Title
National self-insurance and self-protection against adversity: bureaucratic management of security and moral hazard

Permalink
https://escholarship.org/uc/item/5jb7v8nv

Journal
Economics of Governance, 11(2)

ISSN
1435-8131

Authors
Ihori, Toshihiro
McGuire, Martin C.

Publication Date
2010-04-01

DOI
10.1007/s10101-010-0073-z

Peer reviewed
National self-insurance and self-protection against adversity: bureaucratic management of security and moral hazard

Toshihiro Ihori · Martin C. McGuire

Abstract This paper extends existing analyses of self-insurance and self-protection—distinctions first made by Ehrlich and Becker (J Polit Econ 80:623–648, 1972)—that countries may implement at a national level in pursuit of their security. We show that, when no market insurance is available, self-insurance alone raises important new issues as to the definition of “fair pricing” and as to the relations between pricing, optimization, risk aversion, and inferiority that are significantly different from standard, conventional market analysis. We also discover a hitherto unrecognized tendency for misallocation between self-protection and self-insurance when both are available and considered together. Because of external effects running from self-protection to self-insurance, governments ruled by myopic bureaucracies and trying to find the right balance face incentives that encourage extreme, self-inflicted moral hazard, to the detriment of self-protection.

Keywords Self-insurance · Self-protection · Actuarially fair condition · Inferior goods · Public goods

An earlier version of this paper was presented to a Conference on “The Causes and Consequences of Conflict,” Wissenschaftszentrum Berlin (WZB), Germany, March 28–29, 2008, with proceedings in a special issue of Economics of Governance. The authors thank Robin Boadway, Magnus Hoffman, Kai Konrad, other conference participants, and this journal’s referee for insightful comments on earlier versions.

T. Ihori (✉) 
Department of Economics, University of Tokyo, 113-0033 Hongo, Tokyo, Japan 
e-mail: ihori@e.u-tokyo.ac.jp

M. C. McGuire 
Department of Economics, University of California-Irvine, Irvine, CA 92697, USA 
e-mail: mcmcguir@uci.edu
1 Introduction

Among the first to examine combinations of instruments open to individuals to manage risks to their well-being were Ehrlich and Becker ("EB," 1972). EB identified several types of preparation available to expected utility maximizing individuals faced with what we will call "costs of emergency." These costs consist of any mix of (a) probability of loss and (b) magnitude of loss (hereafter together referred to as "risk profile"). Among such preparations were (a) "self insurance" to compensate for or reduce the magnitude of loss (b) "self-protection" to reduce the probabilities of loss—neither of which involved market choice or market insurance—and (c) insurance purchased from others in a market.

It would now be generally agreed that these ideas apply to entire societies attempting to cope with diverse conflicts. Governments need not passively accept risks that production and/or consumption decline in unwelcome situations. The EB distinctions apply with respect to nations also (Ihori 1994). For them diverse instruments exist to manage national adversity including self-insurance and self-protection. Most governments, at present, treat these two as essentially separate.¹ But, as we show, separation of security provision into self-protection and self-insurance is a bad idea. There needs to be much more coordination between these two seemingly disparate functions of government (see McGuire 2000, 2006; McGuire and Becker 2006).

Analysis of multiple instruments of collective risk management must extend the standard market insurance paradigm in several respects. First, sovereign agents will face increasing costs for self-insurance coverage in contrast to the competitive linear prices relevant to individual market insurance. This feature of risk management at the level of nations deserves more attention than usually accorded to it; diminishing returns to self insurance require new approaches to the meaning and measurement of actuarial fairness (as spelled out in the Appendix to this paper), and to identification of risk neutral resource allocations for instance. Second, when instruments both for self-insurance and for self-protection are available to a single decision maker, reducing risks through protection changes the price of self-insurance and thus the two allocations will interact in an unfamiliar manner. This essay will show that prominent among these is a type of self inflicted moral hazard. Moreover, we must integrate into such non-market situations known results concerning inferiority of market insurance (Mossin 1968; Hoy and Robson 1981; Eeckoudt and Gollier 2000; Schlesinger 2000) and of protection (Ihori and McGuire 2006, 2007, 2008a,b).

When an agent provides security by spending on commercial market insurance, then a standard result states that with actuarially fair linear pricing, complete coverage (net of premiums) is purchased from resources available in the good contingency (EB, Mossin 1968; Hirshleifer and Riley 1975). We compare this benchmark result

¹ We agree with EB that distinctions between risk and magnitude of loss are in one sense artificial since any countermeasure most likely will show alloyed improvements along both dimensions. Nevertheless governments typically ignore the interactions among these components. For example, diplomatic or defense policy striving to reduce chances of energy supply disruption may recognize strategic petroleum reserves as a backstop for failure. But it would be extremely rare for the benefit from reducing the need/cost of stockpiling to be counted in a defense planner’s estimate of how to value policy efforts on persuasion, deterrence, bribery, or threat.
with self-insurance by an entire country, and use it also to examine the effect of such self-insurance on the choice of self-protection. This necessitates close attention to fair pricing in a context of diminishing returns; and here we discover that here the definition of actuarial fairness itself is ambiguous.

First we show that if the country provides of self insurance as a national public good and if its spending on this self-insurance is subject to diminishing returns due to decreasing productivity of insurance premiums, then (contrary to the standard market result just mentioned) complete coverage (net of premiums) need not be purchased from resources available in the good contingency, irrespective of the ambiguity of fairness in pricing.

Next, we consider the incentives and conflicts facing a government that can both insure and protect at the same time. We identify three effects of security expenditure: (a) “insurance benefit” provided by the Government’s “insurance branch,” (b) “reduction of dispersion or variance of outcomes” by the Government’s “protection branch,” (c) “insurance cost savings benefits” provided by the protection branch. These distinctions are central to our paper. We show for comparison that if self-insurance can be provided at a market-like fair price, a country that fails to coordinate protection and insurance will never self-protect at all. Failure to recognize the insurance benefits generated by linear priced self-protection leads to its shutdown. Thus, we show that the difficulties of extreme moral hazard can be self inflicted. The incentive to take excessive risk is not due to a desire to pass costs on to a second party, but instead to information deficits and mal-coordination by the providers of protection. In fact, any agent, including entire governments, that is capable of both self-insurance and self-protection but who being a price-taker is also unaware of the effects of protection on the price of insurance will succumb to incentives similar to moral hazard.

We then apply the same idea when insurance is provided by a large agent, say the size of an entire nation, where such self insurance is usually characterized by the decreasing returns. Here, we argue, nations will have access to actuarially fair returns to their insurance outlays. And we show that diminishing productivity of insurance together with actuarial fairness in returns in no way resolves the self-inflicted moral hazard problem.

Our arguments conclude that unless properly coordinated, self-insurance will drive out self-protection completely. This implies that there is far too little consideration given to measures that will improve the odds against war, catastrophe etc (not particularly military measures, all measures). Our arguments lead to the implication that too much attention tends to be given to insurance-type effort and too little to risk improving effort.

This paper consists of four sections. First, we formulate a basic analytical framework. Then Sect. 2 reviews characteristics of optimization under market insurance, self-insurance, and alternatively self-protection for a single country. Section 3 considers interaction between self-insurance and self-protection. Finally Sect. 4 concludes.

2 Analytical framework for a single agent

Let us begin with a narrative to illustrate the concept that our models will try to capture or summarize. Imagine there is an island nation that is subject to flooding.
Whenever a flood happens there is a big loss. No matter how big the flood, the loss is the same, \( L \). To protect itself against this loss the country can reduce the frequency of flooding by building flood-barriers, dikes, channels etc. If it builds no dikes a flood happens every other year. If it builds dikes that are 6 feet tall, the country will be flooded every 6 years. If this country builds dikes 15 feet tall it will be flooded every 11 years. The frequency of flooding depends on the height of its flood barriers. The relationship between cost of dikes and frequency of flood \((1 - p)\) will be known with certainty.\(^2\)

To prepare for this loss, the country can also stockpile food and other necessities. Ignore time discounting and assume to start a linear relationship such that that if a flood happens every other year, each year without a flood the country can set aside \( x \) pounds of goods for the next year when there is a flood and have available in that year \( x \) pound of goods. If a flood happens every 6 years then to have \( x \) pounds available during the flood, the country only needs to give up \( x/5 \) lbs during each of the 5 dry years. If a flood happens every 11 years the country can provide \( x \) lbs during the flood by giving up only \( x/10 \) lb in each of the dry years. In other words, as an initial assumption suppose the country can self-insure at an actuarially fair price, \((1 - p)/p\), irrespective of the scale of provision, “\( x \).”

2.1 Risk profiles and emergency cost

Congruent with the foregoing story we consider a single agent and two contingent states, a good state “1” and a bad state, “0”. Ignoring all insurance and compensation possibilities (that is taking \( L \) as a fixed parameter) expected utility for this agent is given as:

\[
W = pU^1(Y) + (1 - p)U^0(Y - L) \\
C^1 = Y; \quad C^0 = Y - L \\
or \quad W = W(Y, p)
\] (1) (2)

where \( W \) is expected utility, \( C \) is consumption, \( L \) is loss in the bad state, and \( p \) is the chance of a good state. Our analysis will focus on the two canonical types of Ehrlich-Becker (EB) defense; (i) EB’s “self-protection;” which raises \( p \) and reduces \((1 - p)\), (ii) EB’s “self-insurance” which reduces \( L \). Aside from our flood narrative, the variable “\( p \)” might be risk of trade interruption, disease outbreak, environmental calamity, or war. Utility function \( U() \) is assumed to be the same whether luck is good or bad. \( U^1 \) denotes realized utility if the good event happens, and \( U^0 \) if the bad event happens, and \( U_Y \equiv \partial U/\partial Y > 0, U_{YY} \equiv \partial^2 U/\partial Y^2 < 0 \).

To establish the incentives for a single agent Eq. (3) shows the individual budget constraint: where \( Y \) is a fixed income and \( m_k \) denotes allocations \((k = 1, 2)\) to risk

\(^2\) Let “\( d \)” be the number of dry years and “\( w \)” the number of wet years. Then \( p = d/(w + d) \) gives the frequency of dry years or of success, and \((1 - p)\) give the frequency of wet years or of failure.

\( \odot \) Springer
reduction, \( p(m_1) \), and or loss reduction \(-L(m_2)\)—here

\[
Y = C + m_k
\]  

(3)

considered a variable of choice. Therefore, if our concern is with insurance only—with \(-L\) a variable but \(p\) taken to be a parameter—Eq. (1) can be written

\[
\tilde{W} = \tilde{W}(Y, m_k).
\]  

(4)

Equation (4) then shows it can be natural and helpful to consider \(m\) rather than \(L\) to be the national security public good.

2.2 Market insurance: standard result

We will emphasize presently how the structure and context of self-insurance for a large entity such as an entire nation is inherently quite different from market insurance. So it is for later comparative purposes useful to set out in brief summary the standard market insurance model, where again \(m_2\) represents quantity of coverage in bad times.

A basic feature of the standard market insurance model (which distinguishes it from self-insurance) is that the cost of insurance coverage in bad times is linear. For a linear insurance recovery function instead of \(-L(m_2)\) we write (with numeraire income being consumption in good times):

\[
C^1 = Y - m_1 - \pi m_2
\]  

(5)

and

\[
C^0 = Y - m_1 - (L - m_2)
\]  

(6)

where \(\pi m_2\) gives the expenditure on market insurance in good times (measured in units of \(C^1\)), and \(m_2\) represents the amount of insurance coverage purchased at price \(\pi\). Entered as a parameter, \(m_1\) gives the allocation to risk improving self-protection. Then welfare becomes\(^3\)

\[
W = p(m_1)U^1 [Y - m_1 - \pi m_2] + (1 - p(m_1))U^0 [Y - m_1 - (L - m_2)].
\]  

(7)

Then for optimal insurance, maximizing Eq. (7) (or equivalently Eq. (8) from the footnote) with respect to \(m_2\) yields necessary condition Eq. (9).

\[
- p\pi U^1_Y + (1 - p)U^0_Y = 0
\]  

(9)

\(^3\) If instead we took \(m_2/\pi\) to mean the amount of coverage and \(m_2\) contingency-1 expenditures, we would write

\[
W = p(m_1)U^1 [Y - m_1 - m_2] + (1 - p(m_1))U^0 [Y - m_1 - (L - m_2/\pi)].
\]  

(8). Equations (7) and (8) are equivalent in the standard linear case, but we will favor Eq. (7) as more conventional. But in the non-linear case Eqs. (7) and (8) differ importantly, since each implies a different measure of actuarial fairness and risk neutral allocation—issues treated briefly in the Appendix and deserving of further analysis.
If the price of insurance happens to be actuarially fair as in Eq. (10) then the FOC would entail $U_1^1 = U_0^0$ whence

$$\text{Actuarial fair price: } \pi = \frac{(1 - p)}{p}$$

(10)

$U^1 = U^0$ and, therefore, $C^1 = C^0$ or $Y - m_1 - \pi m_2 = Y - m_1 - (\bar{L} - m_2)$. And from this it follows that at the optimum, insurance coverage purchased is

$$m_2 = p\bar{L}$$

(11)

so that the total cost of such fairly priced insurance at the optimum becomes

$$\pi m_2 = (1 - p)\bar{L}$$

(12)

This standard result states that with fair linear pricing, complete coverage (net of premiums) is purchased. We will use this market insurance summary as the benchmark for later to comparison with self-insurance by an entire country, and also in examination of the effect of insurance on the choice of protection. $m_2$ is independent of income; the income effect is zero.

2.3 Self insurance

Self-insurance provided to itself by a large entity such as a nation differs in two important respects from standard market insurance. To show this, we alter notation slightly. Rather than $-L(m)$ where $L(0)$ was a threatened loss if nothing is spent on insurance, we write

$$-L(m) = -[\bar{L} - L(m)]$$

(13)

Now the entire, total, insurance benefit is shown by $L$, and $L(0)$ is given by $\bar{L}$.

2.3.1 Diminishing returns

First of all, self-insurance differs from standard market insurance in that self-insurance function $L$ should show diminishing returns or increasing costs. $L' > 0, L'' < 0$. EB make this assumption also, and refer glancingly to the role of human capital in providing for self-insurance as a source of diminishing returns. We believe (i) that scale considerations appropriate for an entire country along an extensive margin as well as (ii) other cooperating factors of production, as in EB, argue that “self-insurance” has such declining marginal productivity. National self-insurance may often involve actions like stockpiling or standby production maintenance and these surely will show diminishing returns. If $m_2$ is very productive, $-L$ may even conceivably increase.

\footnote{In our story of flood protection and insurance, as greater quantities of consumables are set aside during dry years, their costs of preservation and delivery during good years might increase more than proportionately. For example, as an extreme case, if $p = 1/2$ setting aside $m_2 = 1$ provides 1 unit in bad times, but saving $m_2 = 8$ yields only 4 units in bad times, etc. Here $L = (m_2)^{2/3}$.}
be negative for high values of \( m_2 \) (recognized also by EB as “negative insurance” or as termed here, “gambling.”), so that over some region \(-L(m_2) = -[\overline{L} - L(m_2)] > 0\). However, we ignore this case as it implies a reversal between good and bad contingencies. Declining “productivity” of “\( m_2 \)” thus is the first source of a distinction between sovereign self-insurance vs. lesser scale decentralized market insurance.\(^5\)

Such diminishing returns also will introduce new issues in the formulation of inter-contingency pricing of self-insurance that are absent from market insurance—a fact not recognized in the literature as far as we can tell—new issues that imply further hitherto unrecognized interactions between the form of pricing and the structure of the optimum. Quite arguably, the self-insurance function should be written more generally as \( L(m_2, p, \pi) \) to allow for still more complicated interactions between insurance, risk, and price. This would cause the definitions and formulation of “actuarial fairness” become ambiguous. But we relegate these complications to the appendix and concentrate here on the most salient formulation of pricing when insurance is non-linear with inter-contingency price \( \pi \) as shown in Eq. (14):

\[
W = pU(Y - \pi m_2) + (1 - p)U^0(Y - (\overline{L} - L(m_2))): m_1 \text{ not shown}\tag{14}
\]

where \( \pi \) shows the actuarial price per unit in good times necessary to yield \( m_2 \) units of resources in adversity.

2.3.2 Salience of fair pricing

Complications like writing \( L = L(m_2, p, \pi) \) aside, a second major difference between self-insurance as provided by an entire country and ordinary market insurance is that when a whole nation provides insurance to itself, fair pricing would seem to be the standard case and not an outlier just referenced for comparison. Of course nation’s can make mistakes, have imperfect information etc. But countries in this position are “bargaining with themselves” as to how much insurance and at what price to provide it. They should not in principle have to worry about adverse selection or moral hazard. So they should not give themselves deductibles, “load” prices nor impose arbitrary insurance limits to control fraud.\(^6\) Moreover, it is plausible to assume that the nation as a price maker, not as a price taker, incorporates this actuarially fair condition at its optimization.

2.3.3 The insurance optimization problem

Now similar to EB’s derivation, expected utility Eq. (14) is maximized with respect to \( m_2 \). This gives Eq. (15) as the first order condition. Equation (15) shows the marginal

\(^5\) EB state in passing that self-insurance is independent of risk, but this is surely a mistake. Stockpiling for a seven year recurring famine is surely more expensive than for one that comes every 25 years.

\(^6\) Of course countries have corruption, rent seeking, and numerous misalignments of incentives to concern them.
cost of providing $l$,

\[
- p\pi U_Y^1 + (1 - p)U_Y^0 L'(m_2) = 0
\]

i.e. $[p\pi U_Y^1]$, equal to the marginal benefit of providing $l$, i.e. $(1 - p)U_Y^0 L'$, evaluated at the solution value of $m_2$ If this necessary condition is rewritten as in Eq. (16) then its actuarial meaning becomes clear. The RHS there gives the probability weighted marginal insurance-benefit receipt under adversity while the LHS gives marginal cost of one unit of insurance for the last dollar of premium paid in good times.

\[
U_Y^1/U_Y^0 = [(1 - p)/p\pi]L'(m_2)
\]

2.3.4 Definitions of “Actuarial Fairness”

If self-insurance is actuarially fair (henceforth simply “fair”) as we believe should be the paradigm for an entire country then the concept must be defined. An obvious parallel to the fairness under linear market insurance is resource allocation fairness as in Eq. (10) which is the same as Eq. (17).

Resource allocation fairness: \( \pi = (1 - p)/p \).

An alternative definition of actuarial fairness would incorporate the marginal productivity of resources as applied to the bad contingency. This we label “marginal productivity fairness,” as defined by Eq. (18):

Marginal productivity fairness: \( \pi = L'(1 - p)/p \).

2.3.4.1 Optimal solution allocations

If fairness under marginal productivity—Eq. (18)—obtains then at the optimum $U_Y^1/U_Y^0 = 1$ and thus $U_Y^1 = U_Y^0$, $U^1 = U^0$, $Y^1 = Y^0$, so that the optimum requires complete coverage and the analysis of inferior goods proceeds just as in the linear case. Although this definition of actuarial fairness leads to a nice symmetry it assumes that the self-insuring agent somehow knows its optimal purchase of insurance “in advance,” so as to have knowledge of $l'$ before actually making its allocations. But this seems implausible. Moreover, under Eq. (18) one could not tell whether insurance was fairly priced until the optimum was actually chosen. This supports the first, more conventional, and we believe preferable, definition of actuarial fairness. Under fairness by that first definition Eq. (17) then the optimum simplifies to

\[
U_Y^1/U_Y^0 = L'(m_2)
\]

Equation (19) is simply a familiar equality of MRS and Marginal Rate of Transformation, which obtains irrespective of risks $(1 - p)$ so long as the price of insurance is fair. But note, in contrast to the linear market insurance case no equivalence is implied between $U_Y^1$ and $U_Y^0$. 

\[\text{Springer}\]
These distinctions between “resource allocation fairness” and “marginal productivity fairness” in the pricing of insurance are analyzed in depth in the companion paper to this (Ihor and McGuire 2008b). For the remainder of this paper we focus on “resource allocation fairness” as characterized by Eq. (10) or (17).

2.4 Self protection

Now to return to our flood protection anecdote we suppose a country can reduce the frequency of flooding by building flood-barriers, dikes, channels etc. (See Cornes 1993). The frequency of flooding depends on the height of the flood barriers, and the relationship between cost of dikes and frequency of flood is known with certainty i.e. \((1 - p)\) in our model. Thus the second risk management instrument to consider is self-protection with \(m_1\) spent to reduce the chance of a bad event, \(1 - p\), i.e. to decrease what we call “baseline risk” of \([1 - p(0)]\). Ihori and McGuire (2007) demonstrated (with insurance fixed parametrically as similarly developed in McGuire et al. (1991)) that for self-protection the issue of normality-inferiority is substantially more involved than it has proven to be for self-insurance as analyzed here. We now desire to extend the Ihori-McGuire analysis of self-protection developed for the special case of fixed uninsured loss to the more general case of (1) variable loss and (2) self-insurance where (3) insurance benefit is non-linear, and (4) actuarial fairness interacts with diminishing returns. Whatever the risk-reduction/self-protection function, \(p(m_1), p' > 0\), and \(p'' < 0\) are assumed throughout. 

To begin, we repeat Eq. (14) now including both variables \(m_1\) and \(m_2\). Inserting the condition for resource actuarial fairness, \(\pi = \frac{(1 - p)}{p}\) directly gives:

\[
W = p(m_1)U^1 \left[ Y - m_1 - \frac{1 - p(m_1)}{p(m_1)} m_2 \right] + (1 - p(m_1))U^0 \left[ Y - m_1 - (L - L(m_2)) \right]
\] (14)

The FOC for determining optimal expenditure on self-protection becomes:

\[
\left[ p'(U^1 - U^0) \right] - \left[ pU_Y^1 + (1 - p)U_Y^0 \right] + \left[ (p'/p)m_2U_Y^1 \right] = 0.
\] (21)

We can characterize this optimality condition in the provision of self protection saying that there are “direct” marginal benefits in the form of the gain in utility \(p'(U^1 - U^0)\), “direct” marginal costs \(pU_Y^1 + (1 - p)U_Y^0\) and “indirect” benefits, \(p'm_2U_Y^1/p\),

---

7 Our “baseline risk” corresponds to what is sometimes referred to as “background risk” in economics of insurance analyses. Background risk distinguishes “independent” background risk where \(p(0)\) is not influenced by the value of \(L(0)\) as in our model here, versus “non-independent” background risk where \(p(0)\) and \(L(0)\) are interdependent, and asks how the choice of protection or insurance varies with the independence property (See Schlesinger 2000).

8 Where useful we can re-write Eqs. (2) as (20) where \(L, m_2, \) and \(l\) are now taken to be parameters. Differences in functions \(\hat{W}_Y(\cdot)\) and \(\hat{W}_I(\cdot)\) are implicit in each model, so the notational distinction will be omitted henceforth \(\hat{W} = \hat{W}(C, m_1)\) (20).
comprised of, an unambiguous gain from the decrease in insurance premiums paid for the same $m_2$ coverage received stemming from the lower price implied by lower risk $(1 - p)$. Note that since the sole variable of choice here is $m_1$ any possible implications of the change in $p$ ($m_1$) on subsequent choices of $m_2$ and, therefore, of insurance purchased (by the agent choosing insurance) are irrelevant.

\[
SOC \quad E = p''(U^1 - U^0) - 2p'(U^1_Y - U^0_Y) + \left[pU^1_{YY} + (1 - p)U^0_{YY}\right]
- U^1_{YY} \frac{p'm_2}{p} \left(-1 + \frac{m_2p'}{p^2}\right) + U^1_Y \frac{m_2}{p^2}(pp'' - p'p') < 0
\]  

### 3 Interaction between self-protection and self-insurance

How will and how should a country’s insurance and protection programs be affected by or influence one another? Do they complement or substitute for each other, compete or mutually reinforce? How does the availability and utilization of information among programs for protection versus insurance influence these relationships? These questions seem to have received little attention.

Before addressing the larger question, however, we analyze a special case to illustrate the crucial importance of information availability and accuracy. This is the case where the information and decision processes of the self-insuring/self-protecting agent are independent and isolated from each other. We demonstrate that under such restrictions, especially when insurance is actuarially fair, moral hazard behavior will drive provision of self-protection to zero. In particular we show this to be the case when returns to self-insurance are constant and linear, as would be the case with perfect and competitive market insurance.

#### 3.1 A special example: linear fair market insurance (LFI)

As a benchmark case then, consider linear constant unit cost insurance as in a market where a small agent takes the price of insurance as fixed. As before, for expected welfare write:

\[
W = p(m_1)U^1[Y - m_1 - \pi m_2] + (1 - p(m_1))U^0[Y - m_1 - (\bar{L} - m_2)]
\]  

where $m_2$ and $m_1$ and $\pi$ are as previously defined.

#### 3.1.1 Complete information utilized by the protection branch

Here we assume that the agent who provides insurance does not take the price of insurance $\pi$ as fixed, but rather incorporates the actuarial fair price condition Eq. (10) into his optimization. The standard result for a single coordinated and fully informed agent takes welfare the same as Eq. (7). Fair pricing, optimal self-insurance, and maximization with respect to $m_2$ as in Eq. (11) gives the optimum $m_2$ coverage purchased.
as \( m_2 = p \bar{L} \). Thus complete coverage is purchased net of premium cost. Next, assume self-protection \( p (m_1) \) can also be improved, so fair insurance price declines since risk declines with the increase in \( p \). How much will be spent on protection?

The answer in this standard case depends, on whether a self-protection provider is aware that protection improves the price of insurance. Assume full information—an alternative to the compartmentalized and unshared assumption of the next subsection—for an entity such as a nation that “bargains with itself” over the price of insurance. The protection branch knows that an increases in \( m_1 \) will lower the price of insurance, \( \pi \), by raising \( p \) and also knows this should affect the optimizing behavior of the insurance branch. In a sense, the protection branch is above the insurance branch in a hierarchy. We could suppose that the protection branch moves first, anticipating that its provision of lower risk saves insurance costs but having no knowledge whether the lower premium price it causes will stimulate or curtail the quantity of insurance. Then, the insurance branch moves second.\(^9\) We begin with the necessary condition, differentiating Eq. (7) with respect to \( m_1 \).

\[
p'(U^1 - U^0) - \left[ pU_Y^1 + (1 - p)U_Y^0 \right] + \left[ U_Y^1 m_2 (p'/p) \right] - \left[ \pi pU_Y^0 - (1 - p)U_Y^0 \right] \frac{dm_2}{dm_1} = 0 \tag{23}
\]

Then using the fact that pricing is fair after including \( U_Y^1 = U_Y^0 = U_Y \) Eq. (23) simplifies to

\[
p'[U^1 - U^0] - [pU_Y + (1 - p)U_Y] + m_2 (p'/p)U_Y - \left[ \pi pU_Y - (1 - p)U_Y \right] \frac{dm_2}{dm_1} = 0 \tag{24}
\]

When price is fair the final bracketed term vanishes at the optimum for insurance; therefore, the “protection branch” can ignore this effect (i.e. ignore \( dm_2/dm_1 \)) without undermining the overall optimum, since the effect vanishes when the insurance branch is doing its job properly. But Eq. (24) also includes \( m_2 (p'/p)U_Y^1 \) to indicate that because of the cost savings that \( m_1 \) generates for insurance, the optimal value of \( m_1 \) depends on the choice of \( m_2 \) even at that stationary optimum for \( m_2 \). So the decision of how much to protect, to be optimal, must include a part of the effect of protection on the price of insurance, namely the “insurance cost savings” effect. The “insurance branch,” controlling only \( m_2 \) will not recognize this effect attributable to \( m_1 \); the protection branch must recognize it. In this extended sense, the protection branch is the “leader,” and the insurance branch is the “follower.” If the protection branch fails to recognize it then as shown in the next sub-section the FOC-implementing choices of insurance and protection agents will become incompatible, and \( m_1 \) will

\(^9\) For individual protection and insurance, a hierarchy separation between protection and insurance decisions sounds implausible. The smoker anticipates lower life insurance rates when he quits smoking. He may think of buying more insurance at the lower rates as well. But to impute this sort of foresight to a government however is not at all obvious. Thus our “complete information” case might be regarded as ideal, even utopian.
be driven to zero. But if this effect of $m_1$ is recognized by the protection branch then from $m_2 = \frac{p}{L}$ and $U^1 = U^0$, at the overall optimum the optimal values of $m_1$ and $m_2$ imply

$$p' = \frac{1}{L}$$

(25)

So one of the benefits of spending to raise $p$ when risk is already completely covered by insurance (and will continue to be so covered after $p$ and therefore $\pi$ improves) derives from the decrease in the cost of the optimal amount of insurance coverage when the implied actuarially fair premium declines.\(^{10}\) In this case both $m_1$ and $m_2$ are independent of income; the income effect is zero.

### 3.1.2 Effect of imperfect information: no information sharing between insurance and protection providers

Now consider the protection branch of the country to be a price taker as the benchmark case as in a small open world. We investigate the outcome of non-cooperative Nash equilibrium in such a world by exploring the incentives of the protection agent if it ignores the cost savings it generates for providers of insurance. Here, although it may recognize that $p$ determines $\pi$, assume, nevertheless, the protection branch takes $\pi$ as a fixed parameter for its optimization problem and focuses solely on its role in improving the weights on good and bad outcomes in the expected utility balance. From the insurance provider’s choice of $m_2$ at his optimum, maximization of Eq. (7) with respect to $m_2$ yields as

$$\frac{\partial W}{\partial m_2} \equiv T(m_2) = -p\pi U_Y^1 + (1 - p)U_Y^0.$$ 

(26)

Again crucially, assume that although $m_1$ determines $p(m_1)$, and $p$ determines $\pi$, and $\pi$ determines the cost of insurance coverage $m_2$, nevertheless, the agent choosing $m_1$ is ignorant of these relationships and overlooks the benefit that $m_1$ creates by lowering the cost of existing insurance. That is, assuming that the agent who provides $m_1$ regards the price of insurance $\pi$ as fixed (despite the fact that because of that agent’s decisions this price continually adjusts to satisfy the assumption of actuarial fairness). Then for this case—which might describe behavior among government agencies that is myopic and uncoordinated even though they share the same overall objective function—the protection provider’s welfare maximization is given as:

$$\frac{dW}{dm_1} \equiv T(m_1) = p'(U^1 - U^2) - (pU_Y^1 + (1 - p)U_Y^0).$$

(27)

\(^{10}\) We could have reached the same conclusion more directly by observing that since optimal insurance when fairly priced equalizes income as between contingencies, the optimal value of $m_1$ must be that which maximizes this insured income, or minimizes \{-m_1 - (1 - p(m_1))L\}. This must be the value of $m_1$ such that $p' = 1/L$. 
Now, go back to the provider of insurance and suppose \( T(m_2) = 0 \). Then, from Eq. (26) plus an actuarially fair condition, \([(1 - p)/p] = \pi \) we know \( U_Y^0 = U_Y^1 \); hence \( U^0 = U^1 \). Substituting this into Eq. (27), gives \( T(m_1) = -U_Y^1 < 0 \). Thus, when Eq. (26) and the actuarially fair condition Eq. (10) both obtain, it follows \( T(m_1) = -U_Y^1 < 0 \). Next considering the provider of self-protection again let \( T(m_1) = 0 \). Then from Eq. (26) we know \( p'(U^1 - U^0) = [pU_Y^1 + (1 - p)U_Y^0] > 0 \) and hence \( U^1 > U^0 \). And this implies \( T(m_2) = -pU_Y^1 + pU_Y^0 > 0 \).

To interpret this define curve \( T(m_1) = 0 \) as the locus of \((m_1, m_2)\) which satisfies \( T(m_1) = p'(U^1 - U^0) - [pU_Y^1 + (1 - p)U_Y^0] = 0 \); and define curve \( T(m_2) = 0 \) as the locus of \((m_1, m_2)\) which satisfies \( T(m_2) = -pU_Y^1 + pU_Y^0 = 0 \). Then to summarize:

\[
\begin{align*}
\text{If } T(m_2) &= 0, \quad \text{then } T(m_1) < 0 \quad \text{(28)} \\
\text{If } T(m_1) &= 0, \quad \text{then } T(m_2) > 0 \quad \text{(29)}
\end{align*}
\]

We can never find values along Eq. (28) = 0 for which Eq. (29) = 0, since whenever Eq. (24) = 0, then Eq. (28) < 0. Therefore, at every point along the optimal insurance curve Eq. (28) = 0, the agent providing protection wants to reduce \( m_1 \) and thus increase risk. If he could increase risk so much that \( p < 0 \) and \((1 - p) > 1\), and be compensated for this increase by negative \( m_1\), the agent will want to do this. This is moral hazard in the extreme.

On the other hand, everywhere along the curve \( T(m_1) = 0 \), we see that \( dW/dm_2 > 0 \), meaning that an insurance agent would want to insure more at each value of \( m_1\). Thus it seems—in stark contrast to the incentives for a unitary government agent who coordinates provision of \((m_1, m_2)\)—that a decentralized security agencies would not provide both types of security at the same time. Rather they would normally provide \( m_2 > 0 \) only. The two reaction curves \( T(m_1) \) and \( T(m_2) \) in space \( m_1 - m_2 \) with a Nash solution at \( m_1 = 0, m_2 > 0 \) are shown in Fig. 1, where the closed contours are loci of constant “security.”

![Fig. 1 Uncoordinated nash behavior of insurance and protection agencies](image-url)
3.1.3 Need for inter-program information sharing

Bureaucratic compartmentalization like this seems to us entirely plausible, and not diminished by the assumption that fair pricing is applicable and available. The demands for inter bureau information exchange to achieve a first best optimum are significant, since country wide benefits from \( m_1 \) must be calculated to include cost savings in insurance “premiums” that increases in \( m_1 \) entail. That is the provider of protection must include insurance cost-savings benefits in his calculations to select the correct value of \( m_1 \). This exercise emphasizes the daunting information requirements necessary for the standard result.

3.2 General case: self insurance and self protection

According to the story in our introductory narrative the island nation subject to flooding, can reduce its frequency by building flood-barriers, dikes, channels etc. and can reduce the magnitude of loss by stockpiles, say of food and other necessities, self-insuring at an actuarially fair price, \( \pi = (1 - p) / p \). Here we dwell more generally on how these two instruments interact with each other. In particular we show how the effectiveness of insurance plus protection varies crucially with the quality of information shared within the government and of cooperation among insurance and protection measures, programs, or bureaus.

Since agents individually or in a group may spend on both insurance and protection, we are interested in how any one agent’s incentives interact with respect to the two instruments when both are available. Specifically, taking fair insurance to be the norm, we focus on how anticipation that expenditure on \( p \) will affect \( \pi \) influences the optimal choice of \( m_1 \) and therefore of \( p(m_1) \). First, suppose the government knows all these relationships. It knows the effect of dike height and cost on frequency, and knows it can self-insure at fair prices.

To begin we repeat Eq. (14) now including both variables \( m_1 \) and \( m_2 \), inserting the condition for actuarial fairness, \( \pi = [(1 - p) / p] \) directly, and writing \( W \) in the form where \( m_2 \) indicates quantity of insurance coverage purchased while \( \pi m_2 \) represents total expenditure on insurance coverage:

\[
W = p(m_1)U^1 \left[ Y - m_1 - \frac{m_2(1 - p(m_1))}{p(m_1)} \right] \\
+ (1 - p(m_1))U^0 \left[ Y - m_1 - \{L - L(m_2)\} \right]
\]

(14)

As derived in above,—assuming that the insurance branch accepts \( m_1 \) as a parameter and thus disregards any anticipated change in \( p(m_1) \) and therefore in \( \pi \)—the FOC with respect to \( m_2 \) gives

\[
S = -U^1_Y \left[ Y - m_1 - \frac{m_2(1 - p(m_1))}{p(m_1)} \right] \\
+ \mathbf{L}^0 U^0_Y \left[ Y - m_1 - \{L - L(m_2)\} \right] = 0
\]

(30)
The FOC for determining optimal expenditure on self-protection \( m_1 \) becomes:

\[
\begin{align*}
\left[ p' \left( U^1 - U^0 \right) \right] & - \left[ pU_Y^1 + (1-p)U_Y^0 \right] + \left[ U_Y^1 m_2 p'/p \right] \\
- \left[ p\pi U_Y^1 - (1-p)U_Y^0 L' \right] \frac{dm_2}{dm_1} &= 0.
\end{align*}
\]

(31)

As in the case of linear insurance, at the optimum value \( m_2 \), the bracketed term which includes \( dm_2/dm_1 \) vanishes, such that in a properly decentralized organization of \((m_1, m_2)\) the “protection branch” can ignore it. However we cannot ignore the effects of protection on the price of insurance, so that the third bracketed term in Eq. (31) must be incorporated in the protection branch calculus.

3.2.1 Organization and information sharing between protection and insurance branches

We can consolidate our argument for information sharing among government branches if we think of implementation as solving these two equations:

\[
\begin{align*}
\frac{\partial W}{\partial m_1} &= \phi^1(m_1, m_2) = 0 :\Rightarrow m_1 = f^1(m_2) \quad (32a)
\end{align*}
\]

and

\[
\begin{align*}
\frac{\partial W}{\partial m_2} &= \phi^2(m_1, m_2) = 0 :\Rightarrow m_2 = f^2(m_1) \quad (32b)
\end{align*}
\]

To illustrate the solution draw the “reaction” curves \( f^2(m_1) = m_2 \) and \( f^1(m_2) = m_1 \) over contours of “constant security" as in Fig. 2. At their intersection we would have the true optimal values of \( m_1^* \) and \( m_2^* \).

To see the effects of compartmentalization, suppose, rather than a unitary decision maker, an insurance branch within the government and a protection branch are each charged with providing \( m_2 \) and \( m_1 \) respectively. If we assume full information sharing then that government would understand its welfare to be as shown in Eqs. (32a, 32b) and would strive to implement \( m_1^* \) and \( m_2^* \). How could it do that? Here are some approaches.

\[\text{From Eq. (30)}: \quad \frac{dm_2}{dm_1} = - \frac{\partial S/\partial m_1}{\partial S/\partial m_2} = - \left[ \frac{-U_{YY}^1 \left[ -1 + m_2 p'/p^2 \right] - U_{Y2}^0 L'}{U_{YY}^1 \pi + U_{YY}^2 L' + U_{YY}^0 (L')^2} \right] \frac{dm_2}{dm_1} \geq 0 \quad (33). \]

By SOCs for Eq. (30) the denominator of Eq. (33) is negative. As a special case, if \( L' = 1 \) at the insurance optimum of Eq. (30) so that \( U^1 = U^0; \ U_Y^1 = U_Y^0; \ U_{YY}^1 = U_{YY}^0 \) then the sign of Eq. (33) is > 0 so that just as in the linear insurance case, quantity of coverage \( m_2 \) and of protection are gross complements. We can also simplify Eq. (33) using the definition \( R = -U_{YY}/U_Y \), where \( R \) is the coefficient of risk aversion. Then \( dm_2/dm_1 = - \left[ R_1 U_Y^1 + R_0 U_Y^0 L' - U_{YY}^0 \pi m_2 \right] \left[ 1/(p^2) \right] \left[ 1/(p^2) \right] \) (34). Then, under constant risk aversion, from the FOC the first two terms of the numerator cancel so that increases in expenditures on protection raise optimized quantity of self insurance coverage that is \( dm_2/dm_1 > 0 \) and we know in this case \( m_2 \) is a borderline normal good. Moreover, even if risk aversion is declining so that \( R_1 < R_0 \) still \( m_2 \) may be gross complements since only a very great value of \( R_0 \) will offset the positive weight of the final term in the numerator.
3.2.2 Bureaucratic coordination

A. Black box approach: Just assume that the government makes all the calculations, derives the answer and instructs the insurance branch to spend $m_2^*$ and the protection branch to spend $m_1^*$. This approach avoids the governance problem of externalities as between protection and insurance branches by assuming it away.

B. Central control plus decentralized execution: The government informs each branch of its correct reaction function $m_1 = g_1(m_2)$ and $m_2 = g_2(m_1)$. (Note that the first of these incorporates the effect of $m_1$ on $\pi$ and the insurance-cost-savings attributable to protection outlays). From any point in Fig. 2, each branch moves toward its correct reaction function in Cournot fashion, until the intersection optimum of $m_1^*$ and $m_2^*$ is reached.

C. Hierarchical execution. The government entrusts one branch to make a final “all at once decision.” This avoids the zig-zag iteration implied by B. Instead, the government gives the function $m_2 = f_2(m_1)$ to the protection branch or the function $m_1 = f_1(m_2)$ to the insurance branch. The “Leader-branch” that receives the reaction function then implements it and solves for the overall optimum and provides to the “Follower branch” its solution value of $m_L$ (“L” for leader) after which the “Follower-branch” just follows its reaction function, providing $m_F$ as anticipated by the leader.

To sum up, if a large agent the size of a nation, may spend on both insurance and protection, we expect that both $m_1$ and $m_2$ should (if information is shared and coordinated) be provided at an interior solution. But if information is not correctly distributed and recognized—specifically, if the insurance cost savings from improvements in self protection are not allowed to influence those self-protection decisions—then in aggregate decisions will be suboptimal with losses from extreme moral hazard and corner solutions where only $m_2$ is provided, just like the benchmark case of market insurance in a small open price taking economy. Note that if we ignore the third term in Eq. (31),
the sign of Eq. (31) is likely to become negative. For example, if \( l' \geq 1 \), the sign of Eq. (31) becomes negative and we have the extreme moral hazard and corner solutions.

### 4 Conclusion

This paper has investigated two types of preparation available to expected utility maximizing agents faced with “costs of emergency”, namely self-insurance and self-protection. Self-insurance provided to itself by a large entity such as a nation differs in two important respects from standard market insurance. First of all, the self-insurance function should show diminishing returns or increasing costs whereas market insurance typically has linear or piecewise linear pricing. Second, self-insuring countries far more readily than individuals should have access to actuarially fair prices, which they can then incorporate into their optimization.

If a small price taking agent provides insurance at a market price, then the standard result holds that with pricing fair and linear, complete coverage (net of premiums) is purchased. We have used this standard result as a benchmark to compare with self-insurance by an entire country, and also in our examination of the effect of insurance on the choice of protection.

First, we show that the effect of decreasing productivity (ordinary diminishing returns) when a nation provides self insurance is for complete coverage (net of premiums) not to be purchased except by unlikely chance.

Second we describe the incentive structure when both types of protection are available. In doing this we have focused on the mis-incentives that follow when providers of protection and of insurance (within the same government) ignore the benefits and cost they confer on each other. More specifically, we show that without recognition of such external benefits and costs, the security providers of a country are likely to drive themselves into a corner, in effect, of extreme moral hazard where insurance is over-provided and protection abandoned. Thus, an agent trying to provide both self-insurance and self-protection may find that both will never be provided at the same time because of ignorance of the insurance cost savings attributable to protection outlays. The resolution of this mis-incentive requires that the providers of risk-reducing protection recognize in their calculus the cost-savings they provide to insurers. Since improvements in protection lower the actuarially fair price available to self-insuring agents, the decision of how much protection to provide must incorporate this benefit. Once such benefit is recognized and included in the calculus of protection, the moral hazard corner will be avoided.

Thus, we have shown that separation of security decision-making into self-protection and self-insurance is a bad idea. In fact most governments, at present, treat them as essentially separate. There needs to be much more coordination between these two seemingly disparate functions of government. Also, our arguments conclude that unless properly coordinated, self-insurance will push out self-protection completely, driving \( m_1 \) to zero in our terminology. This says that there is far too little consideration given to measures that will improve the odds against war, catastrophe etc (not particularly military measures, all measures). Our arguments lead to the implication
that too much attention tends to be given to insurance provision and too little to risk improvement.

Open Access  This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

The concept and measurement of actuarially fair pricing is crucial as a benchmark throughout the analysis of insurance. This appendix dwells briefly on an ambiguity in definition of actuarial fairness that arises when insurance is self-provided with diminishing returns.

The paper’s main text assumes one simple relationship between premiums paid, \( \pi m_2 \), during good times and \( l(m_2) \) benefits received during bad times, where the variable of choice is units of coverage \( m_2 \). As pointed out in footnote 3, we could cast the problem in terms of units of expenditure (where we define \( x_2 = \pi m_2 \), “\( x \) for” expenses,” and where \( m_2 = x_2/\pi \) such that the variable of choice is \( x \). Here \( 1/\pi \) would indicate the efficiency of resource transfer across contingencies. But that suggests that in place of Eq. 14 we could write either Eq. (14a) or (14b):

\[
W = pU^1[Y - \pi m_2] + (1 - p)U^0[Y - \{L - L(m_2)\}] : m_1 \text{ not shown} \\
W = pU^1[Y - x_2] + (1 - p)U^0[Y - \{L - L(x_2/\pi)\}] : m_1 \text{ not shown} \tag{14a}
\]

\[
W = pU^1[Y - x_2] + (1 - p)U^0\left[ Y - \left\{ L - \frac{1}{\pi} L(x_2) \right\} \right] : m_1 \text{ not shown}. \tag{14b}
\]

When transfer of resources across contingencies is written as in Eq. (14a) it becomes obvious on inspection that every unit of \( x_2 \) may not transfer into just exactly \( x_2/\pi \) units to the bad contingency. Once this is recognized then the idea of “fair” pricing becomes ambiguous.

In Eq. (14b) \( 1/\pi \) indicates not the productivity or accumulation of \( x_2 \)-resources but the accumulation in bad times of \( l \) units of benefit. Now, in a sense, the \( l \)-function operates during good times to create transfers available only under adversity; so that for the total \( l/\pi \) received in adversity, \( x_2 \) was set aside in the good contingency.\(^\text{12}\)

\(^{12}\) One way to think about this is to assume there is a steady state. We can then frame self-insurance in terms of changes in, or alternative parameters for this steady state. Normalizing the notation of note 1, let \( w = 1, d = n \). Here is the steady state. Every \( n + 1 \) years there is a crisis, requiring that the self-insurance accumulation be utilized. Each year from 1 to \( n \), \$\( m_2 \) is set aside in anticipation of year \( n + 1 \). That is “self-insurance” is fair. At year \( n + 2 \) the process starts over again, and on and on. Therefore \( p = n/[n + 1] \), and \( 1 - p = 1/[n + 1] \), \( \pi = 1/n \) so that \( m_2/\pi = nm_2 \). Changes in the risk of adversity then are given by changes in \( n \). If \( n \) is large adversity happens only rarely so that \( 1 - p \) is small, and if \( n \) is small, the risk of adversity \( 1 - p \) is high. We ignore discounting, and uncertainty.
When insurance is supplied in the market and its marginal costs and benefit are linear, this distinction between Eqs. (14a) and (14b) does not arise or it doesn’t matter; but when \( I \) displays diminishing returns how to describe the cost of insurance seems to be ambiguous. Do diminishing returns apply both to quantities reserved and to the lapse of time between adversities, or only to the former? Between Eqs. (14a) and (14b), the difference seems to lie in “when” the resources transferred across contingencies become productive—before (as in Eq. (14b)) or after (as in Eq. (14a)) the transfer. The novelty of this distinction, or the fact that it seems to have been overlooked, may be because when the insurance function is linear, as in market insurance, no such difference arises.

For either of these options Eq. (14a) or (14b) the idea of actuarial fairness is still important, but its implications with respect to outcomes vary. For example, if marginal utilities just happened to be equalized at the optimum such that \( I' = 1 \) then consumption would also be equalized. It turns out that this would imply \( L(m_2) = (1 - p)L \), in the case of Eq. (14b) or \( L(x_2/\pi) = (1 - p)L \) in the case of Eq. (14a). Although symmetric to market insurance neither of these conditions is required to obtain even when coverage is fairly priced—not required because there is no necessity that \( I' = 1 \). In either case FOCs inform us of optimal protection and allow comparison Eqs. (14a) and (14b), assuming self-insurance to be fairly priced. Generally optimal provision may be greater under formulation Eq. (14a) or (14b) depending on the manner in which \( p \), \((1 - p)\) influence MRS and MRT in each formulation. Moreover the benefit in the bad contingency might depend in a more complicated manner on resources set aside in good times than either Eq. (14a) or (14b). Then we would want to write \( L = L(\pi, p, x_2) \) and the definition of fair insurance correspondingly more ambiguous.

References


Footnote 12 continued

Case 1 Here the total insurance consumed in year \( n + 1 \) is \( nL(m) \). Here diminishing returns apply to each year’s insurance savings individually so that every new year the country begins a new decreasing returns process.

Case 2 Here the total insurance coverage consumed in year \( n + 1 \) is \( L(nm) \). Here \( m_2 \)-savings set aside each year generate less marginal return than the same \( m \) savings of the year before.

In both I and II, greater \( m_2 \) produces more diminishing returns, although in different manner and to different degree. Note however, that with case II, diminishing returns, in addition to being greater for larger values of \( m_2 \), are more severe the rarer the emergency (higher value of \( n \)).
Ihori T, McGuire MC (2008b) Alliance interactions in the provision of mutual insurance and mutual protection: problems of collaboration when the collective good is security