting the intensity of a hidden effort action with e and realized consumption with c, individual welfare is additively separable in the disutility $-v(e)$ and the expected value of utility $u(c)$. Effort decreases welfare at an increasing rate,

[C1] $v'(e) > 0$, $v''(e) > 0 \, \forall \, e > 0$.

To focus on interior optima, we assume that $v'(0) = 0$ and that c can take any value in an open set, such as the real line, throughout which $u(\cdot)$ is strictly concave in consumption,[3]

[C2] $u'(c) > 0$, $u''(c) < 0 \, \forall \, c$.

We allow for two possible realizations of uncertainty over the single period that in the model represents the individuals' lifetime, denoting with $\Delta > 0$ the size of a possible negative shock. Individual resources in consumption terms amount to z with probability $1 - f(e)$, to $z - \Delta$ with probability $f(e)$. We assume that

[C3] $0 < f(e) < 1$, $f'(e) < 0,\ f''(e) > 0 \, \forall \, e > 0:$

the probability of each realization is bounded away from zero and one, and this "full-support" property makes it impossible to infer individual effort from realizations. The strictly negative first derivative implies that moral hazard is always relevant if effort is private information. The positive second derivative implies that effort encounters decreasing returns in reducing the probability of negative shocks, which remains bounded above zero for all effort levels.

The economy is populated by individuals who are ex-ante identical and sufficiently many as to make it unnecessary to draw a distinction between the probability and realized frequency of independently distributed shocks.[4] The realization of each individual's resources is publicly

[3]One might prefer explicitly to restrict consumption to be non-negative. The optimum would then be necessarily interior only if the marginal utility of consumption diverges to infinity at zero.

[4]The logic of our derivations readily extends to similar economies with symmetric information about heterogeneous factor endowments and/or preferences. In that case, however, Pareto-improving policies would need to include ex-ante transfers.

observed. Thus, it is possible to write and enforce private insurance contracts whereby payment of a premium p entitles to a unit of resources upon realization of the negative shock. Of course, $0 < p < 1$ whenever the insurance market is active (otherwise, insurance contracts would entail a sure loss for one of the parties). Each individual chooses to stipulate q such contracts, taking p as given. To represent in a stylized way realistic technological or legal constraints on collection and processing of detailed financial transaction information, we assume that each individual's q is not publicly observable. Hence, the distribution of realized consumption is influenced not only by unobservable effort, but also by similarly hidden insurance purchases.

From each individual's point of view, the relative price of consumption in the two possible contingencies is $p/(1-p)$. From our general equilibrium perspective, p determines the slope with respect to insurance of individual budget constraints, which also depend on contingent resource endowments and income from the production side of the economy. In the economy we model, endowments are parameterized by z and Δ, and the only production that will be of interest is that of insurance services. We allow the insurance production sector of the economy to generate rental income amounting to μ for each of the economy's ex-ante identical individuals. This income depends on equilibrium factor prices as in any competitive general equilibrium economy, in ways that we will not need to make explicit in what follows.

Under the contingent resource constraints

$$c_h = z - qp + \mu,\ c_l = z - \Delta + (1-p)q + \mu = c_h - \Delta + q, \tag{1}$$

the representative individual's problem is that of choosing q and e, taking p and μ as given, in such a way as to maximize

$$\mathcal{U}(q,e) = -v(e) + (1-f(e))u\left(z - qp + \mu\right) + f(e)u\left(z - \Delta + (1-p)q + \mu\right). \tag{2}$$

In equilibrium the insurance price p and income μ depend on the average effort $\bar{e}$, which determines the probability of claims, and on the average volume $\bar{q}$ of insurance transactions, which is also relevant to general equilibrium prices and incomes if production of insurance services is costly.[5] While $p\left(\bar{e},\bar{q}\right)$ and $\mu\left(\bar{e},\bar{q}\right)$ are taken as given in the individual optimization

[5]To see this, it may be useful to inspect the economy's resource constraint when insurance production is not costly, and aggregate resources equal aggregate consumption at $(1-f(\bar{e}))\,z + (z-\Delta)\,f(\bar{e}) = (1-f(\bar{e}))c_h + f(\bar{e})c_l$. With (1), this implies that $(p - f(\bar{e}))\,\bar{q} = \mu$, and if actuarially fair insurance is priced at $p = f(\bar{e})$ it must be the case that $\mu = 0$. In the more general framework we consider, insurance production

problem

$$\left(e^{\dagger}, q^{\dagger}\right) \in \arg\max_{e,q} \mathcal{U}(q, e), \tag{3}$$

in equilibrium it must be the case that $(\bar{e}, \bar{q}) = \left(e^{\dagger}, q^{\dagger}\right)$.

Under the differentiability and monotonicity assumptions in **[C1-C3]**, any optimal choice of effort and insurance must satisfy first-order conditions with respect to effort and to insurance: when $e = e^{\dagger}$, $c_h = z - q^{\dagger}p + \mu$, and $c_l = c_h - \Delta + q^{\dagger}$, then

$$f'(e)\left(u(c_l) - u(c_h)\right) = v'(e) \tag{4}$$

and

$$(1 - f(e))u'\left(c_h\right)p = f(e)u'\left(c_l\right)(1 - p). \tag{5}$$

Recalling that $c_l = c_h - \Delta + q$, actuarially unfair insurance ($f(e) < p < 1$) implies that optimal insurance is positive but partial ($0 < q < \Delta$).

Under additional conditions extensively discussed below, (4) and (5) are sufficient as well as necessary for individual maximization of (2) subject to (1). Then, they can replace (3) as constraints of the social welfare maximization problem that satisfies equilibrium and individual optimality conditions when choosing the representative individual's consumption profile to maximize (2).

The functional form assumptions **[C1-C3]** and the first-order conditions define a continuously differentiable mapping between the individual choice variables and the equilibrium values of p and μ. If an incentive-compatible equilibrium exists, and differentiable functions $e(\bar{q})$, $p(\bar{q}, e(\bar{q}))$, and $\mu(\bar{q}, e(\bar{q}))$ relate the representative individual's optimal effort, the price of insurance, and rents to the aggregate insurance amount $\bar{q}$, then the problem

$$\begin{aligned} &\max_{\bar{q}} \; -v(e(\bar{q})) + (1 - f(e(\bar{q})))\, u\left(c_h\right) + f(e(\bar{q}))u\left(c_l\right) \\ &\text{s.t. } c_h = z - qp\left(\bar{q}, \bar{e}\right) + \mu\left(\bar{q}, \bar{e}\right),\; c_l = c_h - \Delta + q \end{aligned}$$

has first-order condition

$$\left(-v'(e(\bar{q})) - f'(e(\bar{q})\left(u\left(c_h\right) - u\left(c_l\right)\right)\right)\frac{de}{d\bar{q}} + (1 - f(e(\bar{q})))\, u'\left(c_h\right)\frac{dc_h}{d\bar{q}} + f(e(\bar{q}))u'\left(c_l\right)\left(1 + \frac{dc_h}{d\bar{q}}\right) = 0$$

costs imply that $p > f(\bar{e})$, need to be accounted for in the economy's resource constraint, and may imply that $\mu > 0$.

or, using the optimality condition (4) for effort choice at given insurance,

$$-(1-f(e(\bar{q}))u'(c_h^*)\,\frac{dc_h}{d\bar{q}} = f(e(\bar{q}))u'(c_l^*)\left(1+\frac{dc_h}{d\bar{q}}\right), \tag{6}$$

where c_h^* and c_l^* denote constrained-efficient consumption levels.

Since effort is not publicly observed and influences the probability distribution of resources, the social maximization problem cannot achieve the first-best consumption allocation. Less obviously, the economy's competitive equilibrium also generally fails to achieve the best incentive-compatible and physically feasible combination of consumption smoothing and effort, because competitive trade of insurance securities does not account for the effect of hidden insurance on effort incentives and aggregate production. By the envelope theorem the variation of effort caused by changes of insurance has no first-order welfare effect through $v(\cdot)$ and $f(\cdot)$ but, since

$$\frac{dc_h}{d\bar{q}} = -p(\bar{q},e(\bar{q})) - \bar{q}\frac{dp(\bar{q},e(\bar{q}))}{d\bar{q}} + \frac{d\mu(\bar{q},e(\bar{q}))}{d\bar{q}}, \tag{7}$$

the social efficiency condition (6) differs from the individual first-order condition (5) whenever $\bar{q}$ influences consumption levels not only through individually optimal purchases, priced at $p(\bar{q},e(\bar{q}))$, but also through changes of that equilibrium price and of income. It is intuitive, and will be shown formally below, that no such effect would be present in competitive equilibrium if the probability distribution of insurable shocks did not depend on unobservable effort; when atomistic price-taking behavior fails to internalize the welfare implications of $f'(e) < 0$, conversely, then individually optimal consumption profiles are inefficient.

The first-order approach leading to characterization of these distortions as in (7) is very convenient. It is only valid, however, if hidden actions satisfy first-order conditions and a competitive equilibrium exists where endogenous variables are linked to insurance volumes by the well-behaved functions appearing in the social maximization problem. In what follows we first proceed to derive conditions which make it appropriate to characterize individual choices in terms of first-order conditions. Then, we show how a stylized model of insurance production lets structural features determine the equilibrium functions appearing in (6) and (7).

3 Validity of the first-order approach

For the first-order conditions (4) and (5) to be sufficient, it has to be the case not only that $\partial^2\mathcal{U}(q,e)/\partial q^2 \leq 0$ and $\partial^2\mathcal{U}(q,e)/\partial e^2 \leq 0$, or

$$u''(c_l)\, f(e)(1-p)^2 + u''(c_h)\,(1-f(e))\,p^2 \leq 0, \quad f''(e)\,(u(c_l)-u(c_h)) - v''(e) \leq 0, \tag{8}$$

but also that $\left(\partial^2\mathcal{U}(q,e)/\partial q^2\right)\left(\partial^2\mathcal{U}(q,e)/\partial e^2\right) - \left(\partial^2\mathcal{U}(q,e)/\partial q\partial e\right)^2 \geq 0$ or, using (8),

$$-\frac{u''(c_l)\, f(e)(1-p)^2 + u''(c_h)\,(1-f(e))\,p^2}{f'(e)\,((1-p)\,u'(c_l) + pu'(c_h))} \leq -\frac{f'(e)\,((1-p)u'(c_l) + pu'(c_h))}{f''(e)\,(u(c_l)-u(c_h)) - v''(e)}. \tag{9}$$

While the inequalities in (8) follow from assumptions **[C1]**, **[C2]**, **[C3]** for all $c_l \leq c_h$, the cross-derivative sign restriction in (9) does not (Helpman and Laffont, 1975). As $f'(e) \neq 0$ and effort determines the probability distribution of consumption, our assumptions do not imply that expected utility (2) is a globally concave function of both consumption and effort. When the local concavity condition (9) fails to hold at a consumption and effort combination that satisfies the first order conditions, those conditions cannot identify a competitive equilibrium, because a joint variation of insurance and effort locally increases welfare.

To identify plausible and interpretable restrictions that ensure validity of the first-order approach, we proceed in steps. First, we derive a condition that ensures local concavity of (2) when consumption responds sufficiently strongly to the model's insurable shock. We proceed to derive conditions for consumption to respond strongly, and for a locally robust equilibrium to exist. We will then shown that, even though the objective function may well be non-concave at allocations where the first-order conditions are not satisfied, non-local deviations in effort and insurance cannot, under interpretable conditions, be welfare improving.

As a starting point, we formulate a sufficient condition for local concavity of the objective function:

Result 1 *Given [C1], [C2], [C3], if $c_h > c_l$ and*

$$\frac{p^2u''(c_h) - (1-p)^2u''(c_l)}{\left(pu'(c_h) + (1-p)u'(c_l)\right)^2}\,(u(c_h) - u(c_l)) \geq \frac{f'(e)^2}{f''(e)f(e)} \tag{10}$$

then the expected utility loss $-f(e)\,(u\,(c_h) - u\,(c_l))$ is locally concave, and this is sufficient to ensure local concavity of the objective function $\mathcal{U}(q,e)$ in (2).

Proof. Writing $\mathcal{U}(q,e) = -v(e) + u\,(c_h) - f(e)\,(u\,(c_h) - u\,(c_l))$ and noting that $-v(e)$ is concave in e by **[C1]** , $u(c_h)$ is concave in q since $\partial^2 u\,(c_h)\,/\partial q^2 = u''\,(c_h)\,p^2 < 0$ by **[C2]**, and sums of concave functions are concave, we only need to prove that the conditions imply that $-f(e)\left(u\left(c_h\right)-u\left(c_l\right)\right)$ is concave or, equivalently, $f(e)\left(u\left(c_h\right)-u\left(c_l\right)\right) = f(e)\left(u(z-qp+\mu)-u(z-\Delta+(1-p)q+\mu)\right)$ is convex in q and e. For $c_h > c_l$, the monotone first derivative assumption in **[C2]** ensures that $u\left(c_h\right)-u\left(c_l\right)>0$, and therefore that the inequality in (10), where the right-hand side is positive by **[C3]**, implies $u''(c_h)p^2-u''(c_l)(1-p)^2>0$. Hence, both diagonal terms of the Hessian

$$H=\left[\begin{array}{cc} f(e)\left(u''(c_h)p^2-u''(c_l)(1-p)^2\right) & f'(e)\left(-u'(c_h)p-u'(c_l)(1-p)\right) \\ f'(e)\left(-u'(c_h)p-u'(c_l)(1-p)\right) & f''(e)\left(u\left(c_h\right)-u\left(c_l\right)\right) \end{array}\right],$$

are positive. The determinant

$$|H|=f(e)\left(u''(c_h)p^2-u''(c_l)(1-p)^2\right)f''(e)\left(u\left(c_h\right)-u\left(c_l\right)\right)-\left(f'(e)\left(-u'(c_h)p-u'(c_l)(1-p)\right)\right)^2$$

is also positive if (10) holds. ■

This result provides sufficient but not necessary conditions for local concavity.[6] The conditions are violated when insurance is perfect, and consumption is perfectly stable. With $c_h = c_l$ the expected utility loss is zero, but need not be concave if moral hazard is present: when the left-hand side of (10) is zero, the inequality is violated since **[C3]** implies that the right-hand side is strictly positive. When $c_h > c_l$ instead, and an adverse realization reduces utility, then the left-hand side of (10) differs from zero and is positive if $u''(c_h)p^2-u''(c_l)(1-p)^2>0$, i.e. if $u''(c_h)$ is less negative than $u''(c_l)$ by a factor that depends on the insurance premium p. Since concavity fails to hold when joint deviations improve welfare, it is intuitive that decreasing risk aversion should, by reducing the appeal of deviations that vary effort and insurance in opposite

[6]To see that the function is necessarily concave when the conditions of Result 1 are satisfied, rearrange (9), using $f'(e)<0$ and $u(c_h)-u(c_l)+v''(e)/f''(e)>0$, to obtain

$$\frac{p^2u''(c_h)-(1-p)^2u''\left(c_l\right)-\left(p^2/f(e)\right)u''(c_h)}{\left(pu'(c_h)+(1-p)\,u'(c_l)\right)^2}\left(u(c_h)-u(c_l)+\frac{v''(e)}{f''(e)}\right)\geq\frac{f'(e)^2}{f''(e)f(e)}:$$

the left-hand side of this inequality is larger than that of (10) since $v''(e)/f''(e) > 0$ and $-\left(p^2/f(e)\right)u''(c_h)>0$. To see that (10) is not necessary, note that (9) may be satisfied even for complete insurance if it so happens that

$$-\frac{u''(c)}{u'(c)^2}\geq\frac{f'(e)^2}{f''(e)f(e)}\left(\left((1-p)^2+\frac{1-f(e)}{f(e)}p^2\right)\frac{v''(e)}{f''(e)}\right)^{-1},$$

a condition which depends on consumption and effort levels through all of the model's functions, and could only be verified by evaluating them explicitly.

directions, play a role in ensuring that the inequality in (10) is satisfied.

To exploit this insight formally it will be convenient to let the utility function take the hyperbolic absolute risk aversion (HARA) form

[C2'] $u(c) = \zeta\left(\eta + \frac{c}{\sigma}\right)^{1-\sigma}$, with $(1-\sigma)\zeta > 0$.

This makes it possible to write the function of consumption levels on the left-hand side of (10) as a function of the marginal-utility ratio:[7]

$$\frac{\sigma}{\sigma-1}\frac{p^2\frac{u''(c_h)}{u''(c_l)} - (1-p)^2}{\left(p\frac{u'(c_h)}{u'(c_l)} + (1-p)\right)^2}\left(\frac{u(c_h)}{u(c_l)} - 1\right) \equiv g\left(\frac{u'(c_l)}{u'(c_h)}\right). \tag{11}$$

The following properties of this function will be useful in proving our results:

Result 2 *For the function*

$$g(\nu) \equiv \frac{p^2\nu^{-\frac{1+\sigma}{\sigma}} - (1-p)^2}{\left(p\nu^{-1} + (1-p)\right)^2}\frac{\sigma}{\sigma-1}\left(\nu^{\frac{1-\sigma}{\sigma}} - 1\right) \tag{12}$$

(i) $g'(\nu) > 0$	*for* $\nu > 1$ *and* $g(\nu) > 0$	*if* $\sigma > 0$,
(ii) $g(\nu) < 1$	*for all* $\nu > 0$	*if* $\sigma = -1 < 0$ *or* $\sigma \to \infty$,
(iii) $\lim_{\nu\to\infty} g(\nu) > 1$		*if* $0 < \sigma < \infty$.

[7]To see this, note that

$$\frac{p^2u''(c_h) - (1-p)^2u''(c_l)}{\left(pu'(c_h) + (1-p)u'(c_l)\right)^2}\left(u(c_h) - u(c_l)\right) = \frac{u''(c_l)u(c_l)}{(u'(c_l))^2}\frac{p^2\frac{u''(c_h)}{u''(c_l)} - (1-p)^2}{\left(p\frac{u'(c_h)}{u'(c_l)} + (1-p)\right)^2}\left(\frac{u(c_h)}{u(c_l)} - 1\right)$$

and that for HARA utility

$$\begin{aligned}\frac{u(c_h)}{u(c_l)} &= \left(\frac{\left(\eta + \frac{c_h}{\sigma}\right)^{-\sigma}}{\left(\eta + \frac{c_l}{\sigma}\right)^{-\sigma}}\right)^{-\frac{1-\sigma}{\sigma}} = \left(\frac{u'(c_l)}{u'(c_h)}\right)^{\frac{1-\sigma}{\sigma}}, \qquad \frac{u''(c_h)}{u''(c_l)} = \left(\frac{u'(c_l)}{u'(c_h)}\right)^{-\frac{1+\sigma}{\sigma}}, \\ \frac{u''(c)u(c)}{(u'(c))^2} &= \frac{-\zeta\frac{1-\sigma}{\sigma}\left(\eta + \frac{c}{\sigma}\right)^{-\sigma-1}\zeta\left(\eta + \frac{c}{\sigma}\right)^{1-\sigma}}{\left(\zeta\frac{1-\sigma}{\sigma}\left(\eta + \frac{c}{\sigma}\right)^{-\sigma}\right)^2} = \frac{\sigma}{\sigma-1}\end{aligned}$$

if $\sigma \neq 1$. In the $\sigma = 1$ logarithmic utility case, $u''(c_l)u(c_l)/u'(c_l)^2 = -\ln c_l$ and the function defined in (11) takes the form

$$g(\nu) = -\frac{p^2\nu^{-2} - (1-p)^2}{\left(p\nu^{-1} + (1-p)\right)^2}\ln\nu,$$

with $\nu \equiv u'(c_l)/u'(c_h)$, and has the same properties as other HARA functions with $0 < \sigma < \infty$.

Proof. (i) Writing $g(\nu) = \hat{g}(\nu)\frac{\sigma}{\sigma-1}\left(\nu^{\frac{1-\sigma}{\sigma}} - 1\right)$ for $\hat{g}(\nu) = \frac{p^2\nu^{-\frac{1+\sigma}{\sigma}} - (1-p)^2}{\left(p\nu^{-1}+(1-p)\right)^2}$, and noting that $\hat{g}'(\nu) = \hat{g}(\nu)\left(-\frac{1+\sigma}{\sigma}\left(\frac{p^2\nu^{-\frac{1}{\sigma}-2}}{p^2\nu^{-\frac{1+\sigma}{\sigma}}-(1-p)^2}\right) + 2\frac{p\nu^{-2}}{\left(p\nu^{-1}+(1-p)\right)}\right)$,

$$\begin{aligned} g'(\nu) &= \hat{g}'(\nu)\frac{\sigma}{\sigma-1}\left(\nu^{\frac{1-\sigma}{\sigma}} - 1\right) - \hat{g}(\nu)\nu^{\frac{1-\sigma}{\sigma}-1} \\ &= g(\nu)\left(-\frac{1+\sigma}{\sigma}\frac{p^2\nu^{-\left(\frac{1}{\sigma}+2\right)}}{p^2\nu^{-\frac{1+\sigma}{\sigma}} - (1-p)^2} + 2\frac{p\nu^{-2}}{(p\nu^{-1}+(1-p))} + \frac{1-\sigma}{\sigma}\frac{\nu^{\frac{1-\sigma}{\sigma}-1}}{\nu^{\frac{1-\sigma}{\sigma}}-1}\right) \end{aligned}$$

is positive when $g(\nu) > 0$ implies $p^2\nu^{-\frac{1+\sigma}{\sigma}} - (1-p)^2 < 0$, and $\nu > 1$ and $\sigma > 0$ imply $\frac{\sigma}{1-\sigma}\left(\nu^{\frac{1-\sigma}{\sigma}} - 1\right) > 0$.
(ii) For $\sigma = -1$, we have

$$g(\nu) = \frac{1}{2}\frac{(2p-1)\left(\nu^{-2}-1\right)}{\left(\frac{p}{\nu}+(1-p)\right)^2},$$

which has a global maximum of 0.5 at $\nu = (p-1)/p < 0$, and never exceeds unity. For $\sigma \to \infty$ we have

$$\lim_{\sigma\to\infty} g(\nu) = \frac{p^2\nu^{-1} - (1-p)^2}{\left(p\nu^{-1}+(1-p)\right)^2}\left(\nu^{-1}-1\right) = 1 - \frac{1}{\left(p\nu^{-1}+(1-p)\right)^2\nu} < 1\,\forall\nu > 0.$$

These expressions establish that

$$\lim_{\nu\to\infty} g(\nu) = \begin{cases} \frac{1}{2}\frac{1-2p}{(1-p)^2} < 0.5 & \text{for } \sigma = -1, \\ 1 - \frac{1}{\left(p\nu^{-1}+(1-p)\right)^2\nu} < 1 & \text{for } \sigma \to \infty, \end{cases}$$

while
(iii)

$$\lim_{\nu\to\infty} g(\nu) = \begin{cases} \infty & \text{for } 0 < \sigma \leq 1, \\ \frac{\sigma}{\sigma-1} > 1 & \text{for } 1 < \sigma < \infty. \end{cases}$$

■

The behavior of $g(\cdot)$ as ν grows above unity depends on σ, the parameter that determines whether absolute risk aversion increases or decreases with consumption.[8] As the optimal amount of insurance becomes more partial and ν increases further above unity, for $\sigma > 0$ the left-hand side of inequality (10) becomes more positive (if it already is). To ensure local concavity at the optimum, it should exceed the positive right hand side of that condition. The right-hand side is bounded below unity if we strengthen **[C3]** to

[C3'] $0 < f(e) < 1$, $f'(e) < 0$, and $f(e)$ is log-convex: $f''(e)f(e)f'(e)^{-2} \geq 1$,

[8]Absolute risk aversion $-u''(c)/u'(c) = \left(\eta + \frac{c}{\sigma}\right)^{-1}$ is a decreasing function of c when $0 < \sigma < \infty$; it is constant when $\sigma \to \infty$ yields $u(c) = -\eta\exp(c/\eta)$, and increasing if, for example, $\sigma = -1$ yields quadratic utility (see e.g. Gollier, 2001).

to imply that higher effort encounters strongly diminishing returns in reducing the probability of adverse realizations.

Result 2 shows that $g(\cdot)$ cannot exceed unity if $\sigma \to \infty$ implies constant absolute risk aversion, or if $\sigma = -1$ so that utility is quadratic and absolute risk aversion is increasing. When $0 < \sigma < \infty$ implies decreasing risk aversion, conversely, it is possible for insurance to be so incomplete that individual first-order conditions can only hold where the objective function is locally concave, and that a unique consumption and effort combination solves the maximization problem. To see this, note that when the first-order condition (5) for insurance holds, the argument of $g(\cdot)$ is

$$\frac{u'(c_l)}{u'(c_h)} = \frac{1-f(e)}{f(e)}\frac{p}{1-p} \equiv \nu(e,p). \tag{13}$$

Since $\nu(e,p) = 1$ when $p = f(e)$ and $g(1) = 0$ for all admissible values of σ, using (11) and (13) in (10) confirms that (10) is violated under full insurance. To establish that actuarially unfair insurance can ensure local concavity at all allocations that satisfy the first-order condition for insurance, note that the expression in (13) exceeds unity when $f(e) < p$ and that, since $0 < f(e) < 1$ and $0 < p < 1$, it is increasing in e for given p when $f'(e) < 0$ as in **[C3]** implies moral hazard. Intuitively, higher effort decreases the probability of loss and makes it optimal to buy less insurance at an interior optimum, so that a loss realization has larger implications for marginal utility. We then can show:

Proposition 1 *Given **[C1], [C2'], [C3']**, if absolute risk aversion is decreasing then p can exceed $f(0)$ by a sufficiently large amount to ensure that the objective function (2) is concave for all consumption patterns that satisfy interior first-order conditions.*

Proof. The right-hand side of the local concavity condition (10) is bounded below unity by **[C3']** for all e. Its left-hand side is bounded above unity at $e = 0$ by the first-order condition for insurance choice (13) if

$$g\left(\frac{1-f(0)}{f(0)}\frac{p}{1-p}\right) \geq 1, \tag{14}$$

where as defined in (12)

$$g(\nu) \equiv \frac{p^2\nu^{-\frac{1+\sigma}{\sigma}} - (1-p)^2}{(p\nu^{-1} + (1-p))^2}\frac{\sigma}{\sigma-1}\left(\nu^{\frac{1-\sigma}{\sigma}} - 1\right).$$

By Result 2(iii), condition (14) can be satisfied for a finite value of the $g(\cdot)$ function's argument if $0 < \sigma < \infty$ and risk aversion is decreasing. Since by Result 2(i) for $\sigma > 0$ we have that $g'(\cdot) > 0$ when $g(\nu) > 0$ and $\nu > 1$, and the right-hand side of (13) increases with e for $f'(e) < 0$ as in **[C3']**,

$$g\left(\frac{1-f(e)}{f(e)}\frac{p}{1-p}\right) > g\left(\frac{1-f(0)}{f(0)}\frac{p}{1-p}\right) \geq 1 \text{ for all } e > 0. \tag{15}$$

Hence, (14) ensures that the inequality in (10) is satisfied for all $e > 0$ when (13) holds. By Result 1 this is sufficient to ensure local concavity at all such points. This can also be proved noting that the left-hand side of (15) exceeds 1 if $u(\cdot)$ has HARA form and $u(c_h) - u(c_l)$ is log-convex, that log-convexity is preserved by multiplication, and that convexity of the potentially problematic $f(e)(u(c_h) - u(c_l))$ expected utility loss ensures concavity of the individual objective function. ■

Proposition 1 does not establish global concavity of the objective function, nor its quasi-concavity. Its conditions only restrict the shape of the expected utility loss term $f(e)(u(c_h) - u(c_l)$, as the shape of the effort cost function $v(\cdot)$ conveniently plays no role in the sufficient local concavity condition (10) of Result 1. They include not only the plausible and familiar curvature assumptions **[C2']** and **[C3']** and decreasing risk aversion, but also the requirement that the probability $f(0)$ of loss when there is no effort be sufficiently smaller than the insurance price p. The proof first uses Result 1 to show that when the insurance price is unfair and (13) holds, the ratio $\nu(e,p)$ of the marginal utilities in the two states can be large enough to ensure local concavity at $e = 0$. With $v'(0) = 0$, optimal effort cannot be zero when the price of insurance is actuarially unfair, and optimal insurance is only partial. But at all points where the insurance first order condition (13) is satisfied, the marginal-utility ratio $\nu(e,p)$ is larger for $e > 0$ than for $e = 0$, and with decreasing risk aversion Result 2 ensures that the objective function must be locally concave at all such points.

When Proposition 1's assumptions are satisfied, the solution of the individual maximization problem is found at an interior point and is characterized by first-order conditions. The objective function may be non-concave elsewhere, but this is not problematic for the first-order approach. To see this note that since **[C1-3]** implies continuity of individual choices, the objec-

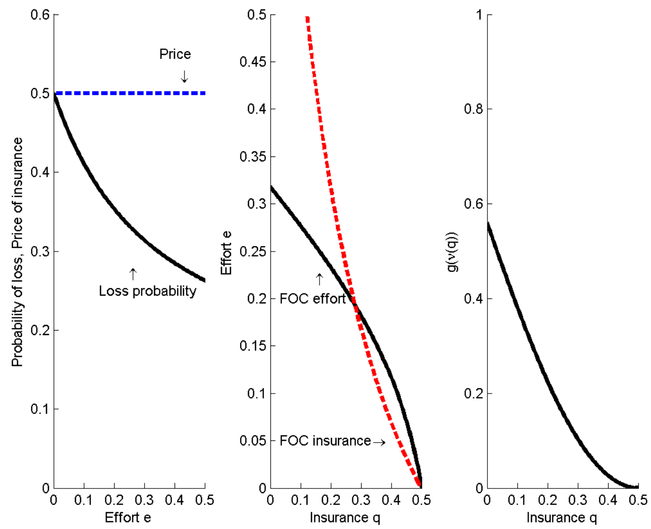

Figure 1: Invalidity of the first-order approach if conditions of Proposition 1 are violated. Source: Authors' calculation. Notes: We make the following parametric assumptions: $u(c) = (c^{1-\sigma} - 1)/(1-\sigma)$, $v(e) = (1/\kappa)e^{\kappa}$, $f(e) = \zeta \exp(-\gamma^2 e/(1+\gamma e))$. The parameter values are $\sigma = 2$, $\kappa = 1.05$, $\zeta = 0.5$, $\gamma = 1.5$, the endowment is $z + \mu = 1$ and the size of the loss is $\Delta = 0.5$.

tive function cannot have multiple interior local optima unless local minima or saddle-points also exist where the first-order conditions hold at non-concave points of the objective function. But if the conditions of Proposition 1 hold, then the first-order conditions can only be satisfied at points where the objective function is locally concave. Thus, there can only be a single local maximum, which must coincide with the global maximum since assumptions **[C1-3]** rule out corner maxima where first-order conditions do not hold.

It is possible to specify conditions that are less stringent, but also less general and transparent than the sufficient condition $g(\nu) \geq 1$ in Proposition 1. It may be of interest to note that when absolute risk aversion is constant then $g(\cdot)$ is monotonic and the derivations in Result 2(ii) establish that, while strictly less than unitary for $\nu > 1$, $g(\cdot)$ is positive if $p < 0.5$, and may exceed $f'(e)^2/(f''(e)f(e))$ if additional functional form restrictions ensure that the latter is strictly less than unitary and suitably small.[9] More generally, the contingent price p need not exceed $f(0)$ so

[9]Functions in the form $f(e) = \zeta \exp(-\xi(e))$ have the properties listed in **[C3']** if $\zeta \exp(-\xi(0)) < 1$, $\lim_{e\to\infty} \zeta \exp(-\xi(e)) > 0$, $\xi'(e) > 0$ and $\xi''(e) \leq 0$ for all $0 \leq e$. While $f'(e)^2/(f''(e)f(e)) = 1$ when $\xi(e)$

much as to satisfy the sufficient condition (14) if $v(e)$ is strongly convex at low levels of e.

Figure 1 illustrates how a violation of Proposition 1's sufficient conditions may make it inappropriate to rely on the first-order condition, under parametric assumptions that do not allow other sources of concavity to compensate for that violation. The left panel shows that effort reduces the probability of a loss, which is the source of the moral hazard problem, and that the price of insurance is fixed at $p = f(0)$, hence it is actuarially fair if the agent exerts no effort. The right panel of the figure plots the function $g(\nu)$, which does not depend on effort and should be larger than unity as a condition of Proposition 1, but in this parametric example is lower than unity for all $q \in [0, \Delta]$. Thus, concavity of the objective function is not guaranteed for combinations of effort and insurance which satisfy the first-order conditions. Indeed, as shown in the middle panel, the first-order conditions are satisfied for two different combinations of effort and insurance in Figure 1. The loci that satisfy the insurance and effort first-order conditions cross at an interior partial insurance point, but also at $e = 0$ and $q = 0.5$: this is on the border of the choice set, but is not ruled out by $v'(0) = 0$ when insurance is actuarially fair. In the numerical example, the objective function's maximum is at the latter point. That relatively uninteresting zero-effort, full-insurance competitive equilibrium, however, cannot exist if actuarially fair insurance is not available, as is realistic, and may be rationalized by insurance production costs.

4 Production of insurance services

Having established that price-taking behavior can support a competitive equilibrium with sufficiently incomplete insurance, we proceed to consider how partial insurance may in turn be implied in equilibrium by costly production of insurance services.

We do not allow non-intermediated trade in individual-specific securities. As individuals

is linear, other functional forms can imply a lower upper bound: for example, if $\xi(e) = \gamma^2 e/(1+\gamma e)$ then $f'(0)^2/(f''(0)f(0)) = \gamma/(\gamma+1)$ at zero effort, and declines to zero as effort diverges to infinity.

can only take long positions in contingent securities, the conditions derived above for local concavity of their objective function only need to be satisfied at first-order conditions with $0 < q$.[10]

We model a standard competitive industry of intermediaries that diversify idiosyncratic risk by holding a representative portfolio of non-exclusive unit insurance contracts, and charge actuarially unfair premia because issuing such contracts and/or processing the resulting claims is costly. Any positive and weakly increasing intermediation cost function may ensure that insurance is sufficiently incomplete to fulfill the key condition of Proposition 1. For our purpose of characterizing equilibria by inspection and manipulation of interior first-order conditions, however, it is convenient to make assumptions that ensure that optimal insurance is positive but partial ($0 < q < \Delta$, $f(e) < p < 1$ for any $e > 0$). To this end, we suppose that the marginal cost $m(x)$ of processing x individual contingent claims is differentiable, increasing, and unitary at $x = 0$:

[C4] $\quad m(0) = 1, \quad m(x) > 0 \, \forall x, \quad m'(x) > 0 \, \forall x,$

so that the average cost of processing claims and payments depends on their amount according to

$$\begin{aligned} a(x) &\equiv \frac{\int_0^x m(y)dy}{x}, \quad a(0) = 1, \\ a'(x) &= (m(x) - a(x)) \frac{1}{x}, \quad 0 < a'(x) < m'(x). \end{aligned} \tag{16}$$

The marginal and average unit cost of contingent payouts are unitary at $x = 0$ and smoothly grow above unity for $x > 0$. While infinitesimally small insurance would allow intermediaries to issue actuarially fair contracts, the cost of processing strictly positive amounts of contingent securities exceeds unity: this represents realistic processing or verification costs, and implies

[10] To see the relationship between our economy's and Bisin and Gottardi's (1999) equilibrium existence conditions it may also be helpful to note that, if short positions in contingent assets were allowed, the functional form assumption we make below would imply strictly positive bid-ask spreads.

that equilibrium insurance premia are actuarially unfair.

Idiosyncratic uncertainty cancels out in the customer base of intermediaries who trade with a large number of individuals. Then, profit maximization implies that the competitive price $p(\cdot)$ of a contract for delivery of a unit of goods upon realization of the negative shock should equal its expected marginal cost. In a symmetric competitive equilibrium with N insurance intermediaries, where the probability of payoff is $f(\bar{e})$ and each intermediary is issuing $x = \bar{q}/N$ contracts, this is

$$p(\bar{q}, \bar{e}) = f(\bar{e}) m\left(\bar{q}/N\right). \tag{17}$$

By **[C4]**, $p(\bar{q}, \bar{e}) > f(\bar{e})$ if $\bar{q}/N > 0$, i.e., if each intermediary operates at a finitely large scale. To ensure this, and to rule out non-intermediated trade, we suppose that organization of risk-sharing entails a fixed $\phi > 0$ cost. Then, the unit insurance premium exceeds the average expected cost $f(\bar{e}) a\left(\bar{q}/N\right)$, and each intermediary's revenues exceed costs by

$$\pi\left(\frac{\bar{q}}{N}, \bar{e}\right) = f(\bar{e}) \left(m\left(\frac{\bar{q}}{N}\right) - a\left(\frac{\bar{q}}{N}\right)\right) \frac{\bar{q}}{N} - \phi. \tag{18}$$

Since $\partial \pi(x, \bar{e})/\partial x = f(\bar{e}) x m'(x) > 0$ by (16) and $\pi(0, \bar{e}) = -\phi < 0$, the free entry condition $\pi(\bar{q}/N, \bar{e}) = 0$ uniquely determines $\bar{q}/N$ for each $\bar{e}$; as N adjusts to ensure that insurance services are produced at the zero-profit efficient scale for every $\bar{q}$, returns to scale are constant. If the per capita number of intermediaries N is given instead, then each of N firms pays rents amounting to (18) to the economy's representative individuals, and returns to scale are decreasing at the level of the aggregate insurance industry.

It is not necessary for our purposes to spell out the microeconomic structure that underlies the cost functions introduced in this section. We measure insurance production costs in terms of consumable resources but, as long as no additional asymmetric-information issues arise in production, it would be simple (and trivial if leisure is perfectly substitutable to consumption in individual utility functions) to express fixed and variable insurance services in terms of labor. The decreasing returns to scale implied by a parametrically given number of firms can be explained in fully standard fashion by use in production of insurance services of a scarce factor, such as managerial ability, that is owned by the economy's representative individuals and is compensated by the amount that, in equilibrium, corresponds to (18).

For our general equilibrium analysis it is instead crucial to characterize the partial derivatives with respect to $\bar{q}$ and $\bar{e}$ of the total cost of insurance production by the N intermediaries,

$$k(\bar{q}, \bar{e}) \equiv f(\bar{e}) a\left(\bar{q}/N\right) \bar{q} + \phi N. \tag{19}$$

It is simple to verify using (16) that, regardless of whether N is fixed or varies to keep $\pi(\bar{q}/N, \bar{e}) = 0$,

$$\frac{\partial}{\partial q} k(\bar{q}, \bar{e}) = f(\bar{e}) m\left(\bar{q}/N\right) = p(\bar{q}, \bar{e}) > 0, \quad \frac{\partial}{\partial e} k(\bar{q}, \bar{e}) = f'(\bar{e}) a\left(\bar{q}/N\right) \bar{q} < 0: \tag{20}$$

the marginal cost of more insurance at given effort corresponds to the insurance premium; more effort at given insurance reduces the incidence of negative shocks and the total cost $a\left(\bar{q}/N\right)\bar{q}$ of insurance services.

5 Hidden insurance in general equilibrium

We proceed to define and study the general equilibrium implied by the insurance demand and supply relationships introduced and characterized above.

Definition 1 *In competitive equilibrium,*

(i) effort $e^\dagger$ and insurance $q^\dagger$ solve the individual maximization problem and satisfy (4), (5) for given $p\left(\bar{q}, \bar{e}\right)$, $\mu\left(\bar{q}, \bar{e}\right)$;

(ii) price-taking intermediaries maximize profits, so that $p = f(\bar{e}) m\left(\bar{q}/N\right)$ by (17) and, by (18), $\mu = f(\bar{e})(m\left(\bar{q}/N\right) - a\left(\bar{q}/N\right))\bar{q} - \phi N$, where N may either be parametrically given or ensure that $\mu = 0$ under free entry;

(iii) equilibrium effort and insurance coincide with individual choices, so that $\left(e^\dagger, q^\dagger\right) = \left(\bar{e}, \bar{q}\right)$.

Part (i) of the definition encompasses the requirement that the first-order approach is valid which, as discussed below, can be established under reasonable and interpretable conditions on the basis of Section 3's analytical results. Part (ii) recognizes that any difference between well-diversified insurers' revenues and costs should appear as μ in individual budget constraints, which ensures physical feasibility of the equilibrium. Aggregating (1), the economy's resource constraint reads

$$(1 - f(\bar{e}))\, z + f(\bar{e})\, (z - \Delta) - ((1 - f(\bar{e}))c_h + f(\bar{e})c_l) = \phi N + f(\bar{e})\, (a\, (\bar{q}/N) - 1)\, \bar{q}$$

in equilibrium, and appropriately accounts for the fact that aggregate resources should cover not only aggregate consumption, but also the cost of insurance services production.[11]

Inserting (17) and (18) in (1), it is also straightforward to verify that

$$c_h = z - (f(\bar{e})a\, (\bar{q}/N)\, \bar{q} + \phi N) = z - k(\bar{q}, \bar{e}) : \tag{21}$$

as in any competitive general equilibrium, insurance revenues are rebated to the representative individual, whose consumption level is reduced by production costs. This simple relationship is key to characterization of the derivatives in (7), and of the wedge that equilibrium effects introduce between the individual and social optimality conditions.

In an equilibrium where $e = \bar{e}$ and $q = \bar{q}$, the effort first-order condition (4) holds and, using (21), reads

$$f'(\bar{e})\, (u(z - \Delta + q - k(\bar{q}, \bar{e})) - u(z - k(\bar{q}, \bar{e})) = v'(\bar{e}); \tag{22}$$

the first-order condition for insurance (13) also holds and, using (17), reads

$$\frac{u'\, (z - \Delta + \bar{q} - k(\bar{q}, \bar{e}))}{u'\, (z - k(\bar{q}, \bar{e}))} = \frac{m\, (\bar{q}/N)\, (1 - f(\bar{e}))}{1 - f(\bar{e})m\, (\bar{q}/N)}. \tag{23}$$

[11] For readers familiar with the pooled securities notion introduced in Bisin and Gottardi (1999) it may be helpful to note that, in the tractable economy we study, μ similarly represents the returns earned by the aggregate counterparts of individual asset trades.

We can show that condition (23) and the functional form assumption **[C4]**, made in Section 4, imply that insurance is positive and partial at the equilibrium crossing point of demand and supply relationships:

Result 3 *For any $\Delta > 0$, [C4] implies that individually optimal equilibrium insurance $\bar{q}$ lies between 0 and Δ.*

Proof. Recalling that $c_l = c_h - \Delta + \bar{q}$, concave utility as in **[C2]** implies that the left-hand side of (23) is a monotonically decreasing function of q for any c_h. Its value is $u'(c_h - \Delta)/u'(c_h) > 1$ at $q = 0$, unity at $q = \Delta$. The right-hand side of (23) is unitary at $q = 0$ since $m(0) = 1$ by **[C4]**. It increases in $q > 0$, diverging to infinity as $p(\cdot) = f(e)m(q/N) \to 1$, with slope

$$\frac{d}{d(q/N)}\left[m(q/N)\frac{1-f(e)}{1-f(e)m(q/N)}\right] = m'(q/N)\frac{1-f(e)}{(1-f(e)m(q/N))^2} > 0$$

where the inequality follows for any q and e from $m'(x) > 0$ for $x > 0$ by **[C4]**. At any effort and for any equilibrium $p(\bar{q},\bar{e})$ and $\mu(\bar{q},\bar{e})$, therefore, (13) is satisfied by a unique level of equilibrium insurance $\bar{q} \in (0, \Delta)$. ■

As long as the first and second-order condition of optimality for individual choices both hold at the equilibrium, it is also possible to show that equilibrium effort is quite intuitively lower when more insurance is available:

Result 4 *The individually optimal equilibrium effort $e^{\dagger} = \bar{e}$ is negatively related to equilibrium insurance volume $\bar{q}$:*

$$\frac{d\bar{e}}{d\bar{q}} = \frac{-f'(\bar{e})\partial\left[u(c_l) - u(c_h)\right]/\partial q}{f''(\bar{e})\left(u(c_l) - u(c_h)\right) - v''(\bar{e}) + f'(\bar{e})\partial\left[u(c_l) - u(c_h)\right]/\partial e} < 0.$$

Proof. The first equality follows from total differentiation of the first-order condition (22) for effort choice at given insurance, with $e = \bar{e}$. To establish the sign of the resulting expression, note that the numerator is positive since $-f'(\bar{e}) > 0$ by **[C3]** and

$$\frac{\partial\left[u(c_l) - u(c_h)\right]}{\partial q} = u'(c_l)\left(1 - \frac{\partial}{\partial q}k(\bar{q},\bar{e})\right) - u'(c_h)\left(-\frac{\partial}{\partial q}k(\bar{q},\bar{e})\right) > 0$$

(more insurance reduces the welfare implications of negative shocks in equilibrium): formally, by (20) more insurance decreases c_h by the insurance premium $\frac{\partial}{\partial q}k(\bar{q},\bar{e}) = f(\bar{e})m(\bar{q}/N)$, which cannot exceed the unitary payoff of insurance contracts. The denominator is negative since $f''(\bar{e})\left(u(c_l) - u(c_h)\right) - v''(\bar{e}) < 0$

by the second order condition for individual effort choice, $f'(\bar{e}) < 0$ by **[C3]**, and

$$\frac{\partial\left[u(c_l) - u(c_h)\right]}{\partial e} = \left(u'(c_l) - u'(c_h)\right)\left(-\frac{\partial}{\partial e}k(\bar{q}, \bar{e})\right) > 0$$

follows from $u'(c_l) > u'(c_h)$ by **[C2]** when $c_l < c_h$, and $\frac{\partial}{\partial e}k(\bar{q}, \bar{e}) = f'(\bar{e})a\left(\bar{q}/N\right)\bar{q} < 0$ by (20). ■

This useful and intuitive pair of results requires sufficiency of the first-order conditions, which by Proposition 1 is ensured in equilibrium by a condition in the form

$$g\left(\frac{1 - f(0)}{f(0)}\frac{f(\bar{e})m\left(\bar{q}/N\right)}{1 - f(\bar{e})m\left(\bar{q}/N\right)}\right) \geq 1, \tag{24}$$

where $g(\cdot)$, as defined in (12), depends on the equilibrium price $p(\bar{q}, \bar{e}) = f(\bar{e})m\left(\bar{q}/N\right)$ as well as on the risk aversion parameter $\sigma > 0$, and is monotonically increasing in its argument. Condition (24) need not be satisfied for every economy of the type we model, and does not explicitly constrain the model's functional parameters, but it does provide useful analytical insights as regards features of the economic environment that support existence of a nontrivial equilibrium with positive effort and partial non-exclusive insurance. The demand for costly insurance, which depends on risk aversion and on the probability and size of the negative shock, should be strong enough as to ensure that the equilibrium marginal cost $m\left(\bar{q}/N\right)$ at which intermediaries supply insurance exceeds unity by a sufficiently large margin.

Whether condition (24) is satisfied also depends on the extent of moral hazard. To see this, note that insurance is positive in equilibrium by Result 3 and, by Result 4, optimal effort is therefore lower than the level $\tilde{e}$ implicitly defined by the first-order condition (4) evaluated at zero insurance. Since $\tilde{e}$ is positive and finite under **[C1-C3]**, $0 < f(\tilde{e}) < f(\bar{e}) < 1$, and the argument of $g(\cdot)$ in (24) is monotonically decreasing in $f(\bar{e})$, the inequality in (24) is certainly satisfied if

$$g\left(\frac{1 - f(0)}{f(0)}\frac{f(\tilde{e})m\left(\bar{q}/N\right)}{1 - f(\tilde{e})m\left(\bar{q}/N\right)}\right) \geq 1.$$

While a small $f(0)$ helps to ensure that price-taking individuals' choice problems are well defined as in Proposition 1, this sufficient condition for the validity of the first-order approach in general equilibrium is more easily satisfied when $f(\tilde{e})$ is large: quite intuitively, in establishing existence of a competitive non-exclusive insurance it is helpful to suppose that positive effort does not much reduce the loss probability, so that the moral hazard problem implied by the negative slope of $f(e)$ is not severe.

We now move on to derive the implications of a hypothetical (or, as discussed in the next section, policy-induced) change in the insurance quantity q. The unambiguously negative relationship between equilibrium effort and insurance established in Result 4 plays a crucial role in our first-order approach to characterization of the equilibrium's constrained inefficiency:

Proposition 2 *In competitive general equilibrium, the social cost of insurance, as measured by the negative effect of insurance on consumption levels, is larger than the marginal cost of insurance for individuals:*

$$\frac{dc_h}{d\bar{q}} = \frac{dc_l}{d\bar{q}} - 1 = -p\left(\bar{q},\bar{e}\right) - f'(\bar{e})a\left(\bar{q}/N\right)\bar{q}\frac{d\bar{e}}{d\bar{q}} < -p\left(\bar{q},\bar{e}\right). \tag{25}$$

Proof. Recognizing that in equilibrium $-\bar{q}p\left(\bar{q},\bar{e}\right) + \mu\left(\bar{q},\bar{e}\right) = -k(\bar{q},\bar{e})$, and using (20), the total derivative (7) of consumption levels with respect to $\bar{q}$ is

$$-p\left(\bar{q},\bar{e}\right) - \bar{q}\frac{dp\left(\bar{q},\bar{e}\right)}{d\bar{q}} + \frac{d\mu\left(\bar{q},\bar{e}\right)}{d\bar{q}} = -p(\bar{q},\bar{e}) - f'(\bar{e})a\left(\bar{q}/N\right)\bar{q}\frac{d\bar{e}}{d\bar{q}}.$$

(25) follows noting that $f'(e) < 0$ by **[C3]**, and $d\bar{e}/d\bar{q} < 0$ by Result 4. ■

As in Pauly's (1974) partial equilibrium and Helpman and Laffont's (1975) general equilibrium with actuarially fair contracts, insurance is too cheap in the economy we model. While exclusive insurance could allow the price of insurance to depend on each individual's incentives to provide effort, the anonymity of atomistic competitive trade implies that the price quoted by each insurance provider takes effort and probabilities as given. It fails to account for the effect of additional insurance on the payoff probability and on the cost of other providers' insurance services. Recalling (5) and (6), inequality (25) readily implies that the socially optimal marginal utility ratio is larger than the ratio implied by price-taking individual optimization. In a competitive equilibrium when competitive financial intermediaries issue non-exclusive unit contracts, consumption's reaction to the resource shock is too small. The representative individual pays the marginal cost of insurance but neglects its effect, through changes in effort, on consumption and insurance production costs.

In (25), the pecuniary externality $-f'(\bar{e})a\,(\bar{q}/N)\,\bar{q}\,(d\bar{e}/d\bar{q})$ is the product of $f'(e) < 0$, times insurance production costs, times the intuitively negative effort effect of insurance shown formally by Result 4. The first and third term are key to establishing constrained inefficiency. In the model (and arguably in reality), insurance production costs explain why insurance is only partial, and do need to be taken into account when, as in the next section, an inefficient competitive equilibrium is viewed from the standpoint of society.

By themselves, however, production costs do not imply any inefficiency. Competitive insurance trade neglects its own effect on the equilibrium prices, and on the rents paid by the insurance sector when its use of a specific factor restricts entry. This pecuniary externality would not imply any inefficiency, however, in the absence of moral hazard. If $f'(\bar{e}) = 0$, in fact, then $f'(\bar{e})a\,(\bar{q}/N)\,\bar{q}\,(d\bar{e}/d\bar{q}) = 0$ regardless of whether $a\,(\bar{q}/N) > 1$ because insurance production is costly.[12] Conversely, insurance is costless and complete in our model if $a(x) = 1$ for all x but the competitive equilibrium, should it exist, would be constrained-inefficient as long as $f'(\bar{e}) < 0$.

6 On policies

In our simple moral hazard economy we have shown that a competitive equilibrium with incomplete non-exclusive or "hidden" insurance can exist, and that it is not efficient when consumption levels are influenced by equilibrium effort and prices through channels that individual choices fail to internalize. The equilibrium's efficiency may however be improved if society,

[12]Costly insurance production is a novel and non-standard feature of our model, but it is very much in the spirit of the remark by Greenwald and Stiglitz (1986, p. 259) that "referring to economies with incomplete markets and incomplete information as 'imperfect' seems to be wrong: we do not refer to economies in which inputs are required to produce outputs as 'imperfect'; and the costs of obtaining information and running markets are no less real costs than other forms of production costs."

while unable to observe individual insurance quantities, does observe transaction prices and can enforce taxes or activate a public insurance scheme alongside the private non-exclusive insurance market.

6.1 Taxes

As in Helpman and Laffont (1975), taxation of observable insurance transactions may address the inefficiency identified by Proposition 2. When each insurance contract is subject to a unit tax τ, individuals pay a premium

$$p = f(\bar{e})m\left(\bar{q}/N\right) + \tau$$

and, in general equilibrium, their resource constraints include tax revenues:

$$\mu = f(\bar{e})\left(m\left(\bar{q}/N\right)\bar{q} - a\left(\bar{q}/N\right)\right) - \phi N + \tau\bar{q}.$$

Comparing (6) and (5), we see that a Pigouvian tax amounting to

$$\begin{aligned}\tau &= \bar{q}\frac{dp(\bar{q}, e(\bar{q}))}{d\bar{q}} - \frac{d\mu(\bar{q}, e(\bar{q}))}{d\bar{q}} \\ &= f'(\bar{e})a\left(\bar{q}/N\right)\bar{q}\frac{d\bar{e}}{d\bar{q}} > 0\end{aligned}$$

aligns the marginal utility ratio implied by the individual first-order condition to that required by social efficiency.

In asymmetric-information economies, it is generally difficult to show that this or other equivalent indirect mechanisms may implement the socially optimal allocation: taxes and subsidies can erase the wedge between individual and social optimality conditions only if the economy's choice problem is locally concave when the first-order conditions are satisfied. In our economy, however, local concavity is preserved when insurance becomes more incomplete than required by Proposition 1's sufficient conditions. Hence, positive taxation of insurance premia can improve efficiency of a decentralized competitive equilibrium that remains well defined if insurance is sufficiently incomplete. In reality, as in our Section 4, this may be rationalized by costly production of insurance services: according to various issues of the *OECD Insurance Statistics Yearbook*, between 1996 and 2005 operating expenses amounted to about one third of private non-life insurance claims in advanced countries. While it is easy to establish that the equilibrium is inefficient in the absence of specific insurance taxes, it is much more difficult to assess the welfare implications of those that do exist in reality (e.g. in the United Kingdom, where an Insurance Premium Tax at rates between 6% and 20% is imposed on common non-life insurance contracts). To do so, one would need to estimate the optimal tax rates that in our simple economy could in principle be computed evaluating the derivatives appearing in the proof of Proposition 2, and might in future research be investigated in the context of more realistic models.[13]

6.2 Public transfers

Next, we consider whether and how insurance may be influenced by a public contingent tax and transfer scheme that pays a possibly negative amount s to all individuals who experience the verifiable negative shock. If the transfers and the scheme's administration costs are funded by a lump-sum levy or rebate $T(f(e)s)$, such that $T(0) = 0$ and $T'(\cdot) > 0$, contingent consumption levels are given by

$$c_h = z - qp + \mu,\ c_l = c_h - \Delta + s + q,$$

where $\mu = f(\bar{e})\left(m\left(\bar{q}/N\right) - a\left(\bar{q}/N\right)\right)\bar{q} - \phi N - T(f(e)s)$.

When free entry in production of insurance services implies constant returns, at given effort the private insurance premium $p = f(\bar{e})m\left(\bar{q}/N\right)$ is not affected by the scheme. Hence, public transfers crowd out private insurance fully unless they change effort through income effects.

[13]In such work it may also be useful to recognize that when multiple consumption goods are (like torches and fire extinguishers) differently associated with behavior that affects the probability of accidents then, as shown by Arnott and Stiglitz (1986) and Greenwald and Stiglitz (1986, Section II.D), commodity taxes can implement allocations that are Pareto-superior to a laissez-faire competitive equilibrium (if one exists) with moral hazard and actuarially fair insurance.

A public contingent transfers scheme can in fact influence effort and insurance when its cost structure differs from that of private insurance, as may be realistic since OECD (2004) data report “tax administration costs per net revenue collection” of only about 5%. If the cost of public insurance is lower than that of private insurance then public transfers affect average consumption levels and, with $u'''(\cdot) > 0$, equilibrium insurance and effort through the resulting change in local risk aversion. The implications of public insurance schemes are more complex, but qualitatively similar, when decreasing returns in the insurance industry let public transfers influence the equilibrium price of insurance.[14]

The welfare impact of such a scheme can be characterized in terms of first-order conditions as long as Proposition 1 applies, which is the case if its conditions hold at $s = 0$ and transfers do not imply smoother marginal utility. The efficiency objectives appropriate for the representative-individual economy we study in this paper would indeed call for the scheme to reduce the excessive consumption smoothing delivered by non-exclusive competitive insurance transactions. While in reality redistribution policies also pursue a variety of other objectives, this may explain some of the cross-country variation of public and private insurance discussed in Bertola and Koeniger (2010).

7 Concluding comments

If insurance against a negative shock is sufficiently incomplete in a single period economy with two possible realizations of uncertainty, we have shown that a competitive equilibrium can exist and be characterized by first-order conditions in the presence of both hidden effort and hidden non-exclusive insurance contracts. While the validity of the first-order approach in asymmetric-information general-equilibrium models with multiple states and hidden assets remains a chal-

[14]In a microfounded general equilibrium model of public and private insurance, the economy’s technology and/or its endowments of sector-specific factors would determine in standard fashion the cost structure of private and public insurance provision.

lenge for future research, our analytical derivations characterize an appealing equilibrium that is supported under standard functional form assumptions by realistically incomplete insurance. As in previous studies of related problems, insurance incompleteness plays an important role in ensuring equilibrium existence. In our model economy, incomplete insurance is implied by costly production of insurance services. The efficiency implications of moral hazard may then be characterized in an otherwise standard general equilibrium setting where individual insurance and effort choices satisfy first- and second-order conditions, but neglect pecuniary externalities that can be corrected by collective policies.

References

Ábrahám, Árpád, and Nicola Pavoni (2005) "The Efficient Allocation of Consumption under Moral Hazard and Hidden Access to the Credit Market," Journal of the European Economic Association 3, 370–381.

Ábrahám, Árpád, Sebastian Koehne and Nicola Pavoni (2011) "On the First-Order Approach in Principal-Agent Models with Hidden Borrowing and Lending," Journal of Economic Theory 146, 1331–1361.

Ales, Laurence, and Pricila Maziero (2009) "Non-Exclusive Dynamic Contracts, Competition, and the Limits of Insurance," Carnegie Mellon University, working paper.

Arnott, Richard, and Joseph E. Stiglitz (1986) "Moral Hazard and Optimal Commodity Taxation" Journal of Public Economics 29, 1–24.

Bertola, Giuseppe, and Winfried Koeniger (2010) "Public and Private Insurance with Costly Transactions" IZA Discussion Papers 5201, CEPR Discussion Paper 8062.

Bisin, Alberto, John D. Geanakoplos, Piero Gottardi, Enrico Minelli, and Herakles Polemarchakis (2011) "Markets and Contracts," Journal of Mathematical Economics 47, 279–288.

Bisin, Alberto, and Danilo Guaitoli (2004) "Moral Hazard and Nonexclusive Contracts," RAND Journal of Economics 35, 306–328.

Bisin, Alberto, and Piero Gottardi (1999) "Competitive Equilibria with Asymmetric Information," Journal of Economic Theory 87, 1–48.

Bizer, David S., and Peter M. DeMarzo (1992) "Sequential Banking," Journal of Political Economy 100, 41–61.

Chetty, Raj, and Emmanuel Saez (2010) "Optimal Taxation and Social Insurance with Endogenous Private Insurance," American Economic Journal: Economic Policy 2, 85–114.

Cole, Harold, and Narayana Kocherlakota (2001) "Efficient Allocations with Hidden Income and Hidden Storage," Review of Economic Studies 68, 523–542.

Gollier, Christian (2001) The Economics of Risk and Time, MIT Press, Cambridge (MA).

Golosov, Mikhail, and Aleh Tsyvinski (2007) "Optimal Taxation with Endogenous Insurance Markets," Quarterly Journal of Economics 122, 487–534.

Gottardi, Piero, and Nicola Pavoni (2011) "Ramsey Asset Taxation under Asymmetric Information," Notes, Bocconi University.

Greenwald, Bruce C., and Joseph E. Stiglitz (1986) "Externalities in Economies with Imperfect Information and Incomplete Markets," Quarterly Journal of Economics 101, 229–264.

Helpman, Elhanan, and Jean-Jacques Laffont (1975) "On Moral Hazard in General Equilibrium Theory," Journal of Economic Theory 10, 8–23.

Hellwig, Martin F. (1983a) "Moral Hazard and Monopolistically Competitive Insurance Markets," Geneva Papers on Risk and Insurance 8, 44–71.

Hellwig, Martin F. (1983b) "On Moral Hazard and Non-Price Equilibria in Competitive Insurance Markets," SFB Discussion Paper No. 109, University of Bonn.

Krueger, Dirk, and Fabrizio Perri (2011) "Public versus Private Risk Sharing," Journal of Economic Theory 146, 920–956.

Loss, Frédéric and Gwenaël Piaser (2013): "Linear Prices Equilibria in a Non-Exclusive Insurance Market," Working Paper, CNAM.

Mirrlees, James A. (1971) "An Exploration in the Theory of Optimum Income Taxation," Review of Economic Studies 38, 175–208.

OECD (2004) Tax Administration in OECD Countries: Comparative Information Series, Paris, OECD.

Pauly, Mark V. (1974) "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," Quarterly Journal of Economics 88, 44–62.