

Insuring the Uninsurable: Brokers and Incomplete Insurance Contracts

Neil A. Doherty and Alexander Muermann*

The Wharton School, University of Pennsylvania

October 2005

Abstract

How do markets spread risk when events are unknown or unknowable and where not anticipated in an insurance contract? While the policyholder can “hold up” the insurer for extra contractual payments, the continuing gains from trade on a single contract are often too small to yield useful coverage. By acting as a repository of the reputations of the parties, we show the brokers provide a coordinating mechanism to leverage the collective hold up power of policyholders. This extends both the degree of implicit and explicit coverage. The role is reflected in the terms of broker engagement, specifically in the ownership by the broker of the renewal rights. Finally, we argue that brokers can be motivated to play this role when they receive commissions that are contingent on insurer profits. This last feature questions a recent, well publicized, attack on broker compensation by New York attorney general, Elliot Spitzer.

Key Words incomplete insurance contracts, brokerage, contingent commissions, reputation

JEL Classification G22, G24, L14

*Doherty and Muermann: Department of Insurance and Risk Management, The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104-6302, USA, email: doherty@wharton.upenn.edu - muermann@wharton.upenn.edu. We wish to thank Christian Laux, Georg Noeldeke, Soenje Reiche and seminar participants at the Risk Theory Society Meeting 2005, the conference “Risk Transfer between (Re-)Insurers, Banks, and Markets” at the Center for Financial Studies, the World Risk and Insurance Economics Congress 2005, and the University of Alabama for valuable comments.

1 Introduction

Our objective is to show how insurance markets can provide an orderly mechanism by which policyholders can receive transfers for losses that are non verifiable and, possibly were never anticipated when coverage was written. Given the pace of technical, social and legal change, such fuzzy risks are cropping up with some frequency. Examples include the asbestos claims that have cost the insurance industry tens of billions of dollars, the sudden appearance of toxic mold in insurer claims, various forms of "cyber loss", and possibly the 9/11 losses which, despite the war exclusion on many policies, were not disputed by insurers and reinsurers. We will show that brokers play a central role in extending insurance markets to cover non-verifiable losses.

A clue to the role of brokers lies in their contractual arrangement with insurers.¹ It is normal for brokers to "own the renewal rights" on the book of business they place with the insurer. This means that the broker has the renewal rights to the policies it places with that insurer. The broker is free to recommend to its clients that they renew with the current insurer or switch to a rival. Contrary, the insurer revokes any right to direct solicitation of business places through the broker. This provision vests the broker with considerable hold up power. Why would brokers seek this hold up power, how would they exercise it and why would insurers choose to vest this power in the broker?

In our model, insurance contracts are written that cover verifiable losses. However, non verifiable losses also can arise. Brokers then encourage insurers to make transfers for the non verifiable losses, against the threat that the business will be withdrawn and the future rents lost to the insurer. In this way, the market for risk transfer is extended to include losses that are not contractible. For their part, insurers are willing to expose their reputations to this hold up threat because they can extract rents from the extended informal coverage. For the broker's part, they can participate in those rents and, indeed, a profit based commission will make the threats they hold over insurers credible.²

Previous literature has stressed several roles for brokers including search agents who match trading

¹Ironically, it is the completeness of the broker-insurer contract that supports the incompleteness of the insurance contract that we now describe.

²Some observers have noticed a trend towards the ex post negotiation of claims. Apparently, insurers are now more likely to dispute large claims, to offer less than 100 cents on the dollar, or to try to get away without paying. Richard and Barbara Stewart (2001) have labeled this the "loss of certainty effect" and Kenneth Abraham (2001) has talked of the "de facto big claims exclusion". One reason for such disputes is that large claims threaten insurer solvency and such offers may be seen to resemble workouts in which distressed non-insurance firms negotiate with creditors. But the issue here is with the willingness, not the ability, to pay. These writers see the "big claims exclusion" as degradation of the insurance market because risk-averse consumers will place a lower value on such uncertain insurance. Indeed, they see a potential downward spiral of the insurance market if this practice continues. Contrary, we argue that ex post negotiation might represent an expansion of the insurance market to include transfer for non verifiable losses.

partners (e.g. Rubinstein and Wolinsky, 1987) and information intermediaries who are able to transmit credible information and thereby lower costs of adverse selection (Biglaiser, 1993). In addition to these roles, we show that brokers act as a clearing house of the reputations of insurers. This role enables brokers to “hold up” insurers for ex post reasonable transfers for non verifiable losses. Given appropriate compensation, brokers can exercise a credible threat to insurers which leads to more complete insurance markets. Our work parallels that of Kingston (2005) who shows that brokers can help sustain cooperation in markets where lack of trust can lead to a prisoner’s dilemma in which parties default on their trades. This examination of brokers is topical. Brokers have been assailed recently by the New York attorney general, Elliot Spitzer, and at the center of this assault is the compensation structure for brokers. Contrary to Spitzer’s assertions, we will show that the compensation structures criticized by Spitzer, can lead to real benefit to policyholders.

1.1 Non-Verifiable Losses

At the heart of our paper are non-verifiable or non-contractible events. By these we mean events that are incapable of inclusion in the policy because they can not be anticipated; or events which are simply too complex to include in the contract, or events or circumstances which could be included but, due to time inconsistency or the prospect of new information, the parties believe it preferable to bargain after the fact.³ Non-verifiable losses can also refer to the size or cause of the loss. An insurance policy might be specific about how a claim is to be settled (damage to a home or its contents might be limited to the repair cost or the cost of replacement with something of similar condition), but the insurer and insured might wish to leave open the possibility that settlement can be more generous as idiosyncratic circumstances dictate, or less generous if there is suspicion of claim fraud (which is difficult to prove).

Consider unanticipated losses. The parties consider the possibility that some unanticipated losses might occur. These unanticipated losses are not insurable in a formal contract because they cannot be specified and, even if they could be specified, they might be unsuited to insurance perhaps because they would incite severe ex post moral hazard, or because they are undiversifiable. However, there is another class of unanticipated losses for which, had they been anticipated, the parties would concur that they are insurable.

³Some policies specify the perils and losses that are covered. If a loss occurs that is not specified, then it is not covered. Other policies work in the opposite direction, they cover everything that is not included. The latter does provide a structure for including the unanticipated, but does so at a cost – it is open ended and becomes very difficult to price. Moreover, having such open policies complicates the insurer’s financial and risk management.

Let us simply describe these as ex post insurable.⁴ If the parameters by which they could be deemed ex post insurable could be pre-specified, then there is nothing to prevent the parties from conditioning the insurance on these parameters. For example, suppose that the only thing that separated an ex post insurable from an ex post uninsurable event, was diversification. An ex post insurable event might be one that hit only one policyholder but an event impacting many policyholders would be ex post uninsurable. If things were this simple, a contract could be written conditioning coverage on the insurer's surplus (rather like a mutual contract). But the circumstances that determine whether an event is ex post insurable may not be that easy to pre-specify or that easy to anticipate. For example, consider toxic mold, which burst onto the insurance scene as an unanticipated loss recently. Not only can it be an undiversifiable loss but, going forward, its coverage carries significant moral hazard. The fear is that insurance may be seen as a substitute for proper repair and maintenance of property. It may not be practical to write into contracts enforceable exclusions based, not only on the peril which is unanticipated, but on the moral hazard it might engender. Table 1 illustrates the criteria often listed in insurance textbooks for insurability (the loss is anticipated, measurable, etc.) and suggests how such factors might determine whether an unanticipated loss is ex post insurable. Thus we suggest that a surprise event may be insurable going forward if it is measurable (ex post), there is low correlation, low moral hazard, and information is symmetrically distributed between the parties.

Classification of Losses	Ex ante insurable	Ex post insurable	Ex post uninsurable
Anticipated	Yes	No	No
Measurable	Yes	No	No
Observable	Yes	Yes	No
Low correlation	Yes	Yes	No
Low moral hazard	Yes	Yes	No
Symmetric information	Yes	Yes	No

Table 1

⁴Berliner (1982) defines insurable losses as satisfying the following criteria. Losses are anticipated, measurable and, after the fact observable. There must be low correlation and little moral hazard or adverse selection. Losses that are not anticipated but satisfy the other conditions are considered ex post insurable.

1.2 Examples of Incomplete Insurance Contracts

Consider the following examples of incomplete contracts. The first involves reinsurance. Traditionally, the reinsurance market has been considered to be one based on relationships. Contracts were not very detailed and relationships between the insurer and reinsurer tended to persist over many years. In this relationship, the parties link their fortunes. If an unusual or uncovered loss arises, the reinsurer will often pay without raising a fuss. But there is a corresponding obligation of “payback”. The payback feature makes the contract something like a debt contract. Jean-Baptiste and Santomero (2000) argue that long term implicit contracts are efficient because they allow evolving information to be included in pricing. But reinsurance contracts may be incomplete in other ways. Contracts rarely specify the underwriting and claims settlement practices to be adopted by the primary insurer. The reinsurer will usually monitor the relationship but permit the primary insurer considerable flexibility in exploiting its core skills. Finally, reinsurance contracts often are not as specific in defining coverage as are the direct policies from which they are derived. This allows some ex post flexibility for the reinsurer to respond to losses that may not be covered. For example, despite the war exclusion on many reinsurance policies, reinsurers uniformly responded to the 9/11 losses of their primary insurance clients without seeking to evoke this exclusion.

The feature of the reinsurance market that is of particular interest here is the presence of brokers. If a reinsurer behaves badly to the primary insurer, the broker will know of it. The consequence for the reinsurer might be not only a loss of that contract, but a diversion of other business from the broker to other reinsurers. This leveraging of reputation enhances the “hold-up” power of the primary insurer.⁵

The second example of contract incompleteness lies in the procedures used to inaugurate coverage in the brokered reinsurance and high-end commercial insurance markets. Coverage is often syndicated across a number of carriers. To set this in place, the broker approaches a number of insurers asking for a commitment to provide coverage for a portion of the risk. To make such a commitment, the underwriter signs a binder (or slip at Lloyds) indicating how much coverage is offered. The interesting thing about this binder is the lack of specificity about the coverage.⁶ Sometimes, the binder will refer to a particular policy form indicating

⁵While we are stressing the relationships in the reinsurance markets, there are signs of an apparent shift to a more commoditized approach, which is highlighted by, though not confined to, insurance securitization. In contrast to relationship-based transactions, insurance securitizations are embodied in precise contractual form leaving little room for discretionary performance and ex post bargaining.

⁶In a dispute between the leaseholders and insurers of the World Trade Center, the central issue is whether there was one “occurrence” or two. This is of importance because policies had not been issued at the time of the loss and coverage offered under binders was specific enough to give policy limits per occurrence but not specific enough to avoid dispute as to the meaning of “occurrence”. Had a policy been issued, then there would have probably been a precise definition of the term.

the general form of coverage. Other times the binder will simply commit the underwriter to the same type of coverage offered by another insurer. But sometimes, the coverage is still to be agreed. This practice permits coverage to be arranged, or renewed without waiting for the contract to be issued. Coverage need not be delayed pending the agreement of all contract details. The potential downside is that the lack of specificity offers leeway for dispute.

As a third illustration of incomplete insurance, some insurers such as Chubb Insurance company, have made and protected a reputation for going the “extra mile” to ensure its personal (and other) policyholders are happy with their claims settlements. The strategy is to resolve ambiguity over amount or coverage more in the policyholder’s favor. Thus, there is a willingness to go beyond the narrow limits of the contract to ensure that the policyholder is adequately compensated. This flexibility introduces a degree of contract incompleteness, which is resolved after the fact in the claim negotiation. Chubb’s distribution system is a set of independent agents and brokers. These intermediaries “own” the renewal rights and can advise clients to move business if they become unhappy with Chubb’s claims performance. Chubb entrusts its reputation to these agents and brokers in order to pre-commit to a “policyholder friendly” claim settlement strategy.

In each of these illustrations, the relationship is brokered. This is important in shaping the discretion the parties can exercise. For example, suppose Chubb fails to make a suitably generous offer to settle a claim, despite the fact that it promotes itself in this basis and charges higher premiums to cover such settlements. The broker can now exercise discretion in its response.⁷ If it suspects that the claim is fraudulent or inflated, it might condone Chubb’s lowball offer or refusal to pay. On the other hand, if the broker believes that Chubb is failing to meet the reasonable expectations of a deserving policyholder, it might withdraw business from Chubb. Thus, the reputation costs to Chubb from its unsatisfactory claim offer could be very high. There is a corresponding mechanism to discipline the policyholder. If the policyholder plays fast and loose on its claims, the broker may withdraw its services – indeed the broker’s reputation with insurers is one of its most valuable assets and brokers may have a financial incentive to drop troublesome policyholders.

Alternatively, suppose insurance was offered on a binder but this minimal document was not specific in the terms of coverage. A loss occurs before the policy is issued and the insurer balks on payment.

⁷Kingston, 2005, gives a parallel example of 18th century marine insurance in England. Given slowness of communications, captains of vessels were given considerable leeway in changing routs and even return cargoes. Thus, marine contracts were largely incomplete and this incompleteness extended to insurance contracts. As Kingston notes, “reputation was key” to successful underwriting and quotes a contemporary broker that “the private underwriters will settle a loss for a man of character, where they will not for a man who they suspect.

In choosing how to react to this breakdown, the broker must decide whether the event was one which, in principle, is insurable and which a reasonable person would be expected to be covered. If the answer is affirmative, the broker might extract a large reputation toll from the insurer. But, if the policyholder “is trying it on” – to hold up the insurer to pay for a loss that is not an insurable loss (e.g. poor business performance) – the broker will not impose a reputation cost on the insurer for refusing to pay.

By acting as a guardian of reputations, the broker can help secure a market for non-contractible losses. The formal and informal incentive structure for the broker reinforces this value-creating role. There is some ambiguity as to whether an insurance broker is the agent of the policyholder or of the insurer. The policyholder chooses whether to seek the services of a broker who will advise on insurance, on the choice of insurers, and will place contracts with specific insurers. The policyholder also is concerned with the treatment of its claims and usually expects the broker to become involved. The broker’s involvement can range from a monitoring of the settlement to exerting pressure on the insurer to go beyond the strict contract wording. Thus, the ability of the broker to create a market for non-verifiable losses will enhance the demand for its services. The issue of legal agency is clouded by the payment for the service and by the incentives this creates. The normal structure is for the broker to receive a commission from the insurer based on the premium. However, many policyholders, particular large commercial clients, negotiate a fee with the broker (for which the proportional commission is an offset) related to the perceived value added. Such an arrangement provides a mechanism for the broker to be directly rewarded for creating insurance coverage for non-verifiable losses and for disciplining insurers that behave poorly towards their policyholders.

But brokers also act on behalf of insurers. Insurers often have contingent fees for brokers under which an additional compensation is paid to the broker based on the revenues and/or profitability of the book of business the broker holds with the insurer. Insurers typically compete amongst each other in the design of these profit and revenue sharing schemes and there is evidence that brokers do indeed respond to these incentives in their placement and cancellation decisions, (see Wilder, 2002).

In our formal model below, brokers respond to these bilateral incentives in the following way. Brokers offer an implicit deal to policyholders that they will use their leverage over insurers when unusual claims arise. For their part, insurers understand that such leverage will be used and that this adds value for the policyholder. Thus, insurers price for this effect. When a verifiable loss occurs it is paid according to the contract. When a non-verifiable loss arises, the broker determines whether the loss is ex post insurable or

not. If the broker determines the loss to be ex post insurable, the insurer and policyholder negotiate for a settlement. If the insurer fails to make a reasonable offer, the broker inflicts reputation penalty on the insurer. With such a breakdown, the broker no longer feels that its clients are getting good service from this insurer and it diverts business to rivals or extracts concessions from the misbehaving insurer in the form of lower future prices on the business it places. If the broker determines the loss to be ex post uninsurable, no claim is made. The broker could also inflict a reputation penalty on the policyholder for trying to push an ex post uninsurable loss onto the insurer. For example, if the broker feels that the loss is due to fraud, or is unreasonably inflated (i.e., uninsurable), the broker could inflict a reputation cost on the policyholder. The broker now is less willing to represent such policyholders because they may hurt its contingent fees from the insurer or compromise its reputation with the insurer and therefore its ability to place higher quality policyholders.

2 Related Literature

Our paper is related to the literature on costly state verification and falsification in the way that observability and/or verifiability of states of the world is restricted which impacts the types of “contracts” that can be written.⁸ The literature on costly state verification (Townsend, 1979, Bond and Crocker, 1997) considers situations in which (some) states are only observable to the policyholder but that the insurer can verify these state ex-post at a cost. After a claim is filed, the insurer now has the choice to audit the claim. A “contract” is then represented by a premium, a verification set which specifies those claims to be monitored, and a transfer from the insurer contingent on the policyholder’s claim and the auditing outcome. The literature on costly state falsification (Crocker and Morgan, 1998) considers situations in which states are also only observable to the policyholder but they cannot be verified at any cost. It is, however, costly to the policyholder to falsify claims. Ex-ante, the policyholder and insurer sign a “contract” specifying a premium and an ex-post transfer from the insurer contingent on the policyholder’s claim. In all these models, full

⁸Our paper also relates to the literature on implicit risk sharing (e.g. Kimball, 1988, and Kocherlakota, 1996). These papers are applicable to village communities (e.g. farmers, Kimball) or to two households (Kocherlakota) where individuals are in close contact and information about defect becomes easily known (e.g. through a blacklist). It is assumed that no future trade will take place once a party defects and this threat is assumed to be credible. Our paper is different in two ways. First, we assume that an institutional structure, the insurance market, is in place to organize risk sharing. We then address the question how risk sharing can be extended to include non contractible events? In contrast to the papers above, we thus look at the coexistence of explicit and implicit risk sharing which leads to interesting interactions. Second, we show how risk sharing can be improved through the broker serving as a coordination mechanism. This allows for scenarios in which information does not flow easily, i.e. outside village communities. Interestingly, not only can the broker improve implicit but also explicit risk sharing.

coverage cannot be achieved as either the verification cost to the insurer or the falsification cost to the policyholder is embodied in the equilibrium contract. In both approaches, it is assumed that the transfer schedule is enforceable in front of a court. In this paper, we investigate situations with an even higher degree of “incompleteness”, namely events that do not allow for any transfer schedules to be enforceable ex-post by a court. The insurer has no contractual obligation to indemnify the policyholder ex-post. The mechanism we present in this paper is based on the hold-up power the policyholder derives from future rents that the insurer is able to extract from an ongoing relationship. As those rents are costly, i.e. the policyholder has to pay an implicit loading, full coverage is precluded by this mechanism. The broker’s ex-post judgement and leverage influences the size of hold-up and so affects the transfers from the insurer to the policyholder.

The role of the broker in our model in some ways resembles the role of the courts in Anderlini et al (2003 a ,b).⁹ Both courts and brokers must exercise judgment in deciding whether an unanticipated loss should be covered under the contract and in both cases they are guided by the efficiency gains this “precedent” implies for future contracts. In Anderlini et al, the court’s desire for efficiency gain is implied by its adoption of a social welfare function. In our model, brokers seek future efficiency gain because they can capture rents directly from value added. However, there are differences. In our case, there is a modeled efficiency gain in the current period. The parties can choose to contract with one or another of a number of competing brokers. In making this choice, the degree of hold up, and therefore the terms of ex post bargaining, can be bounded by the parties. The second difference lies in the nature and expertise of the institutions. The nature of the ex post judgment (whether a revealed loss is ex post insurable) is technically quite complex. Courts, by their nature, have no core skills to address such problems and rely on the expertise imported by the parties in an adversary process. Brokers do have these core skills and derive rents directly from the creation of value. But perhaps the biggest difference between our model and theirs lies in its purpose. Whereas they address unanticipated events that frustrate the purpose or outcome of a contract (i.e., they are an unintended nuisance), we address a situation where the parties know unanticipated events can happen and we are trying to expand the domain of the contracting relationship to encompass such events (i.e., to provide insurance). This is why, for us, it is important to address efficiency gains in the current period.

Perhaps, most closely related is the contemporaneous paper of Kingston (2005). He considers an economy where people can defect on trades. In a single period, this is a prisoner’s dilemma. However, in a repeated

⁹The papers of Anderlini et al. and ours both build on the incomplete contract literature (see Grossman and Hart, 1986, Hart and Moore, 1990, Hart, 1995).

game, traders may be disinclined to default if their reputation damage prevents gains from future trades. Without intermediaries, the dissemination of information on defaults may be limited and the reputation sanctions correspondingly small. However, if trades are conducted through intermediaries (without contact between buyer and seller) the intermediaries can act as a clearing house for information. Given frequent interaction of each party with the broker, the sanction for default can be enhanced thereby lowering the incentive to cheat. We also use intermediaries as a clearing house for information on performance and allow the broker to impose a sanction for non performance. The difference lies mostly in objective. Whereas Kingston is looking generally for an operative mechanism to promote trust when contracts cannot be enforced, we are focused on the insurance market. Thus, we look in detail at the structure of the partially complete contracts and market relationships that permit transfers to be made for non verifiable losses. Moreover, instead of assuming a “nuclear” trigger for sanctions, we look at how information is updated and how brokers are compensated to see whether threatened sanctions are indeed credible.

3 “Insuring” Non-Verifiable Losses without Brokers

In this section, we consider a simple world with a risk-averse individual and multiple, risk-neutral insurers. The individual is endowed with initial wealth w_0 and exposed to a loss, L , with probability p . The loss may be verifiable (and therefore insurable under a conventional policy) with a probability q or non-verifiable but ex-post insurable with a probability $1 - q$. The non-verifiable loss is observed by both the individual and the insurer, but cannot be contracted upon. For example, this may be a loss that could not have been anticipated at the time the contract was written but, once it has occurred, it is clearly observed by all parties.¹⁰

Now consider the following infinite period problem. The individual buys a contract from a competitive insurer to cover an amount c of the verifiable loss (probability pq). However, the premium exceeds the actuarial value of the verifiable loss, $P > pqc$. Now suppose that the insurer expects that the policyholder will renew this coverage indefinitely. With these renewal expectations, the insurer would make rent having a present value of $(P - pqc)/r$ where r is the discount rate.

The expected rent only will be realized if the policyholder does renew the contract. This “hoped-for”

¹⁰For simplicity, we will assume that all non-verifiable losses are ex-post insurable. The addition of ex-post uninsurable risks to our model would create a background risk.

rent provides “hold up” power to the policyholder. Suppose a non-verifiable loss occurs. The size of the loss is L , but the policyholder believes (for reasons we will examine presently) it would be appropriate for the insurer to make a non-contractual payment of b to the policyholder. Would the insurer be willing to make a non-contractual payment for this non-verifiable loss in order to retain the customer and secure these future rents?

We start by assuming that the policyholder has a prior expectation that the insurer will indeed make a payment of b for a non-verifiable loss. At inception, the policyholder has the same prior for any insurer from whom it might write the contract, but the policyholder selects one, the incumbent. All insurers charge the same premium. If a non-verifiable loss arises and the insurer makes a payment, the policyholder’s prior for such payment in the future from this insurer is revised upwards. Thus, the posterior that the generous incumbent will make such settlements in the future is higher than for potential rivals. Given that all insurers charge the same premium going forward, the policyholder chooses to renew the policy with the incumbent. However, if the insurer does not pay the non-verifiable loss, the policyholder revises the prior downwards. Thus, the posterior for the incumbent is below the prior for rivals and, given the same premium, the policyholder will switch to a rival.

Any possible subgame perfect equilibrium to this game must involve either

1. insurers charge a premium in excess of the expected value of verifiable losses, $P > pqc$, and will choose to make a payment, b , should a non-verifiable loss occur. To ensure that expectations are met in equilibrium, the prior that such payment will be made is one; or
2. insurers charge a premium which just covers the expected value of the verifiable loss, $P = pqc$. For these insurers, the prior is zero.

Consider now a steady state in which, in any future year, policyholders might have an expectation that a payment of b will be made against the unverifiable loss. In this case, future rents are reduced by the expectations of future “type- b ” payment. Thus, assuming b to be constant over time, the insurer will make the payment, b , only if b is less than future rents discounted at the interest rate r , i.e. if

$$b \leq \frac{P - pqc - p(1 - q)b}{r}.$$

If the policyholder has all the bargaining power, then the premium P which includes future “type- b ”

payment is given by

$$P = pqc + p(1 - q)b + rb.$$

The premium is the sum of the actuarially fair value, $pqc + p(1 - q)b$, and a loading, rb . This loading represents the future rents that provides the insurer with the incentive to pay the transfer b in case of a unverifiable loss.

Would the policyholder and insurer wish to engage in this partially incomplete contract in which there is conventional insurance coverage together with the ability of the policyholder to hold-up the insurer to pay non verifiable, and therefore not contracted losses? If so, then such an arrangement provides explicit insurance for contractible losses and implicit insurance for the non-verifiable (non-contractible) losses.

Proposition 1 *In the model outlined above, it is optimal for the policyholder to partially insure both types of losses and to buy more coverage on the verifiable than to implicitly generate on the non-verifiable event, i.e. $b^* < c^* < L$ for all r . Furthermore, there exists a critical discount rate $\bar{r} > 0$ such that it is optimal to generate a transfer if the discount rate is below \bar{r} and not to generate a transfer otherwise, i.e. $b^* > 0$ if $r < \bar{r}$ and $b^* = 0$ if $r \geq \bar{r}$.*

Proof. See Appendix A.1. ■

For $r < \bar{r}$, we thus have a subgame perfect equilibrium where insurers charge a premium in excess of the expected value of verifiable losses and will choose to make a payment for a non-verifiable loss. Policyholders have a prior of one that such payment will be made. For all $r \geq \bar{r}$, we have a subgame perfect equilibrium where only verifiable events are covered and policyholders have a prior of zero that payments for non-verifiable losses are made.

The intuition behind these results is as follows. In order to generate a “hold-up” and thereby payments for unverifiable losses, the policyholder has to pay a loading rb . The premium is therefore unfair and full “insurance” of the unverifiable loss is not optimal.¹¹ Purchasing full coverage of the verifiable loss would then lead to a higher marginal utility in the state of the unverifiable event. The policyholder thus finds it optimal to transfer wealth into that state by not buying full coverage of the verifiable loss and thereby reducing the premium. Last, the discount rate measures the level of the loading, rb , and for high discount rates the policyholder will find it optimal to not generate the transfer b .

¹¹The result still holds if there is only a non-verifiable loss, i.e. if $q = 0$, and is therefore robust with respect to different correlation structures between the verifiable and unverifiable loss.

4 “Insuring” Non-Verifiable Losses with Brokers

In the bilateral model above, the policyholder could hold up the insurer for a payment against the non-verifiable loss, by threatening to withdraw his/her future business. This implies that, for the hold up to result in a positive transfer, future rents must be strictly positive. Indeed, the risk averse policyholder is willing to endow the insurer with positive rent in order to create a mechanism to hedge the non-verifiable loss as long as the interest rate is not too high.

The problem with this model is that it results in either a very small transfer, because future profits on the policy are small, or that the policyholder must bestow very large rents on the insurer in order to get a useful level of coverage for potentially large losses. In principal, we can mitigate this trade off between coverage and rent, by changing the information assumptions. Instead of allowing only the policyholder to observe the settlement, b , we allow all transfers to be observed by all the insurer’s policyholders. With similar assumptions about beliefs, (i.e., all policyholders upgrade (downgrade) their priors about future settlements, when the incumbent makes (fails to make) a settlement of b_{br}^*) then all the firm’s policyholders would switch insurers if the insurer failed to pay b_{br}^* for any one of them suffering a loss. This increases the hold up power of the claimant which allows larger coverage to be obtained for non-verifiable losses, whilst keeping the rent contribution in each premium fairly low.

The enhanced model outlined in the previous paragraph suggests that brokers may not be strictly necessary to create a market for non-verifiable losses. However, the informational assumptions are strong; all policyholders must observe all of the incumbent insurers claim payments to all policyholders and be prepared to switch if even one policyholder is denied. We clearly need some coordination mechanism to harness the hold-up power of multiple policyholders, while permitting the insurer to extract rents for exposing itself to this ex-post hold up. We present the insurance broker as the coordinating device.

There are multiple agents of three types: risk-averse policyholders, risk-neutral insurers, and brokers. Each policyholder initially secures a broker to arrange insurance and offers a compensation, k , to the broker which we assume to be a linear function of the insurer’s rent. This compensation ensures that the broker has an ex ante interest in securing an incomplete contract with expectations of a settlement for non-verifiable losses (recall that these rents provide the ex post hold up power from which the settlement is made). Suppose that each broker arranges incomplete contracts between m such policyholders and an incumbent insurer. Such contracts will offer explicit coverage of c_{br}^* for verifiable losses and an expected settlement of

b_{br}^* for non verifiable. Brokers can advise policyholders to move that business and thereby extract a penalty from the incumbent insurer for failure to make an appropriate payment for a non-verifiable loss. Each policyholder observes its settlement of a non-verifiable claim, or the broker's penalty on the insurer if it is not paid. Since each insurer has portfolio with each broker, the broker's reputation for enforcing (failing to enforce) penalty quickly becomes known to all insurers. If the broker is known not to apply a penalty, then the insurer makes future payment of $b_{br} = 0$.

Initially, each policyholder has a uniform prior that the incumbent broker, and rival brokers, will use its leverage over insurers to encourage settlement of non verifiable losses.¹² If the insurer fails to make a settlement and the incumbent broker then fails to extract a penalty from the insurer, the policyholder revises the prior downwards. This posterior is now lower than the common prior for rival brokers and the consumer switches brokers.

Now some proportion of these policyholders will suffer an non verifiable loss and the policyholders will have an expectations that the broker will facilitate a settlement of b_{br}^* . The value b_{br}^* is limited by the rents and the broker is expected to impose a penalty such that the insurer will weakly better off settling at this value. We have estimated the penalty under the assumption that all ex post bargaining power lodges with the policyholder. This is not necessary for the general structure of our results, but it permits considerable simplification. If the insurer fails to make this settlement and the broker fails to extract the penalty, the affected policyholders cancel their business and switch to another broker. Because the brokers compensation increases with the number of clients, then the broker has an incentive to present the penalties to the insurer and enforce them if the settlement is not reached. Anticipating the ex post disposition of the non-verifiable claims, the policyholder's ex ante problem is to secure, with the help of the broker, an incomplete contract with explicit and implicit coverage of c_{br}^* and b_{br}^* .

Now suppose that the size of the book of business, m , is large such that we can approximate the number of non-verifiable losses occurring in each period by the expected number, $mp(1 - q)$. The incentive constraint for the insurer to pay the transfer b is then

$$mp(1 - q)b \leq m \frac{P - pqc - p(1 - q)b}{r}$$

¹²Recall that we have assumed all non verifiable losses to be ex-post insurable. If there also were ex-post uninsurable losses, the broker would be called to use its judgement to distinguish and to use its hold-up power only for ex-post insurable losses. Further structure could be added to our model by allowing brokers to penalize policyholders who tried to falsify ex-post uninsurable claims to appear as though they were insurable.

where the right hand side represents the future rents for the insurer from retaining the entire book of business. Assuming policyholders have all the bargaining power, the net premium paid to the insurer is then

$$P = pqc + p(1 - q)b + p(1 - q)rb.$$

We observe that the implicit loading, $p(1 - q)rb$, is smaller than the implicit loading that is needed in the bilateral relationship without the broker, rb . Ex-ante the policyholder chooses the level of insurance coverage, c , of the verifiable loss and the transfer, b , of the non-verifiable loss to maximize expected utility subject to the premium structure above and the compensation paid to the broker. The compensation structure is linear in the insurer's expected profits, i.e.

$$k = \gamma p(1 - q)rb.$$

The overall payment for each policyholder can then be written as

$$\hat{P} = P + k = pqc + p(1 - q)b + (1 + \gamma)p(1 - q)rb.$$

Defining

$$\hat{r} = (1 + \gamma)p(1 - q)r$$

yields

$$\hat{P} = P + k = pqc + p(1 - q)b + \hat{r}b.$$

The premium structure is thus the same as in the bilateral case which leads to results that are qualitatively equivalent to Proposition 1.

Proposition 2 *In the market with the broker, it is optimal for the policyholder to partially insure both types of losses and to buy more coverage on the verifiable than to implicitly generate on non-verifiable event, i.e. $b_{br}^* < c_{br}^* < L$ for all r . Furthermore, there exists a critical discount rate $\bar{r}_{br} > 0$ such that it is optimal to generate a transfer if the discount rate is below \bar{r}_{br} and not to generate a transfer otherwise, i.e. $b_{br}^* > 0$ if $r < \bar{r}_{br}$ and $b_{br}^* = 0$ if $r \geq \bar{r}_{br}$.*

Next, we compare the optimal transfer b_{br}^* and coverage c_{br}^* generated through the broker with the ones

in a bilateral relationship, b^* and c^* . We observe that difference is based on a pure price effect. The overall implicit loading \hat{r} is strictly lower than in the bilateral case if and only if

$$(1 + \gamma)p(1 - q) < 1.$$

We thus have to examine the comparative statics of the transfer b and coverage c under changes of the implicit loading. Insurance can be a Giffen good under decreasing absolute risk aversion if the income effect outweighs the substitution effect. To abstract away from those wealth effects, we assume that the policyholder's preferences exhibit constant absolute risk aversion (CARA). Under this assumption, the following proposition shows that coverage for both verifiable and non-verifiable losses can be obtained more efficiently through the broker than under the bilateral case.

Proposition 3 *Suppose that policyholder's preferences exhibit CARA and that $(1 + \gamma)p(1 - q) < 1$. Then $\bar{r}_{br} > \bar{r}$. Furthermore, $b_{br}^* > b^*$ and $c_{br}^* > c^*$ for all $r < \bar{r}_{br}$ and $b_{br}^* = b^* = 0$ and $c_{br}^* = c^*$ for all $r \geq \bar{r}_{br}$.*

Proof. See Appendix A.2. ■

These results show that the brokers can play an important coordinating mechanism in securing the incomplete contract. In particular, by pooling risk and using the hold up power from their whole book of business, the broker can secure implicit coverage for the non-verifiable loss on more advantageous terms and this will lead to a higher values of b^* and c^* .

5 Summary and Comments on Contingent Commissions

The propositions are summarized in Figure 1. Without brokers, a transfer payment of the non verifiable loss, b^* , is negotiated but it is bounded by the future rents on the policy which might be small. The present value of these rents is of course interest rate sensitive, and the transfers are diminishing in the interest rate and eventually disappear. These rents are bestowed on the insurer (even under competition) to create the hold-up required for the ex post transfer for non verifiable losses. The downside to the rents is that the insurance on non-verifiable losses is actuarially unfair and therefore only partially insured, $b^* < L$. This creates a negative spill-over effect on insurance of verifiable losses. To reduce the gap in marginal utilities, it is optimal to also partially insure verifiable losses, $c^* < L$. As explicit insurance is relatively cheaper than

implicit insurance, more contractual insurance is bought than implicitly generated through rents, $c^* < b^*$.

With brokers, the “renewal rights” to a book of business are transferred to the broker under its brokerage contract with the insurer. This magnifies the hold-up which results in larger negotiated transfers of non-verifiable (but ex post insurable) losses as shown by b_{br}^* . Moreover, the interest domain over which transfers are made, is expanded from \bar{r} to \bar{r}_{br} . Not only do brokers enable more complete transfers of the non-verifiable loss, they permit an efficiency gain through diversification. This arises because the brokers can use the future rents from policyholders not having a loss, to expand its hold-up on behalf of those who do suffer a loss. The increased hold-up power reduces the implicit premium loading and thereby increases the transfers for non-verifiable losses, $b_{br}^* > b^*$. This, in turn, reduces the negative spill-over effect – the spread in marginal utilities is lower – and increases the transfers for verifiable losses, $c_{br}^* > c^*$. Thus, with brokers, more contractual insurance for verifiable losses is bought and this is accomplished with lower transaction costs imposed on the extra-contractual market for non-verifiable losses.

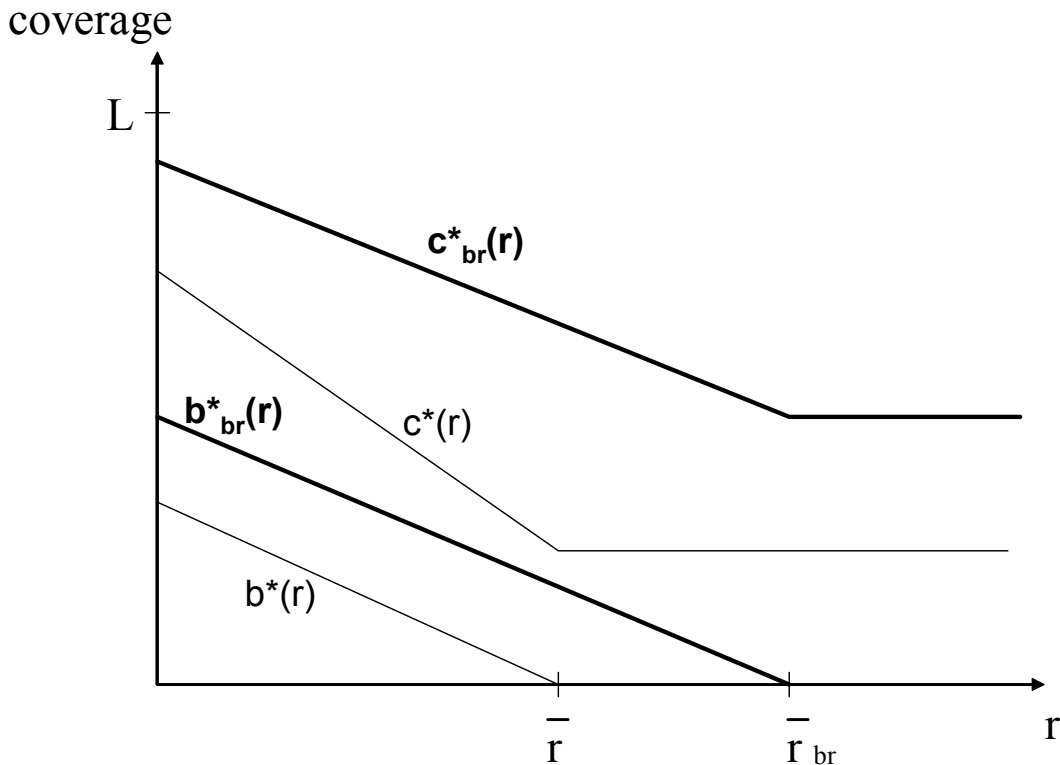


Figure 1

The role defined here for brokers raises some topical challenges. Brokers have recently been under a well publicized attack from the attorney general of New York, Elliot Spitzer. This attack has decimated the value of the world's largest broker and largest insurer, forced the resignation of both their CEO's, led to several criminal indictments, and has resulted in all four of the largest brokers abandoning a major form of compensation. At the heart of the issue is the compensation structure for brokers. In addition to commissions (proportional to premiums), brokers often receive "contingent commissions". These are payments by the insurer to the broker based on some measures of the book of business transacted. Most common are contingent commissions based on the profit of the broker's book, but volume based contingent commissions and renewal based structures also are found. The allegation of Spitzer, is as follows. The broker is an agent of the policyholder. Therefore, broker compensation linked to the welfare of the insurer must compromise its obligations to its principle (the policyholder).

Our results challenge this simplified conclusion. We have shown that brokers can benefit policyholders by expanding the market for non verifiable losses. To do so, they must exercise a credible threat to sanction insurers that do not respond to such losses. This is achieved by constructing a similar incentive compatibility constraint for the insurer and broker such that they both gain from the ex post transfer, i.e. that the broker shares in the future rents from the book it holds with the insurer. In our formal model, the broker compensation is linear in the rents. This contradicts Spitzer's blanket assertion that "kickbacks" from insurers are inimical to the interests of policyholders. While, of course, there are other dimensions to the contractual relationships not addressed here, and compensation also must address these, our model shows that contingent fees can lead to an expansion of insurance markets to include informal coverage of non verifiable losses and this is beneficial to policyholders.¹³

¹³There are other benefits from profit based contingent commissions. If brokers have better information about policyholder risk than insurers, brokers can send a signal to insurers about risk type and thereby mitigate adverse selection. Profit sharing with insurers will render the signal credible (see Cummins and Doherty, 2005).

A Appendix: Proofs

A.1 Proof of Proposition 1

The policyholder solves the following program

$$\max_{c,b} E[u(w)] = (1-p)u(w_0 - P) + pqu(w_0 - P - L + c) + p(1-q)u(w_0 - P - L + b)$$

subject to

$$P = pqc + p(1-q)b + rb.$$

The first derivative of expected utility with respect to b is

$$\frac{\partial E[u(w)]}{\partial b} = \left[\begin{array}{c} -(1-p)(p(1-q) + r)u'(w_0 - P) - pq(p(1-q) + r)u'(w_0 - P - L + c) \\ + p(1-q)(1-p(1-q) - r)u'(w_0 - P - L + b) \end{array} \right].$$

The second derivative is negative and expected utility is thus globally concave in b for all levels of c . The FOC thus determines the unique global maximum $b^* = b^*(c)$. Evaluating the first derivative at $b = L$ yields

$$\frac{\partial E[u(w)]}{\partial b} \Big|_{b=L} = -ru'(w_0 - P) + pq(p(1-q) + r)(u'(w_0 - P) - u'(w_0 - P - L + c)) < 0$$

for all $c \in [0, L]$, i.e. $b^*(c) < L$ for all $c \in [0, L]$ and therefore $b^* < L$. Evaluating the first derivative at $b = 0$ yields

$$\frac{\partial E[u(w)]}{\partial b} \Big|_{b=0} = \left[\begin{array}{c} -(1-p)(p(1-q) + r)u'(w_0 - P) - pq(p(1-q) + r)u'(w_0 - P - L + c) \\ + p(1-q)(1-p(1-q) - r)u'(w_0 - P - L) \end{array} \right].$$

For any $r \geq 1 - p(1 - q)$, we have

$$\frac{\partial E[u(w)]}{\partial b} \Big|_{b=0} < 0$$

for all $c \in [0, L]$, i.e. $b^*(c) = 0$ for all $r \geq 1 - p(1 - q)$. For $r = 0$ we get

$$\frac{\partial E[u(w)]}{\partial b} \Big|_{b=0, r=0} = \left[\begin{array}{c} (1-p)p(1-q)(u'(w_0 - P - L) - u'(w_0 - P)) \\ + pqp(1-q)(u'(w_0 - P - L) - u'(w_0 - P - L + c)) \end{array} \right] > 0$$

for all $c \in [0, L]$, i.e. $b^*(c) > 0$ for $r = 0$. Differentiating $\frac{\partial E[u(w)]}{\partial b} \Big|_{b=0}$ with respect to r yields

$$\frac{\partial}{\partial r} \frac{\partial E[u(w)]}{\partial b} \Big|_{b=0} = -(1-p)u'(w_0 - P) - pqu'(w_0 - P - L + c) - p(1-q)u'(w_0 - P - L) < 0.$$

This implies that for each c there exists a unique $\bar{r}(c) > 0$ such that $b^*(c) > 0$ for all $r < \bar{r}(c)$ and $b^*(c) = 0$ for all $r \geq \bar{r}(c)$.

The first derivative of expected utility with respect to c is

$$\frac{\partial E[u(w)]}{\partial c} = pq(-(1-p)u'(w_0 - P) + (1-pq)u'(w_0 - P - L + c) - p(1-q)u'(w_0 - P - L + b)).$$

The second derivative is negative and expected utility therefore globally concave in c for all levels of b . The FOC thus determines the unique global maximum $c^* = c^*(b)$. Evaluating the first derivative at $c = L$ yields

$$\frac{\partial E[u(w)]}{\partial c} \Big|_{c=L} = pqp(1-q)(u'(w_0 - P) - u'(w_0 - P - L + b)).$$

As $0 \leq b^* < L$, we have $\frac{\partial E[u(w)]}{\partial c} \Big|_{c=L} < 0$ and thus $c^* < L$. Evaluating the first derivative at $c = b$ yields

$$\frac{\partial E[u(w)]}{\partial c} \Big|_{c=b} = pq(1-p)(u'(w_0 - P - L + b) - u'(w_0 - P)).$$

As $0 \leq b^* < L$, we have $\frac{\partial E[u(w)]}{\partial c} \Big|_{c=b} > 0$ and thus $c^* > b^*$. Finally, define $\bar{r} = \bar{r}(c^*)$.

A.2 Proof of Proposition 3

The FOCs for $b^*(r)$ and $c^*(r)$ are

$$\left[\begin{array}{c} -(1-p)(p(1-q)+r)u'(w_0 - P(r)) - pq(p(1-q)+r)u'(w_0 - P(r) - L + c^*(r)) \\ +p(1-q)(1-p(1-q)-r)u'(w_0 - P(r) - L + b^*(r)) \end{array} \right] = 0$$

and

$$-(1-p)u'(w_0 - P(r)) + (1-pq)u'(w_0 - P(r) - L + c^*(r)) - p(1-q)u'(w_0 - P(r) - L + b^*(r)) = 0$$

where

$$P(r) = pqc^*(r) + p(1-q)b^*(r) + rb^*(r).$$

Implicitly differentiating both FOCs with respect to r yields

$$\left[\begin{array}{c} -P'(r) \left\{ \begin{array}{c} -(1-p)(p(1-q)+r)u''(w_0 - P(r)) - pq(p(1-q)+r)u''(w_0 - P(r) - L + c^*(r)) \\ +p(1-q)(1-p(1-q)-r)u''(w_0 - P(r) - L + b^*(r)) \end{array} \right\} \\ -c^{*'}(r)pq(p(1-q)+r)u''(w_0 - P(r) - L + c^*(r)) - pqu'(w_0 - P(r) - L + c^*(r)) \\ +b^{*'}(r)p(1-q)(1-p(1-q)-r)u''(w_0 - P(r) - L + b^*(r)) \\ - (1-p)u'(w_0 - P(r)) - p(1-q)u'(w_0 - P(r) - L + b^*(r)) \end{array} \right] = 0$$

and

$$\left[\begin{array}{c} -P'(r) \left\{ \begin{array}{c} -(1-p)u''(w_0 - P(r)) + (1-pq)u''(w_0 - P(r) - L + c^*(r)) \\ -p(1-q)u''(w_0 - P(r) - L + b^*(r)) \end{array} \right\} \\ +c^{*'}(r)(1-pq)u''(w_0 - P(r) - L + c^*(r)) - b^{*'}(r)p(1-q)u''(w_0 - P(r) - L + b^*(r)) \end{array} \right] = 0$$

Under the constant coefficient of absolute risk aversion R_a and using the FOCs we derive

$$\left[\begin{array}{c} R_a c^{*'}(r) pq(p(1-q)+r)u'(w_0 - P(r) - L + c^*(r)) \\ -R_a b^{*'}(r)p(1-q)(1-p(1-q)-r)u'(w_0 - P(r) - L + b^*(r)) - u'(w_0 - P(r) - L + c^*(r)) \end{array} \right] = 0$$

and

$$b^{*'}(r)p(1-q)u'(w_0 - P(r) - L + b^*(r)) = c^{*'}(r)(1-pq)u'(w_0 - P(r) - L + c^*(r))$$

The last equation implies

$$b^{*'}(r) = c^{*'}(r) \frac{(1-pq)u'(w_0 - P(r) - L + c^*(r))}{p(1-q)u'(w_0 - P(r) - L + b^*(r))}$$

and

$$\text{sign}(b^{*'}(r)) = \text{sign}(c^{*'}(r)).$$

Substitution into the first equation yields

$$c^{*'}(r) = -\frac{1}{R_a(1-p-r)}$$

We thus deduce that

$$\begin{aligned}\text{sign}(b^{*'}(r)) &= \text{sign}(c^{*'}(r)) < 0 \text{ for all } r < 1-p \\ \text{sign}(b^{*'}(r)) &= \text{sign}(c^{*'}(r)) > 0 \text{ for all } r > 1-p.\end{aligned}$$

As $b^{*'}(\bar{r}) < 0$, we have $\bar{r} < 1-p$ and thus $\text{sign}(b^{*'}(r)) = \text{sign}(c^{*'}(r)) < 0$ for all $r < \bar{r}$.

The implicit rate of interest in the market with the broker

$$\hat{r} = (1 + \gamma)p(1 - q)r$$

is lower than the one in the bilateral case which completes the proof.

References

- [1] Abraham, K.S., 2001, *The Insurance Effects of Regulation by Litigation*, Washington DC, Brookings Institution.
- [2] Anderlini, L., L. Felli and A. Postlewaite, 2003a, “Courts of Law and Unforeseen Contingencies”, *working paper*, University of Pennsylvania, Department of Economics.
- [3] Anderlini, L., L. Felli and A. Postlewaite, 2003b, “Should Courts Always Enforce What Contracting Parties Write”, *working paper*, University of Pennsylvania, Department of Economics.
- [4] Berliner, B., 1982, *Limits of Insurability of Risks*, Prentice-Hall.
- [5] Biglaiser, G., 1993, “Middlemen as Experts”, *RAND Journal of Economics* 24, 212-223.
- [6] Bond, E.W. and K.J. Crocker, 1997, “Hardball and the Soft Touch: the Economics of Optimal Insurance Contracts with Costly State Verification and Endogenous Monitoring Costs”, *Journal of Public Economics* 63, 239-264.
- [7] Crocker, K.J. and J. Morgan, 1998, “Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts”, *Journal of Political Economy* 106, 355-375.
- [8] Cummins, J.D. and N.A. Doherty, 2005, “The Economics of Insurance Intermediaries”, American Insurance Association.
- [9] Grossman, S.J. and O. Hart, 1986, “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration”, *Journal of Political Economy* 94, 691-719.
- [10] Hart, O., 1995, *Firms, Contracts, and Financial Structure*, Clarendon Press, Oxford.
- [11] Hart, O. and J. Moore, 1990, “Property Rights and the Nature of the Firm”, *Journal of Political Economy* 98, 1119-1158.
- [12] Jean-Baptiste, E.L. and A.M. Santomero, 2000, “The Design of Private Reinsurance Contracts”, *Journal of Financial Intermediation* 9, 274-297.
- [13] Kocherlakota, N.R., 1996, “Implications of Efficient Risk Sharing without Commitment”, *Review of Economic Studies* 63, 595-609.
- [14] Kimball, M.S., 1988, “Farmers’ Cooperatives as Behavior Toward Risk”, *American Economic Review* 78, 224-232.
- [15] Kingston, Ch., 2005, “Intermediation and Trust”, *working paper*, Department of Economics, Amherst College.
- [16] Rubinstein, A. and A. Wolinsky, 1987, “Middlemen”, *Quarterly Journal of Economics* 102, 581-593.
- [17] Stewart, R.E. and B.D. Stewart, 2001, “The Loss of Certainty Effect”, *Risk Management and Insurance Review* 4, 29-49.
- [18] Townsend, R., 1979, “Optimal Contracts and Competitive Markets with Costly State Verification”, *Journal of Economic Theory* 21, 265-293.
- [19] Wilder, J., 2002, “Competing for the Effort of a Common Agent: Contingency Fees in Commercial Lines Insurance”, *working paper*, MIT, Department of Economics.