Why changes in PBGC and FDIC premiums should not fully reflect changes in underlying risk (with some application to long-term private insurance contracts)

Abstract

The degree of risk adjustment in both FDIC and PBGC premiums appears to be much smaller than actuarially fair. We explore why this is using a stylized theoretical model of multi-period insurance contracts in the presence of moral hazard where the risk status of insureds changes over the life of the contract. We show how our model extends prior work on social insurance with moral hazard. If insureds value stable premiums, and there is moral hazard, the optimal multi-period insurance contract for full insurance allocates greater premiums to higher-risk states, and lower premiums to lower-risk states, but the optimal allocation of premiums across risk states will usually not be actuarially fair. The degree of risk adjustment rises with the extent of moral hazard and falls as risk aversion rises. We extend our analysis to examine optimal risk classification in private insurance in the presence of moral hazard, with similar results. We also discuss practical considerations which further reduce the desirability and feasibility of actuarially-fair risk adjustments in premiums for the FDIC and PBGC.

1. Introduction

In multi-period insurance contracts where information about the true risks of insureds are symmetrically revealed over time, a tension often arises between the concept of premium smoothness – that is, how the premium changes over time as the policy progresses and the underlying risk of the insured changes – and premium fairness – that is, how accurately the premium charged over any period compensates the insurer for the risk they are taking on over that period. On the one hand, insureds might legitimately prefer a premium that does not change much, if at all, over the life of the contract. Indeed, premiums that are extremely sensitive to underlying risk may render an insurance policy worthless to insureds.[1] On the other, insurers might well prefer to charge premiums at each point in time that fairly reflect the risk that they are taking on at that time. Reflecting risk in premiums at each point in time helps control adverse selection or adverse retention if commitment between the insurer and the insured is unilateral,[2] or if commitment is bilateral,[3] perceived unfairness. But also, as we will show, in the presence of moral hazard, adjusting premiums to reflect underlying risk more accurately helps reduce the overall cost of insurance by giving insureds an incentive to behave well, controlling moral hazard. This paper explores this tension, primarily in the context of public insurance programs such as the Federal Deposit Insurance Corporation (FDIC) and the Pension Benefit Guaranty Corporation (PBGC), but with some application to other insurance markets as well.

The tension between premium smoothness and premium fairness is best explained by referring to the model of McCarthy and Neuberger (2005b), developed in the context of

---

1 For example, few informed consumers would likely buy a whole life insurance policy that could underwrite them accurately every morning, and charged them a daily premium equal to that day's risk. Premiums for healthy individuals would be negligible, but as individuals neared death, their daily premiums would approach the sum assured under the policy.
2 By unilateral commitment, we mean that if insureds pay premiums, the insurer is obliged to honor the contract, but if insureds decline to pay the premiums, the insurer cannot compel them to do. Almost all private-sector insurance in the US has unilateral commitment.
3 By bilateral commitment, we mean that the insurer can compel the insured to pay premiums due in terms of the policy. Social insurance is usually bilateral.

the Pension Protection Fund (PPF), the UK equivalent to the PBGC. They show that the only stable multi-period premium structure which (1) can be introduced into a population with a varying mix of risks (2) charges each insured a premium in each period based only their risk in that period (so is not forward-looking or path dependent) and (3) which is actuarially fair for all insureds on an *ex ante* lifetime basis, and so embodies no *ex ante* cross-subsidies between different risks, is the premium that is actuarially fair for each insured in each period.

For this and other reasons, much academic work has focused on measuring what an actuarially-fair premium should be for the PBGC and the FDIC. Marcus (1985) used an option-pricing model to price pension insurance, a model extended by McCarthy and Neuberger (2005a) to incorporate systemic risk and the correlation between default rates and pension funding levels. Pennacchi (1987), Flannery (1991), Blair and Fissel (1991), Kendall (1992), Duan and Yu (1994) and Lee et al (2015) use similar approaches to price FDIC insurance under various assumptions.

The main contribution of this paper is to explore this issue in more detail: how and why should insurers adjust premiums in long-term contracts as the underlying risk of insureds changes? Even abstracting from the practical difficulties of adjusting premiums in a fully actuarially-fair way, which we will briefly discuss, we show that insurers should not even *aspire* to adjust premiums to ensure that they are actuarially fair at each point in time. Rather, we show theoretically that the optimal premium structure depends on three factors: (1) the extent to which moral hazard on the part of insureds affects their risk; (2) the extent to which insureds prefer smooth over variable premiums (which, in the case of the FDIC and the PBGC depends on their ability to pass changes in premiums on to customers and employees by changing prices and/or wages as well as other internal frictions, discussed below), and (3) whether the policy has bilateral or unilateral commitment (the FDIC and PBGC have bilateral commitment, but almost all private US insurance is unilateral).

Our work extends prior work on moral hazard in social insurance by examining the case where the probability of loss (rather than the claim amount) depends on moral hazard.

In an influential paper, Baily (1978) that the presence of moral hazard in claim amounts reduces the optimal insurance level, thereby creating a form of co-insurance that serves to control moral hazard.[4] Chetty (2006) extends this work by showing that similar results obtain in a much wider set of models. The core insight of these papers is that moral hazard in social insurance should use co-insurance to increase the degree of risk-sharing between insurer and insured in order to encourage good behavior, and that the degree to which risk should be shared between insurer and insured depends *inter alia* on risk aversion and the degree of moral hazard. In this paper, we extend this work to the case where moral hazard concerns the probability of loss (rather than the claim amount), and premiums may differ across different risk states – thus introducing reclassification risk into social insurance, and show that similar, but not identical, results obtain.

The closest any author has come to examining this issue in the social insurance literature is Hendren (2021), who shows that willingness-to-pay for insurance increases when allowance is made for uncertainty in information flows that may change the insured's risk type. However, he too does not explore the implications of his result for multi-period insurance policies when new information about the insured's risk is symmetrically revealed during the term of the contract.

We also apply our work to the case of private insurance, which, in the US has mainly unilateral commitment. This extends the work of Hendel and Lizzeri (2003) by including moral hazard. They, like us, examine a model of multi-period insurance contracts where there is learning about the risk-state of customers. The focus of their model is to explain how insurers use front-end loading to deal with the difficulties posed by unilateral commitment in the second period. They find that long-term contracts are optimal, and that additional information about changes in the insured's risk state is not fully reflected in premiums in later periods. Front-end loading of contracts is used to bond insureds to insurers in the case of asymmetric commitment, with the optimal

4 The author is grateful to an anonymous referee for pointing out the connection between the current paper and this strand of the economics literature.

degree of front-end loading differing by the degree to which insureds expect their income to grow. We find that moral hazard changes the amount of the bond that policyholders post with insurers: the greater the amount of moral hazard, the lower the bond posted in the first period.

Other previous work on optimal private insurance contracts under moral hazard differs substantially from our work. Most focus on one period. As summarized in Winter (2013), this work shows that where moral hazard affects the probability of a loss but not its amount, the optimal contract is full insurance with a deductible (Holmstron, 1979). Where moral hazard affects the distribution of the amount of a loss, but not its probability, the optimal contract is full insurance of small losses, but an increasing share of self-insurance as the amount of the loss rises (Rees and Wambach, 2008). Given these results, one approach could be to examine varying amounts of co-insurance or deductibles, where the amount of risk borne by the insurer falls in some way as the insured's risk deteriorates (e.g. by raising the deductible or by reducing cover at higher claim levels). Although these approaches have some theoretical appeal, in the case of the FDIC and the PBGC they are likely to be of limited practical use for reasons we discuss in section 2.

The paper proceeds as follows. In the next section, we discuss in more detail the case of the PBGC and FDIC, as this is a primary focus of this paper. We then present the model set-up, before examining the case of bilaterally enforceable contracts with and without moral hazard in the third section. The next section dispenses with the assumption of bilateral enforceability, and so is more applicable to private insurance contracts. The final section is a conclusion. To improve readability, all mathematical proofs are relegated to the appendix.

2. Application to FDIC and PBGC

The PBGC has been concerned about moral hazard almost since its inception. In the absence of PBGC insurance, Sharpe (1976) shows that rational employees would demand increases in cash wages to compensate them for increased riskiness in pension promises. The presence of the so-called 'PBGC put' weakens this link, allowing employers to make pension decisions that increase shareholder value at the expense of the PBGC. There are at least two sources of such moral hazard. First, employers have wide discretion about pension funding. Deliberately choosing to underfund pension plans increases the value of PBGC insurance by more than the extent to which the value of employee claims falls. Second, employers have wide discretion about the investment strategy followed in their pension funds. Choosing to invest in riskier assets would have the same effect. See Brown (2008) for a discussion.

Despite this concern, the PBGC has never charged pension fund sponsors premiums that differ based on their default risk. Initially, a constant premium was charged per member, supplemented from 1988 with a premium based on plan underfunding (so representing claim severity). But after much consideration of whether to base premiums on default risk, this has not so far been implemented.[5]

The FDIC provides a more interesting example. For nearly the first 60 years of its existence, the FDIC charged all banks the same premium (per $ of deposits), so, like the PBGC, incorporating some allowance for claim severity but none for claim frequency. Risk-based pricing was first introduced in 1993, largely in response to concerns about moral hazard and cross-subsidies between strong and weak banks (Fissel, 1994).[6] Table 1 shows premium rates and 5-year failure rates under various FDIC premium assessment systems since that time. It should be clear from the table that assessment rates are almost always much flatter than failure rates: under the 1996-2006 system, for less-than-adequately-capitalized banks, premium rates varied across supervisory sub-groups by a factor of 2.7, while failure rates between 1985 and 2000 varied over the same groups by a factor of 12.5.[7] Statutory reform in 2006 was followed by a substantial alteration to the premium system, but with a similar outcome: failure rates

5 The UK equivalent of the PBGC, the Pension Protection Fund (PPF), has used risk-based pricing since inception.
6 Among international deposit insurance systems, risk-rating is still rare; the FDIC is somewhat of an outlier in that it adjusts premiums for default risk at all. See Table A.1.4 of Demirgüç-Kunt et al (2005).
7 The figures are calculated by taking the assessment (failure) rate in the right-hand column of the table and dividing it by the assessment (failure) rate in the left-hand column of panel A (panel B) of Table 1 in the 'less than adequately capitalized' row.

between the riskiest and the safest banks varied by a factor of 4.6 but premium assessment rates only by a factor of 1.5.[8] While these calculations are approximate for various reasons (failure rates are only a proxy for claim amounts as claim severity may differ across risk categories; the time periods over which failures are measured do not coincide exactly with the premium structure at any point in time; individual banks move between assessment categories over time), the overall message of the table should be clear.

Before discussing the mathematics of the model in the next section, we comment on two aspects of the model that are related to our application of it to the PBGC and the FDIC. First, the FDIC and PBGC charge premiums to financial intermediaries (banks and pension fund sponsors) but provide cover to households. Our analysis abstracts from this issue by focusing on the financial intermediary's preference for smooth over variable premiums rather than household risk aversion. Although in practice this choice is uncontroversial,[9] in an academic setting it requires some justification. In a frictionless world, shareholders are risk-neutral and financial intermediaries could pass any changes in premiums directly to customers and employees. Modelling the insured as risk averse as we do relies on two frictions: first, Froot, Scharfstein and Stein (1993) describe imperfections creating incentives for corporate risk management (managerial incentives, convex tax schedules, asymmetric information leading to difficulties accessing external capital, among others), and second, imperfections in banking and employment markets that prevent intermediaries from passing on changes in premiums to their customers and/or employees. It is likely that both of these sources of imperfection rise as insured credit quality falls. We therefore model the varying impact of these frictions on intermediaries as risk aversion over premiums in the usual way. This has the further

8 The figures are calculated analogously to the method described in footnote 6, but using panels C and D.

9 In their final rule establishing the 2011 risk-based premium system, the FDIC (2011) highlighted that "… more stable and predictable effective assessment rates [are] a feature that industry representatives said was very important…" (page 43, with similar language again on page 59), indicating that in practice, banks attach significant weight to stable FDIC premiums. The PPF also expended significant effort in ensuring their premiums did not change too much in response to changes in credit-worthiness, suggesting that the same would be true for sponsors of DB pension funds and PBGC premiums.

advantage of increasing the generality of the model to settings other than the PBGC and the FDIC, which is one important reason we use it.[10]

Second, in response to moral hazard, insurers have other contract choices besides retaining full insurance and altering the premium as we assume here. For instance, they could raise the deductible or the degree of co-insurance in response to increases in the policyholder's risk. While such an approach may be acceptable in competitive private insurance markets, we argue that in the case of the FDIC and the PBGC it would be highly undesirable. One reason is that doing this would simply pass the *ex post* risk on to individual depositors and pension fund members. This may be less effective in controlling moral hazard than changing premium rates charged to financial intermediaries *ex ante*. In the case of FDIC insurance in particular, such an approach may even have the effect of undermining one of the primary *raisons d'être* of the FDIC (preserving financial stability) by increasing the possibility of bank runs as depositors became aware that their coverage limits had fallen. In fact, for precisely this reason, such a premium structure would violate one of the principles of deposit insurance laid out by the FDIC (FDIC, 2014), that the premium assessments be kept confidential. Also, both approaches are tantamount to withdrawing insurance precisely as the insured event becomes more likely. While this may be acceptable in competitive private insurance markets where individual insureds choose their cover type, to pension fund members and depositors where coverage is mandatory and there is no choice about insurance terms, this will – probably rightly – be viewed as inappropriate and self-defeating, given the public purposes that this insurance is intended to serve. We note that during the financial crisis, deposit insurance coverage limits actually *rose* in many countries, including the US and the UK, as deposit insurers strove to promote financial stability over this tumultuous period.

Another issue worth discussing is that our model requires each individual policy to be actuarially-fair. While it is evident that this criterion would apply to a private insurer,

10 Other choices are possible. One is a state-dependent value function of premiums, where the 'cost' per dollar of premiums to insureds rises as their risk rises. This is functionally equivalent to our model.

it is not obvious that the same should be true for public monopoly insurers such as the PBGC or FDIC. As we alluded to earlier, a great deal of academic work has focused on what an actuarially-fair premium should be in the case of the FDIC and the PBGC. There are two reasons for this. First, actuarial fairness reduces cross-subsidies between different classes of risk, preventing, for example, stronger risks from subsidizing weaker ones, and perceived unfairness of premiums on the part of policyholders of different financial strengths. Second, premiums that are not actuarially fair reduce market discipline acting on financial intermediaries, and may lead to aggregate misallocation of capital or other resources. Actuarial fairness across policyholders is therefore an ideal that the FDIC and PBGC should strive to meet.[11]

A final point worth noting is that it is not obvious that the FDIC or PBGC should aim to maximize the utility of the insured given that they are statutory monopolies (captured in the maximization problem we state formally in the next section). We argue that a benevolent public insurer should strive to maximize consumer surplus, which would be achieved by replicating the market outcome as modeled above. In fact, provided the policy is actuarially fair, the insurer has nothing to lose by following such a strategy, provided that it can borrow or lend costlessly, as we assume.

3. Model set-up

We now turn to discussing our mathematical model. It contains the following features. First: a risk-averse insured that can purchase a multi-period insurance policy to fully cover themselves against a peril that could arise in any or all periods. Second: the underlying risk of the peril to the insured fluctuates across different periods, in a manner that is symmetrically observable to both the insurer and the insured. This gives the insurer the ability to contract at time 0 on premiums by the risk of the insured as the contract progresses. Third: some element of moral hazard, that is, a hidden but

11 Looking ahead, we note in the conclusion that the results in this paper imply that an optimal premium structure that is neither forward nor backward-looking cannot be fully actuarially-fair on a lifetime basis across all policyholders simultaneously (as shown by McCarthy and Neuberger 2005b). The optimal premium structure must therefore embody some degree of unfairness, and the PBGC/FDIC must trade off this unfairness against premium smoothness in an aggregate sense. But because the tradeoff depends on the distribution of underlying risks in the first period, this is a much more complex problem than the simple one we examine here. We therefore leave consideration of this important point to future work.

costly action that the insured can undertake after the contract has been purchased to reduce the probability of the loss occurring. Finally, a risk-neutral insurer that is self-financing. We therefore abstract from issues not relevant to the topic under investigation: we assume a two period model, where the insured has no access to capital markets,[12] but the insurer may, and that interest rates and subjective discount factors are zero.[13]

2.1 Insured

We first discuss the insured. In each period, the insured faces the possibility of a loss of amount 1. The loss could occur in each period, in both or in neither. Losses in each period are independent. However, in the first period, the insured is in risk state M, and the probability of loss is p_M. In the second period, the insured's risk either rises to p_H (the high-risk state) or falls to p_L (the low-risk state), with $p_H > p_M > p_L$. The insured can choose the probability that they move to the low-risk state in period 2. We call this choice variable p. The probability that the insured moves to the high-risk state in period 2 is then $1-p$. In all periods, the insured's risk state and the underlying probabilities are public knowledge. Any choice of p has a cost $c(p)$, where $c(p)$ a known function with support $[\underline{p};\overline{p}]$ where $0<\underline{p}<\overline{p}<1$.[14] To ensure an interior solution, we assume that $c'(p)>0 \ \forall \mathrm{p} \in (\underline{p};\overline{p}]$ (so increasing the probability of entering the low-risk state is costly) and that $c''(p)>0 \ \forall \mathrm{p} \in (\underline{p};\overline{p}]$ (so there are decreasing returns to effort). To ensure interior solutions we assume that $c'(\underline{p})=0$ and $c'(\overline{p})=\infty$.

We assume that the insured has constant income in each period x, and no access to savings or borrowings, and that therefore any income in excess of the loss is consumed at the end of the period once the loss is known. With access to capital markets, the

12 Allowing the insured to access capital markets allow them to smooth *average* premiums over time but would not allow them to hedge against changes in risk-based premiums in each period.
13 If the discount factor equals the interest rate, the model outcomes will be unchanged regardless of their value. Differences between the discount factor and the interest rate primarily affect the balance of insurance premiums between the first and second periods, rather than the allocation of premiums between risk states in the second period.
14 In applying the model to FDIC and PBGC premiums, insured income may depend on the risk state (for instance, if depositors abandon a risky bank). We, however assume that the costly action is non-pecuniary (i.e. $c(p)$ lies outside the argument of the utility function). However, the insights would be exactly the same if we made the costly effort pecuniary, at the cost of more difficult algebra. We thank an anonymous referee for making this point.

insured would be able to save in anticipation of any average change in premiums, but would not be able to hedge against changes in state-dependent premiums. The insured has an additively separable lifetime utility function composed of felicity functions $u(\cdot)$ in each period, and no subjective discount factor. The insured is assumed to prefer more to less and to be risk-averse, so $u'(\cdot) > 0$ and $u''(\cdot) < 0$.

The expected lifetime utility of the insured without insurance is for a given p is then:

$$V_U(\{p_M;(p_L, p_H)\}; p) = [p_M + p.p_L + (1-p)p_H]u(x-1) \\ +[1-p_M + p(1-p_L) + (1-p)(1-p_H)]u(x) - c(p)$$

2.2 Insurer

We now discuss the insurer. In the most general possible case, that we simplify in subsequent sections for reasons that we will discuss, the insurer provides a two-period, bilaterally enforceable insurance contract to the insured for full insurance. By bilaterally enforceable, we mean that both parties enter into the contract in period 1 and cannot renege in period 2. Since, by assumption, the risk-state of the insured is costlessly observable to the insurer, the insurer can adjust the premiums charged in each state. We therefore denote the premium structure of the policy as $\{\pi_M;(\pi_L,\pi_H)\}$, where π_M is the premium in the medium-risk state in the first period, and (π_L,π_H) are the low- and high-risk-state premiums in the second period. We assume that the policy as a whole is actuarially fair (but not necessarily so in each state), by imposing a zero-profit or self-financing constraint on the insurer. Interest rates are assumed to be zero.

$$\pi_M + p.\pi_L + (1-p)\pi_H = p.p_L + (1-p)p_H + p_M\ ,$$

where the left-hand side represents the expected value of the premiums and the right-hand side the expected value of the claims. Note that although the insurer is unable to control the action of the insured, it is aware of the incentives faced by the insured, and can therefore infer the probability that the insured will enter each state.

The expected lifetime utility of the insured with full insurance for a given p is then

$$V_I((\pi_M;\{\pi_L,\pi_H\}); p) = u(x-\pi_M) + pu(x-\pi_L) + (1-p)u(x-\pi_H) - c(p)\,.$$

2.3 Optimisation problem

We now discuss the optimization problem. First, we assume that the insured chooses the optimal probability of ending in the low-risk state in period 2, in period 1. Without insurance, the problem of the insured is then:

$$V_U^*((p_M;\{p_L,p_H\})) \equiv \max_p V_U((p_M;\{p_L,p_H\});p).$$

With insurance the problem of the insured becomes:

$$V_I^*((\pi_M;\{\pi_L,\pi_H\})) \equiv \max_p V_I((\pi_M;\{\pi_L,\pi_H\});p).$$

The insurance company, on the other hand, wishes to choose a premium schedule $(\pi_M;\{\pi_L,\pi_H\})$ subject to a zero-profit constraint and a participation constraint, expressed as

$$V_I^*((\pi_M;\{\pi_L,\pi_H\})) \geq V_U^*((p_M;\{p_L,p_H\})).$$

The participation constraint simply records the fact that the insured's expected lifetime utility must improve when they purchase insurance. A competitive market in insurance policies would drive an insurer to maximize the insured's objective function, subject to the participation and zero-profit constraints, labeled 'IOP' for 'insurer optimization problem' and written as:

$$\max_{\pi_M,\pi_L,\pi_H} V_I^*((\pi_M,\{\pi_L,\pi_H\})) \quad s.t. \quad V_I^*((\pi_M,\{\pi_L,\pi_H\}) \geq V_U^*((p_M\{p_L,p_H\}))$$
$$\pi_M + \hat{p}.\pi_L + (1-\hat{p})\pi_H = p_M + \hat{p}.p_L + (1-\hat{p})p_H, \qquad \text{(IOP)}$$

where $\hat{p}$ is the value of p chosen by policyholders in response to the premium structure $(\pi_M;\{\pi_L,\pi_H\})$. $\hat{p}$ is therefore properly regarded as a function of the premium structure, and so could be written $\hat{p}(\pi_M;\{\pi_L,\pi_H\})$. Note that $\hat{p}$ is not a function of the probability of loss in the high and low-risk states, as, under the assumption of full insurance, the individual is insulated from any and all losses. To save notation, we suppress its arguments.

2.4 Moral hazard

We now discuss how we represent moral hazard in this model. While various approaches are possible, for reasons that we will demonstrate in a subsequent section we use the difference between the probabilities of loss in the high-risk and low-risk states, so $p_H - p_L$.

This difference measures the degree to which information about the insured's risk is revealed in a way that is observable to the insurer for rating purposes in period 2, and to the insured when evaluating different insurance policies. For instance, if p_H and p_L are very close to one another, not much uncertainty is removed by the new information that is revealed in period 2. Further, given a function $c(p)$, the potential for moral hazard is then small because any action exerted by the insured will only have a small effect on the probability of loss. On the other hand, if p_H and p_L are far apart, the new information revealed in period 2 is very informative for rating purposes, and, for a fixed $c(p)$, the potential for moral hazard is larger. In the extreme case, if $p_H - p_L = 1$, then insurance is redundant in the second period because all risk is resolved at the end of the first period.

Having laid out the basic structure of our model, we now apply it first to the case of bilaterally enforceable contracts in the next section, and then to the case of unilaterally-enforceable contracts in the subsequent one.

4. Bilaterally enforceable contracts

In this section, we simplify the general model presented in the previous section by setting $\pi_M = p_M$. We do this for two reasons. First, allowing π_M to vary away from p_M introduces the possibility that the insurer can spread cost across time periods. In the case of bilaterally-enforceable contracts, this complicates interpretation of the analysis without altering any of the fundamental insights of the model. Second, restricting the analysis to only the premiums in the high and low-risk states allows us to derive very similar results to those found by Baily (1978) and Chetty (2006).

Proposition 1: If prudence and higher-order terms in the utility function are sufficiently small that they can be disregarded,[15] *and relative risk aversion is locally constant, the following approximate relationship holds:*

$$\xi \frac{(\pi_H - \pi_L)}{c_H} = \varepsilon_{q,\pi_L - p_L} ,$$

where c_H *is the individual's consumption in the high-risk state,* ξ *is the coefficient of relative risk aversion and* $\varepsilon_{q,\pi_L - p_L} \equiv \frac{d\log(q)}{d\log(\pi_L - p_L)}$ *is the elasticity of the ratio of the probability of being in the low-risk to the high-risk state in the second period w.r.t. the excess premium paid in the low-risk state.*

Corollary 1.1: In the absence of moral hazard, the optimal premiums in the high-risk and low-risk states are equal.

Corollary 1.2: The more risk averse the customer, the lower the optimal premium dispersion between high and low-risk states in the second period, ceteris paribus.

Corollary 1.3: The greater the moral hazard (in the sense of the difference between p_H *and* p_L*), the greater the optimal premium dispersion in the second period, ceteris paribus.*

Proposition 1 is a reformulation of Chetty (2006) and Baily (1978) in the case where moral hazard concerns the probability of a loss (rather than the amount, as they model). They found that where moral hazard concerns the severity of a loss, the percentage change in consumption multiplied by the coefficient of relative risk aversion equaled the elasticity of the claim amount w.r.t. to the level of insurance. Proposition 1 shows that where moral hazard is around the probability of the claim, what matters is the elasticity of the odds ratio w.r.t the excess premium paid in the low-risk state (which is not the elasticity in the total expected claim, which is the result of Chetty (2006) and Baily (1978)). But the optimal degree of co-insurance still depends on the

15 An equivalent result that allows for prudence is presented in Appendix B. We note that Chetty (2006) also used a second (third) order expanstion of the utility function to prove his results.

risk aversion of the insured, with the higher the degree of risk aversion, the lower the optimal degree of co-insurance because the insured is willing to pay a higher premium (caused by greater moral hazard) to preserve smoother premiums – and hence smoother consumption –in the second period.

Corollary 1.1 is a restatement of a well-known result in insurance: the optimal premium schedule is to charge an equal premium in all risk states in the second period. We emphasize it only because in our context, it has some surprising implications. If there is no moral hazard and contracts are bilaterally enforceable, the insurer optimally *disregards* the emergence of known information about the true risk of the insured when setting premiums in the second period. The insured, on the other hand, willingly enters into a contract in period 1 where they agree to pay *more* than the actuarially-fair premium in the low-risk state in period 2, but *less* than the actuarially-fair premium in the high-risk state in period 2. In effect, the optimal long-run insurance contract includes full insurance against any reclassification risk in the second period, and spreads risk fully between different risk states in the second period. This contract improves their lifetime utility relative to a contract that includes reclassification risk, and which is therefore actuarially-fair in all states of the world. Of course, the insurer is willing to sell such a contract only because it is enforceable in the low-risk state in the second period, and because, by assumption, it can spread premiums costlessly across risk states. Such a contract is preferred by insureds over a series of one-period contracts where they pay the actuarially-fair premium in each state, so long-term contracts are preferred to short-term ones.

Insert Figure 1 near here.

Figure 1 illustrates the relationship in the second period between the premiums of the optimal contract and the underlying risk. The probability of loss in each state is shown along the horizontal axis, and the premiums are shown on the vertical axis. Actuarially-fair premiums in each state, shown as AF_L and AF_H, must lie on the 45-degree line. The optimal premiums in each state, where there is no moral hazard, and contracts are enforceable, denoted A_L and A_H lie on a horizontal line. This is a consequence of the nature of this insurance policy, spreading risk fully between different states, and different time periods.

Corollary 1.2 shows that as the coefficient of risk aversion rises, the optimal spread of premiums in the second period must fall, *ceteris paribus.* And corollary 1.3 shows that as moral hazard rises (in the sense of the difference between the probability of loss in the high- and low-risk states), then the premium spread across risky states must rise. Figure 2 shows the optimal allocation of premiums across risk states in the case where there is moral hazard. Premiums are lower in the low-risk state, and higher in the high-risk state, but smoother than actuarially fair. The higher risk-aversion, the closer the premiums in the two states, and the further from actuarially-fair. The greater moral hazard, the closer premiums are to actuarially-fair, but the less equal they are.

Insert Figure 2 near here.

Corollaries 1.2 and 1.3 raise an interesting question: could moral hazard be so great that the the optimal premium spread in the second period is greater than actuarially fair? The next result shows that this is not the case: the maximum premium dispersion is actuarially-fair.

Proposition 2: If the insured would prefer the actuarially-fair contract $\{(\pi,\pi)\}$ *over remaining uninsured, then the optimal contract offered by the insurer has* $\hat{\pi}_L < \hat{\pi}_H$ *. Furthermore, the optimal premium spread between high-risk and low-risk states in the second period is less than actuarially fair, that is* $\hat{\pi}_H - \hat{\pi}_L < p_H - p_L$ *.*

Proof: See appendix.

The participation constraint deserves some discussion. In the absence of moral hazard, the participation constraint would be guaranteed due to the fundamental theorem of insurance. However, moral hazard introduces the possibility that the insured would be better off remaining uninsured, particularly if $p_H \gg p_L$, and would prefer to engage in some other form of risk management, such as avoidance. We disregard these cases as degenerate in the proposition by assuming that the insured would prefer the level contract to remaining uninsured.

We have now shown that the optimal premium structure when moral hazard is present is to create some classification risk in the second period by introducing a spread between the premium in the high risk and the low-risk states. Risk is still spread across different risk states, but not completely. The insurer no longer entirely disregards information about changes in the insured's risk state, but does not incorporate it fully into premiums either. The dispersion in the premiums gives the insured an incentive to undertake costly effort to reduce the probability of loss, and, under the assumption that $c'(\underline{p})=0$, the insured willingly purchases such insurance over insurance with level premiums because it reduces the overall cost of insurance. In effect, re-rating premiums is a form of co-insurance, sharing risk between the insured and insurer.

Our assumption of full insurance allows this work to be related to work on moral hazard in single-period contracts in an interesting way. Even though our model has the *probability* of a claim changing over time, the assumption of full insurance turns the moral hazard into moral hazard regarding the *amount* of premium that should be charged in the second period. The multi-period contract has two pieces: insurance against risk in the first period, and insurance against changes in the amount of the premium for full insurance in the second. As shown by Rees and Wambach (2008), the optimal contract for this type of risk with moral hazard is an increasing degree of co-insurance as the amount of the claim rises. This is equivalent to our finding that the premium in the second period should be risk-adjusted, but not fully.

5. Where contracts are only unilaterally enforceable

Up to this point, we have assumed that contracts are bilaterally enforceable, given our focus on FDIC and PBGC premiums. For completeness, however, we now turn to the case of unilateral commitment, which holds in most private-sector insurance contracts. Here, the insured can stop paying premiums at any time, in which case the insurance lapses and cover ceases, but the insurer is bound to offer cover if the premiums are paid as agreed. Note that the insurer still sells a two-period contract in period one, in which the policyholder agrees to pay the premiums laid out in the policy (which may depend on the observed risk state in the second period). Policyholders who approach the insurer in the second period seeking cover just for that period need not be offered the same terms.

We model asymmetric enforceability by adding further state-contingent participation constraints in the second period. We now require that in each state, the premium should be less than or equal to the actuarially-fair premium in that state. Hence, $\hat{\pi}_L \leq p_L$ and $\hat{\pi}_H \leq p_H$. Under these conditions, reneging on this contract in the second period and purchasing actuarially-fair single-period insurance from an alternative provider will never be cheaper. It is therefore never optimal for the insured to renege. However, under these assumptions, $\hat{\pi}_M \geq p_M$ to ensure that the contract is actuarially-fair. We therefore need to abandon our assumption that the insurer cannot redistribute resources across time. But having done so, we can write the next proposition.

Proposition 3: If prudence and higher-order terms in the utility function are sufficiently small that they can be disregarded, and relative risk aversion is locally constant, the following approximate relationship holds:

$$\xi \frac{(\pi_H - \pi_M)}{c_H} = \frac{d \log(1-p)}{d \log(\pi_M - p_M)} = \varepsilon_{1-p, \pi_M - p_M} \ ,$$

where ξ *is the coefficient of relative risk aversion, and* $\varepsilon_{q, \pi_M - p_M} = \dfrac{d \log(q)}{d \log(\pi_M - p_M)}$ *is the elasticity of the odds ratio w.r.t. the difference between the premium and the probability of loss in the first period.*

Corollary 3.1: In the absence of moral hazard, the optimal premiums in the high-risk state and the first period are equal.

Corollary 3.2: The more risk averse the customer, the lower the optimal premium dispersion between the high-risk state and the first period, ceteris paribus.

Corollary 3.3: The greater the moral hazard (in the sense of the difference between p_H *and* p_M*), the greater the premium dispersion between the first period and the high-risk state in the second period, ceteris paribus.*

Proof: See appendix.

Like Proposition 1, proposition 3 is an alteration of the formula of Baily (1978), in this case for where commitment in the second period is unilateral. Here, however, it shows that the optimal change in consumption between the first time period and the high-risk state in the second period is a function of the coefficient of relative risk aversion and the elasticity of the probability of being in the high-risk state wrt the excess premium paid in the first period.

Corollary 3.3 shows that the present of moral hazard raises the size of the bond that the policyholder willingly posts in the first period. This is an extension of the case of Hendel and Lizzeri (2003), who examine the case where moral hazard does not exist, but the probabilities of loss change over the period of the policy.

Once bilateral enforceability is no longer possible in the second period, then the optimal contract has the insured overpay in the first period, creating a bond between the insured and the insurer. In the second period, if the low-risk-state occurs, the insured pays an actuarially-fair premium, which is equal to the premium they could obtain if they reneged. But if the high-risk state occurs, the premium they would pay is lower than the actuarially-fair premium, representing a return of the bond.

Proposition 4: If the insured would prefer the actuarially-fair but level contract $\{(\pi,\pi)\}$ *over remaining uninsured, then the optimal contract offered by the insurer has* $\hat{\pi}_L = p_L$, $\hat{\pi}_H < p_H$ *and* $\hat{\pi}_M > p_M$. *Furthermore, this contract is preferred to a series of successive one-period contracts with actuarially-fair premiums in each state.*

Proof: See appendix.

Although we stress again that unilaterally-enforceable contracts are not the primary focus of this paper, in Tables 2 and 3 we briefly summarize the theoretical implications of our model and show how these match observed rating structures in multi-period insurance policies of different types.

Insert Tables 2 and 3 near here.

6. Discussion and conclusion

We have shown that under bilaterally enforceable contracts, insurers should reclassify risks over the life of multi-period contracts as a way of controlling moral hazard. The optimal rating structure depends on the ease of moral hazard, and the risk aversion of the insured. The greater the aversion of the insured to variation in premiums, the smoother premiums should be and the more the insurer should ignore specific information about the risk of the insured in the second period when setting premiums. The greater the extent of moral hazard, the greater the premium dispersion in the second period. Insureds prefer this structure to a level premium despite their aversion to changes in premiums because it lowers the overall cost of insurance plus the cost of controlling moral hazard. The conclusion of our model is that actuarially-fair premium risk adjustments charged to insureds with different underlying risks in multi-period contracts will be optimal only under very specific circumstances that likely do not hold in practice. Where work has examined the dynamics of insurance over multiple periods, it has typically assumed that the underlying loss distribution (other than any effects of moral hazard) is stationary, and consequently finds that long-term contracts offer no advantage over short-term contracts in dealing with moral hazard. Where the underlying loss distribution is not stationary, however, we find that long-term contracts are preferred to a series of short-term ones. Our work raises interesting questions as to whether risk classification in social insurance might be generally welfare-enhancing where there is moral hazard.

Our results raise the possibility that making insurance contracts bilaterally enforceable may improve insured welfare even in private insurance markets. The reason is that insureds may be willing to sacrifice the option to seek cheaper insurance in the second period in exchange for the smoother premiums that only bilateral commitment can provide. This will be especially valuable where the possibility of substantial changes to premiums exists. Exploring the implications of our model for the desirability of unilateral commitment will be the focus of future work.

While our model provides one justification for PBGC and FDIC premium structures that do not adjust premiums in an actuarially-fair way as the underlying risk of policyholders changes, there are other important reasons why this should be the case. First, risk-adjusted premiums will result in counter-cyclical premiums, both at the level of individual banks and pension fund sponsors, and at an aggregate level. Counter-cyclical premiums imply that individual banks and the banking industry as a whole will be faced with demands for higher premiums at precisely the time that they are least able to bear the cost, and when these higher premiums are most likely to restrict lending, and this lending is most likely to be valuable to their customers. In fact, as shown in FDIC (2011), page 13, chart 1, the risk-adjusted premium structures adopted by the FDIC after 1993 (FDIC, 2020) led to highly countercyclical premiums during banking crises. PBGC premiums, although not directly risk-adjusted are probably also counter-cyclical due to the correlation between corporate defaults, the level of the equity market and pension funding, as McCarthy and Neuberger (2005a, 2005b) make clear in the context of the PBGC's UK equivalent, the PPF. The undesirability of counter-cyclical premiums provides a further justification for assuming that financial intermediaries prefer premium structures that are stabler and therefore not fully actuarially-fair. Of course, reducing the spread of premiums away from actuarially fair, as suggested by our model, reduces this counter-cyclicality but does not eliminate it fully.

A second reason why fully risk-adjusting premiums is undesirable is that charging actuarially fair premiums to the riskiest banks and pension fund sponsors may increase the likelihood of bank or sponsor failure – precisely the event that the insurance system is designed to mitigate – and hence actuarially-fair premiums may simply be uncollectable in practice. For this reason, subsidies to the riskiest insureds may be unavoidable. For the FDIC and PBGC to remain solvent, they must therefore collect higher than actuarially-fair premiums during stable times in order to pre-fund the costs of defaults later – precisely the optimal premium structure described by our model. Both of these issues are discussed in FDIC (2020).

A final point worth noting is that our model assumes that premium risk adjustment is the only way insurers have to control moral hazard. In the case of the FDIC and PBGC, this is not the case. For banks, prudential regulation (including capital requirements and investment regulation) and regular inspections and monitoring are important tools to control moral hazard. In the case of the PBGC, pension funding and investment regulations, as well as provisions that allow the PBGC to involuntarily terminate a plan (under certain circumstances) serve the same function.

But these practical concerns simply strengthen the conclusion of this paper, which shows that PBGC and FDIC premiums should not be fully risk-adjusted in the presence of moral hazard. A necessary consequence of this choice, however, as McCarthy and Neuberger (2005b) show, is undesirable *ex ante* cross-subsidies between insureds of different quality at the time the system is introduced. The optimal degree of risk adjustment is thus a trade-off between the efficiency and other costs of these cross-subsidies and the cost to insured of variation in premiums, and the benefits of lower moral hazard, leading to lower claims and therefore lower premiums in the long term. Fully assessing this trade-off requires us to move from this two-period model to a dynamic model of the entire portfolio of risks, too long to be accommodated here, and which we therefore leave for future work.

7. References

Baily, M., 1978, "Some aspects of optimal unemployment insurance", *Journal of Public Economics,* 10: 379-402.

Blair CE and GS Fissel, 1991, "A Framework for Analyzing Deposit Insurance Pricing", *FDIC Banking Review,* 25, pp 25-36.

Brown, Jeffrey R., 2008 "Guaranteed Trouble: The Economic Effects of the Pension Benefit Guaranty Corporation", *Journal of Economic Perspectives*, 22(1), Winter: 177-198.

Chetty, Raj, 2006, "A general formula for the optimal level of social insurance", *Journal of Public Economics,* 90: 1879-1901.

Demirgüç-Kunt, A, Baybars Karacaovali and Luc Laeven, 2005, "Deposit Insurance around the World: A Comprehensive Database", World Bank Policy Research Working Paper 3628, June.

Duan, Jin-Chuan and Yu, Mon-Te, 1994, "Forbearance and pricing deposit insurance in a multi-period framework", *Journal of Risk and Insurance*, 61(4) pp 575-591

FDIC, 2011, Final Rule on Assessment*, 12 CFR 327, https://www.fdic.gov/news/board/2011rule1.pdf*

FDIC, 2014, "Core principles for effective deposit insurance schemes", https://www.fdic.gov/bank/international/iadi/coreprinciples11-2014.pdf

FDIC, 2020, "A History of Risk-Based Premiums at the FDIC", Report No. 2020-01.

Fissel, GS, 1994, "The Risk-Adjusted Premium System of the FDIC", *FDIC Banking Review,* 28, pp 25-36.

Flannery, Mark J, 1991, "Pricing deposit insurance when the insurer measures bank risk with error", 15 pp 975-998.

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein. "Risk management: Coordinating corporate investment and financing policies." *Journal of Finance*, 48.5 (1993): 1629-1658

Hendel I and A Lizzeri, 2003, "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance", *The Quarterly Journal of Economics*, 118(1) pp 299–328.

Hendren, Nathaniel, 2021, "Measuring Ex Ante Welfare in Insurance Markets, *The Review of Economic Studies*, 88(3): 1193–1223.

Holmström, Bengt, 1979, "Moral Hazard and Observability", *The Bell Journal of Economics*, Vol. 10, No. 1 (Spring, 1979), pp. 74-91

Kendall, Sarah B, 1992, "A note on the existence and characteristics of fair deposit insurance premia", *Journal of Banking and Finance,* 16 pp 289-297.

Lee, Shih-Chen, Lin, Chien-Ting and Tsai, Ming-Shann, 2015, “The pricing of deposit insurance in the presence of systemic risk”, *Journal of Banking and Finance*, 51, pp 1-11.

Marcus, Alan, 1985, "Spinoff/terminations and the value of pension insurance," *Journal of Finance*, 40(3): 911-924.

McCarthy, David G. and Anthony Neuberger, 2005a, “Pricing Pension Insurance”, Fiscal Studies, Volume 26, Issue 4, pages 471-489.

McCarthy, David G. and Anthony Neuberger, 2005b, “The Pension Protection Fund”, Fiscal Studies, Volume 26, Issue 2, pages 139-167

Pennacchi, George G., 1987, Á re-examination of the Over- (or under-) pricing of deposit insurance, *Journal of Money, Credit and Banking,* 19 (3) pp 340-360

Rees, Ray and Achim Wambach, 2008, “The Microeconomics of Insurance”, Foundations and Trends in Microeconomics, 2008, vol. 4, issue 1–2, 1-163

Winter, R.A., 2013. “Optimal insurance contracts under moral hazard.” In *Handbook of insurance* (pp. 205-230). Springer, New York, NY.

8. Figures and Tables

FIGURE 1: OPTIMAL PREMIUM ALLOCATION WHERE CONTRACTS ARE ENFORCEABLE AND THERE IS NO MORAL HAZARD. PREMIUMS ARE EQUAL IN EACH STATE

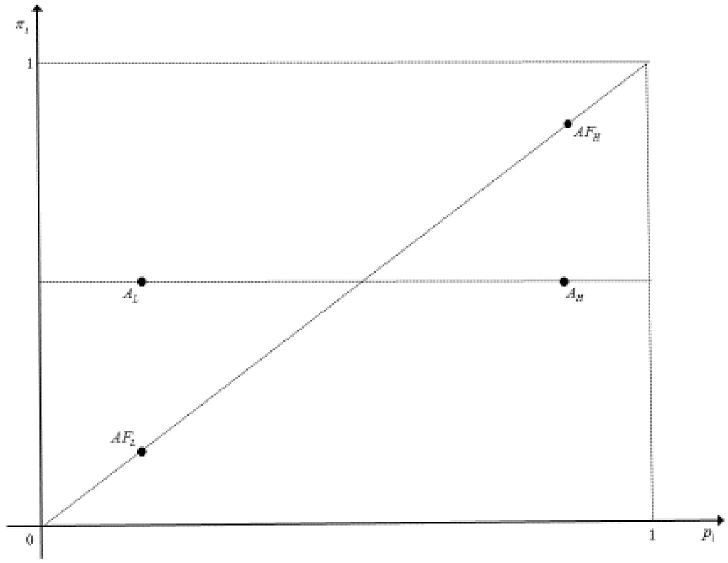

FIGURE 2: OPTIMAL PREMIUM ALLOCATION WHERE CONTRACTS ARE ENFORCEABLE AND THERE IS MORAL HAZARD. PREMIUMS ARE LOWER IN THE LOW-RISK STATE AND HIGHER IN THE HIGH-RISK STATE, BUT ARE SMOOTHER THAN ACTUARIALLY-FAIR.

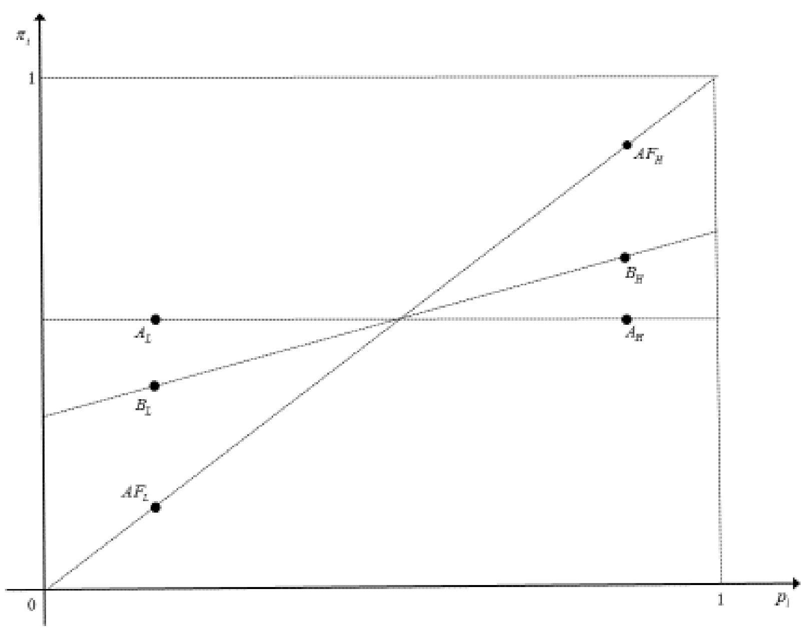

TABLE 1: FDIC ASSESSMENT RATE SCHEDULES AND 5-YEAR FAILURE RATES

Panel A: Assessment rate schedule: 1996-2006 (once fund ratio reached 1.25%)

	SUPERVISORY SUBGROUP		
Capital group	Healthy (CAMELS 1 or 2)	Supervisory concern (CAMELS 3)	Substantial supervisory concern (CAMELS 4 or 5)
Well capitalized	0 b.p.†	3 b.p.	17 b.p.
Adequately capitalized	3 b.p.	10 b.p.	24 b.p.
Less than adequately capitalized	10 b.p.	24 b.p.	27 b.p.

Panel B: Five-year failure rates by assessment category: 1985-2000

	SUPERVISORY SUBGROUP		
Capital group	Healthy (CAMELS 1 or 2)	Supervisory concern (CAMELS 3)	Substantial supervisory concern (CAMELS 4 or 5)
Well capitalized	0.77%	2.67%	6.78%
Adequately capitalized	2.03%	5.51%.	14.43%
Less than adequately capitalized	2.30%	7.10%	28.84%

Panel C: Assessment rate schedule: 2007-2008

	SUPERVISORY SUBGROUP		
Capital group	Healthy (CAMELS 1 or 2)	Supervisory concern (CAMELS 3)	Substantial supervisory concern (CAMELS 4 or 5)
Well capitalized	5-7 b.p.	10 b.p.	28 b.p.
Adequately capitalized	10 b.p.	10 b.p.	28 b.p.
Less than adequately capitalized	28 b.p.	28 b.p.	43 b.p.

Panel D: Five-year failure rates by assessment category: 2007-2012

	SUPERVISORY SUBGROUP		
Capital group	Healthy (CAMELS 1 or 2)	Supervisory concern (CAMELS 3)	Substantial supervisory concern (CAMELS 4 or 5)
Well capitalized	1.98%*	4.82%	13.09%
Adequately capitalized	4.82%	4.82%	13.09%
Less than adequately capitalized	13.09%	13.09%	60.60%

NOTE: † The FDIC was prevented by statute from charging banks in this category any premium at all once the fund ratio reached 1.25%. * Reported figure is an average across four quartiles of risk within risk category I. Assessment rates between 1993 and 1995 varied from 23 b.p.'s for the lowest risk category to 31 b.p.'s for the highest risk category and so are not shown. Source: FDIC (2020) Tables 2-5.

Table 2: Multi-period insurance contracts in the presence of reclassification risk

	Unilateral commitment	Bilateral commitment
Larger moral hazard affecting reclassification risk	Health insurance Auto insurance Home insurance	DB pension insurance (PBGC) Deposit insurance (FDIC) Insurance guarantee funds US Private Mortgage Insurance
Smaller moral hazard affecting reclassification risk	Term life insurance Whole life insurance	OASDI

Table 3: Multi-period insurance contracts in the presence of reclassification risk: model predictions

	Unilateral commitment	Bilateral commitment
Larger moral hazard affecting reclassification risk	Front-end-loaded contracts. Large risk-adjustment in low-risk states. Some risk-adjustment in high-risk states.	Limited risk adjustment in low and high-risk states, but less than actuarially fair.
Smaller moral hazard affecting reclassification risk	Front-end-loaded contracts. Large risk adjustment in low-risk states and little or no risk-adjustment in high-risk states.	No risk adjustment in high or low-risk states.

Appendix A

Proposition 1: If prudence and higher-order terms in the utility function are sufficiently small that they can be disregarded,[16] *and relative risk aversion is locally constant, the following approximate relationship holds:*

$$\xi \frac{(\pi_H - \pi_L)}{c_H} = \varepsilon_{q,\pi_L - p_L},$$

where c_H *is the individual's consumption in the high-risk state,* ξ *is the coefficient of relative risk aversion and* $\varepsilon_{q,\pi_L - p_L} \equiv \frac{d\log(q)}{d\log(\pi_L - p_L)}$ *is the elasticity of the ratio of the probability of being in the low-risk to the high-risk state in the second period w.r.t. the excess premium paid in the low-risk state.*

Corollary 1.1: In the absence of moral hazard, the optimal premiums in the high-risk and low-risk states are equal.

Corollary 1.2: The more risk averse the customer, the lower the optimal premium dispersion between high and low-risk states in the second period, ceteris paribus.

Corollary 1.3: The greater the moral hazard (in the sense of the difference between p_H *and* p_L*), the greater the optimal premium dispersion in the second period, ceteris paribus.*

Proof:

The policyholder's problem, given a premium structure is:

$$V(\pi_L, \pi_H) = \max_{\{c_L; c_H; p\}} pu(c_L) + (1-p)u(c_H) - c(p) \text{ s.t. } x - c_L \geq \pi_L \text{, and } x - c_H \geq \pi_H$$

The insurer's problem is to choose a premium structure that maximizes consumer utility:

$$\max_{\{\pi_L; \pi_H\}} V(\pi_L, \pi_H) \text{ s.t. } (1-p)p_H + p.p_L = (1-p)\pi_H + p.\pi_L ,$$

16 An equivalent result that allows for prudence is presented in the appendix.

recognizing that p is a function of the chosen premium structure. At an interior optimum, the optimal premium structure must satisfy

$$\frac{dV}{d\pi_L}(\hat{\pi}_L) = 0 \,,$$

recognizing that π_H is a function of π_L because of the budget constraint, and writing V using Lagrange multipliers as

$$V(\pi_L) = \max_{\{c_L;c_H;\lambda;\gamma_L;\gamma_H\}} pu(c_L) + (1-p)u(c_H) - c(p) + \gamma_L(x - c_L - \pi_L) + \gamma_H(x - c_H - \pi_H)$$

$$\frac{dV}{d\pi_L} = -\gamma_L - \gamma_H \frac{d\pi_H}{d\pi_L} = 0$$

$$\frac{d\pi_H}{d\pi_L} = -\frac{\gamma_L}{\gamma_H}$$

but $\gamma_L = pu'(c_L)$ and $\gamma_H = (1-p)u'(c_H)$ so

$$\frac{d\pi_H}{d\pi_L} = -\frac{pu'(c_L)}{(1-p)u'(c_H)} = -q\frac{u'(c_L)}{u'(c_H)} \text{, where } q = \frac{p}{1-p} \,,$$

the odds ratio of the low-risk to the high-risk state.

From the budget constraint, we have:

$$\pi_H = p_H + q(p_L - \pi_L) \text{, so } \frac{d\pi_H}{d\pi_L} = (p_L - \pi_L)\frac{dq}{d\pi_L} - q \,.$$

Combining these two gives:

$$q\frac{u'(c_L)}{u'(c_H)} = q + (\pi_L - p_L)\frac{dq}{d\pi_L}$$

Writing $u'(c_L) = u'(c_H) + u''(c_H)(c_L - c_H)$ and remembering that $c_L - c_H = \pi_H - \pi_L$ gives

$$q\left(1 + \frac{u''(c_H)}{u'(c_H)}(\pi_H - \pi_L)\right) = q + (\pi_L - p_L)\frac{dq}{d\pi_L} = q + (\pi_L - p_L)\frac{dq}{d(\pi_L - p_L)} \,,$$

so $$\xi\frac{(\pi_H - \pi_L)}{c_H} = \frac{d\log(q)}{d\log(\pi_L - p_L)} = \varepsilon_{q,\pi_L - p_L} \quad (*),$$

where ξ is the coefficient of relative risk aversion and ε_{q,π_L} is the elasticity of the probability of being in the low-risk to the probability of being in the high-risk state w.r.t. the low-risk-state premium. The right-hand side is the elasticity of the odds ratio of the low-risk to the high-rise state with respect to the difference between the low-risk premium and the probability of loss in the low-risk state.

To prove corollary 1.1, we note that in the absence of moral hazard, $c(p) \equiv 0$ and hence $\varepsilon_{q,\pi_L - p_L} \equiv 0$. Since $\xi > 0$, this means that $\pi_H - \pi_L = 0$, so the premiums in the low-risk and the high-risk states are equal.

To prove corollary 1.2, we note that although the right-hand side of this equation will change as risk aversion changes, because those changes only occur through changes in p, they are second order. This therefore shows that ξ and $\pi_H - \pi_L$ are inversely-related.

To prove corollary 1.3, we note that $\frac{d\ln(q)}{d\ln(\pi_L - p_L)} = \frac{\pi_L - p_L}{\pi_L}\frac{\pi_L}{q}\frac{dq}{d\pi_L} = \frac{\pi_L - p_L}{\pi_L}\varepsilon_{q,\pi_L}$. As p_L rises, the right hand side of (*) therefore falls, meaning that for fixed ξ, the premium spread $\pi_H - \pi_L$ must fall.

Proposition 2: If the insured would prefer the actuarially-fair contract $\{(\pi,\pi)\}$ *over remaining uninsured, then the optimal contract offered by the insurer has* $\hat{\pi}_L < \hat{\pi}_H$. *Furthermore, the optimal premium spread between high-risk and low-risk states in the second period is less than actuarially fair, that is* $\hat{\pi}_H - \hat{\pi}_L < p_H - p_L$.

Proof:

Writing

$$V_I^*(\pi_L,\pi_H) = \max_p pu(x-\pi_L) + (1-p)u(x-\pi_H) - c(p)\ ,$$

the optimal contract will be defined as the maximum over:

$$\max_{\pi_L,\pi_H,\lambda} V_I^*(\pi_L,\pi_H) + \lambda((1-p)p_H - (1-p)\pi_H + p.p_L - p.\pi_L)$$

Taking first order conditions yields:

$$\frac{dV_I^*}{d\pi_L} + \lambda(-p + \frac{dp}{d\pi_L}(\pi_H - \pi_L - (p_H - p_L))) = 0\ ,\ \text{or}\ \ \lambda = \frac{dV_I^*}{d\pi_L} / (\frac{dp}{d\pi_L}(\pi_H - \pi_L - (p_H - p_L)) - p)\ .$$

$$\frac{dV_I^*}{d\pi_H}+\lambda(-(1-p)+\frac{dp}{d\pi_H}(\pi_H-\pi_L-(p_H-p_L)))=0\text{ , or }\lambda=\frac{dV_I^*}{d\pi_H}/(\frac{dp}{d\pi_H}(\pi_H-\pi_L-(p_H-p_L))-(1-p))\ .$$

In each of the first two FOC's, the first term represents the expected utility cost of marginally altering each variable, while the second term, multiplied by the Lagrange multiplier, represents the effect of altering each variable on the budget constraint. In each of these FOC's, the shadow cost is itself divided into two pieces. The first term inside the brackets is a direct effect of changes in premiums on the budget constraint. The second term inside brackets is an indirect effect, caused by the behavioral change of the insured in response to the change in the premium structure. These indirect effects are large when the difference between the spread of the risk in the high and low-risk states and the spread of the premiums in those states – that is, the term $p_H-p_L-(\hat{\pi}_H-\hat{\pi}_L)$ – is large, or when the probability of ending in the low-risk state is very sensitive to changes in the premium structure.

For instance, if we assume that $c(p)=x$ for all p, so there is no moral hazard, then the terms representing the indirect effects all vanish, and we are left, as before, with an optimal contract that equalizes marginal utility across all states by having premium that are equal in all states. When there is moral hazard, it is the indirect, behavioral effects that drive the optimal premium away from this simple structure.

Setting the values of λ implied by the two first-order conditions equal to one another yields:

$$\frac{\frac{dV_I^*}{d\pi_L}}{p+\frac{dp}{d\pi_L}((p_H-p_L)-(\pi_H-\pi_L))}=\frac{\frac{dV_I^*}{d\pi_H}}{(1-p)+\frac{dp}{d\pi_H}((p_H-p_L)-(\pi_H-\pi_L))}$$

However, from the agent's optimization problem (and remembering the Envelope theorem), we must have that

$$\frac{dV_I^*}{d\pi_L}=-pu'(x-\pi_L)\text{ and }\frac{dV_I^*}{d\pi_H}=-(1-p)u'(x-\pi_H)\ ,$$

so the previous equation simplifies to:

$$\frac{u'(x-\pi_L)}{1+\frac{dp}{d\pi_L}\frac{(p_H-p_L)-(\pi_H-\pi_L)}{p}} = \frac{u'(x-\pi_H)}{1+\frac{dp}{d\pi_H}\frac{(p_H-p_L)-(\pi_H-\pi_L)}{1-p}}.$$

To prove the proposition, we note that since from the agent's optimization we have that

$$\hat{p} = (c')^{-1}(u(X-\pi_L)-u(X-\pi_H)),$$

it follows that $\frac{d\hat{p}}{d\pi_L} < 0$ and $\frac{d\hat{p}}{d\pi_H} > 0$. Intuitively, as the low-risk premium rises, holding the high-risk premium constant, this reduces the incentive individuals have to expend costly effort to increase the probability of entering the low-risk state. Raising the high-risk-state premium while holding the low-risk state premium constant has the opposite effect.

Assuming that $p_H - p_L - (\hat{\pi}_H - \hat{\pi}_L) > 0$, that is, that the premium spread between high and low-risk states is less than actuarially fair in the second period (we will prove this later), and given the signs of the derivative terms, we must have that:

$$u'(X-\hat{\pi}_L) < u'(X-\hat{\pi}_H), \text{ or } \hat{\pi}_L < \hat{\pi}_H.$$

Once moral hazard has been introduced, we therefore have shown that $\hat{\pi}_L < \hat{\pi}_H$.

Now, we show that the premium spread is less than the actuarially-fair spread in the second period by contradiction. If we assume that the premium spread is actuarially-fair, so $p_H - p_L - (\hat{\pi}_H - \hat{\pi}_L) = 0$, the two first-order conditions then imply that $\hat{\pi}_L = \hat{\pi}_H$ meaning that $p_H - p_L - (\hat{\pi}_H - \hat{\pi}_L) = p_H - p_L > 0$, which is a contradiction.

The participation constraint can be verified in two stages. First, the insured, by definition, prefers the insurance contract $\{(\hat{\pi}_L, \hat{\pi}_H)\}$ to the actuarially-fair contract with level premiums $\{(\tilde{\pi}, \tilde{\pi})\}$. Note that this contract clearly exists, and is unique because there will be only one level premium $\tilde{\pi}$ which satisfies the zero-profit constraint, and such a $\tilde{\pi}$ is feasible. Second, the insured prefers the contract $\{(\tilde{\pi}, \tilde{\pi})\}$ to remaining uninsured, by assumption.

Proposition 3: If prudence and higher-order terms in the utility function are sufficiently small that they can be disregarded, and relative risk aversion is locally constant, the following approximate relationship holds:

$$\xi \frac{(\pi_H - \pi_M)}{c_H} = \frac{d\log(1-p)}{d\log(\pi_M - p_M)} = \varepsilon_{1-p,\pi_M - p_M} ,$$

where ξ *is the coefficient of relative risk aversion, and* $\varepsilon_{q,\pi_M - p_M} = \frac{d\log(q)}{d\log(\pi_M - p_M)}$ *is the elasticity of the odds ratio w.r.t. the difference between the premium and the probability of loss in the first period.*

Corollary 3.1: In the absence of moral hazard, the optimal premiums in the high-risk state and the first period are equal.

Corollary 3.2: The more risk averse the customer, the lower the optimal premium dispersion between the high-risk state and the first period, ceteris paribus.

Corollary 3.3: The greater the moral hazard (in the sense of the difference between p_H *and* p_M*), the greater the premium dispersion between the first period and the high-risk state in the second period, ceteris paribus.*

Proof:

Due to the enforceability constraints, we note that $\hat{\pi}_L = p_L$. We can therefore treat the insurer's problem as optimizing over $\hat{\pi}_H$ and $\hat{\pi}_M$.

The policyholder's problem, given a premium structure is:

$$V(\pi_M, \pi_H) = \max_{\{c_M; c_H; p\}} u(c_M) + pu(x - p_L) + (1-p)u(c_H) - c(p) \text{ s.t. } x - c_M \geq \pi_M \text{, and } x - c_H \geq \pi_H$$

The insurer's problem is to choose a premium structure that maximizes consumer utility:

$$\max_{\{\pi_M; \pi_H\}} V(\pi_M, \pi_H) \text{ s.t. } (1-p)p_H + p_M = (1-p)\pi_H + \pi_M ,$$

recognizing that p is a function of the chosen premium structure. At an interior optimum, the optimal premium structure must satisfy

$$\frac{dV}{d\pi_M}(\hat{\pi}_M) = 0 \,,$$

recognizing that π_M is a function of π_H because of the insurer's budget constraint, and writing V using Lagrange multipliers as

$$V(\pi_M) = \max_{\{c_M;c_H;\lambda;\gamma_M;\gamma_H\}} u(c_M) + (1-p)u(c_H) - c(p) + \gamma_M(x - c_M - \pi_M) + \gamma_H(x - c_H - \pi_H)$$

$$\frac{dV}{d\pi_M} = -\gamma_M - \gamma_H \frac{d\pi_H}{d\pi_M} = 0 \text{, so } \frac{d\pi_H}{d\pi_M} = -\frac{\gamma_M}{\gamma_H} \,.$$

but $\gamma_M = u'(c_M)$ and $\gamma_H = (1-p)u'(c_H)$ so

$$\frac{d\pi_H}{d\pi_M} = -\frac{u'(c_M)}{(1-p)u'(c_H)} \,.$$

From the budget constraint, we have:

$$(1-p)p_H + p_M = (1-p)\pi_H + \pi_M \text{, so } \frac{d\pi_H}{d\pi_M} = \frac{-(1-p) - (p_M - \pi_M)\frac{dp}{d\pi_M}}{(1-p)^2} \,.$$

Combining these two gives:

$$\frac{u'(c_M)}{u'(c_H)} = 1 + \frac{p_M - \pi_M}{(1-p)} \frac{dp}{d\pi_M} = 1 + \frac{\pi_M - p_M}{(1-p)} \frac{d(1-p)}{d(\pi_M - p_M)} = 1 + \frac{d\log(1-p)}{d\log(\pi_M - p_M)}$$

Writing $u'(c_M) = u'(c_H) + u''(c_H)(c_M - c_H)$ and remembering that $c_M - c_H = \pi_H - \pi_M$ gives

$$\frac{u'(c_H) + u''(c_H)(c_M - c_H)}{u'(c_H)} = 1 + \xi \frac{(\pi_H - \pi_M)}{c_H} = 1 + \frac{d\log(1-p)}{d\log(\pi_M - p_M)} \,,$$

so $$\xi \frac{(\pi_H - \pi_M)}{c_H} = \frac{d\log(1-p)}{d\log(\pi_M - p_M)} \quad (*),$$

where ξ is the coefficient of relative risk aversion. The right-hand side is the elasticity of the probability of landing in the high risk state with respect to the difference between the first-period premium and the probability of loss in the first period, so the bond posted by the insured in the first period.

To prove corollary 3.1, we note that in the absence of moral hazard, $c(p) \equiv 0$ and hence $\varepsilon_{1-p,\pi_M - p_M} \equiv 0$. Since $\xi > 0$, this means that $\pi_H - \pi_M = 0$, so the premiums in the first period and the high-risk states are equal.

To prove corollary 1.2, we note that although the right-hand side of this equation will change as risk aversion changes, because those changes only occur through changes in p, they are second order. This therefore shows that ξ and $\pi_H - \pi_M$ are inversely-related.

To prove corollary 1.3, we note that $\frac{d\ln(1-p)}{d\ln(\pi_M - p_M)} = \frac{\pi_M - p_M}{\pi_M}\frac{\pi_M}{q}\frac{d(1-p)}{d\pi_M} = \frac{\pi_M - p_M}{\pi_M}\varepsilon_{1-p,\pi_M}$. As p_M rises, the right hand side of (*) therefore falls, meaning that for fixed ξ, the premium spread $\pi_H - \pi_M$ must fall.

Proposition 4: If $c'(\underline{p}) = 0$ *and* $c''(p) > 0$ $\forall$ $p \in [\underline{p};\overline{p}]$, *and* $c'(\overline{p}) = \infty$ *and the insured would prefer the actuarially-fair but level contract* $\{(\pi,\pi)\}$ *over remaining uninsured, then the optimal contract offered by the insurer has* $\hat{\pi}_L = p_L$, $\hat{\pi}_H < p_H$ *and* $\hat{\pi}_M > p_M$. *Furthermore, this contract is preferred to a series of successive one-period contracts with actuarially-fair premiums in each state.*

Proof:

As an immediate consequence of the unilateral enforceability constraints, $\hat{\pi}_L \le p_L$ and $\hat{\pi}_H \le p_H$, it is clear that actuarial fairness of the overall contract requires $\hat{\pi}_M \ge p_M$. The only reason that the insured would prefer such a contract to successive single-period actuarially-fair contracts is if the premium spread in the second period is less than the actuarially-fair premium spread. Clearly, this is incompatible with $\hat{\pi}_H = p_H$, since the enforceability constraint $\hat{\pi}_L \le p_L$ would imply a greater-than actuarially-fair premium spread in the second period. Therefore, $\hat{\pi}_H < p_H$. Because $p_H > p_M > p_L$, the actuarially-fair contract that also had $\hat{\pi}_L < p_L$ would be dominated by a contract that had $\hat{\pi}_L = p_L$ and that increased $\hat{\pi}_M$ to compensate. Therefore $\hat{\pi}_L = p_L$, $\hat{\pi}_H < p_H$ *and* $\hat{\pi}_M > p_M$. The participation constraint can be verified in the same manner as before.

From the first part of the proposition, the successive one-period contracts are feasible for the insurer. However, the insured gets higher utility from the optimal two-period policy described in the proposition. Therefore long-term contracts are preferred.

Appendix B

This restates proposition 1 in the case where prudence is too large to be assumed away.

Proposition 1: If higher-order terms than prudence in the utility function are sufficiently small that they can be disregarded, and relative risk aversion is locally constant, the following approximate relationship holds:

$$\xi \frac{(\pi_H - \pi_L)}{c_H}\left[1 + \frac{1}{2}\gamma \frac{(\pi_H - \pi_L)}{c_H}\right] = \varepsilon_{q,\pi_L - p_L},$$

where c_H *is the individual's consumption in the high-risk state,* ξ *is the coefficient of relative risk aversion,* γ *is the coefficient of relative prudence and* $\varepsilon_{q,\pi_L - p_L} \equiv \frac{d\log(q)}{d\log(\pi_L - p_L)}$ *is the elasticity of the ratio of the probability of being in the low-risk to the high-risk state in the second period w.r.t. the excess premium paid in the low-risk state.*