



HAL
open science

Self-insurance and market insurance substitutability: an established tenet reconsidered

Jean-Marc Bourgeon, Pierre Picard

► **To cite this version:**

Jean-Marc Bourgeon, Pierre Picard. Self-insurance and market insurance substitutability: an established tenet reconsidered. 2024. hal-04401333

HAL Id: hal-04401333

<https://hal.science/hal-04401333>

Preprint submitted on 17 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Self-insurance and market insurance substitutability: an established tenet reconsidered

Jean-Marc Bourgeon* Pierre Picard†

January 17, 2024

Abstract

We analyze the interaction between self-insurance and market insurance when accident losses are multivalued. We show that self-insurance and market insurance may be complementary when self-insurance expenses do not affect much the probability distribution of large losses and the loading factor is high. This contrasts sharply with the conclusion of Ehrlich and Becker (1972) who establish the substitutability between self-insurance and market insurance when the cost of an accident is single-valued. The results are derived theoretically and illustrated through numerical simulations.

Keywords: insurance, self-insurance, contracts.

JEL Classification Numbers: D86, G22, D80.

*INRAE and CREST- Ecole Polytechnique. E-mail: bourgeon@agroparistech.fr

†CREST- Ecole Polytechnique. E-mail: pierre.picard@polytechnique.edu.

1 Introduction

The article of Ehrlich and Becker (1972) on risk prevention and insurance is a widely recognized cornerstone of insurance economic theory. Ehrlich and Becker (1972) distinguish self-insurance from self-protection. Self-insurance (or loss reduction) consists in making preventive expenses that decrease the damages caused by accidents that may occur, while self-protection (or loss prevention) refers to costly measures that reduce the probability of such accidents. They show that self-insurance and market insurance are substitutable, while self-protection and market insurance may be complementary. In concrete terms, when a decrease in the price of market insurance makes it more attractive, individuals reduce self-insurance and they may increase self-protection. The objective of this paper is to reconsider this alleged substitutability between market insurance and self-insurance and to highlight its lack of robustness. As illustrated by debates on the insurability of climate-related risks in the global warming context, this is not a purely academic issue: when the price of market insurance is affected by insurers' costs and by taxes and subsidies, while self-insurance exerts externalities on social welfare, their interaction is often a crucial element that should be taken into account in government decision-making.

Ehrlich and Becker (1972) analyze self-insurance in a two-state model, in which a risk-averse individual faces the risk of an accident that may occur with a given probability. He may protect himself against the potential damages of such an accident, through market insurance or self-insurance. Both risk-management tools consist in substituting wealth from the no-accident state to the accident state. In the case of market insurance, this substitution of wealth between states of nature goes through the payment of an insurance premium paid in both states in compensation for an indemnity larger the premium received in the accident state. In the case of self-insurance, preventive expenses also borne in both states allow the individual to reduce the damages that would result from an accident. In such a setting, it comes as no surprise that an increase in the price of market insurance leads the individual to rely more on self-insurance through a standard effect of relative prices on customer's optimal choices.

In what follows, we show that this substitutability between market insurance and self-insurance does not extend beyond the two-state model considered by Ehrlich and Becker (1972). Indeed, consider the case where the insurance contract is a straight deductible, i.e. an insurance policy that covers the difference (if positive) between the incurred damages and the deductible. We know from Arrow (1971) that such a contract is optimal when the price of insurance is affected by a constant loading factor.

To fix ideas, assume that the losses incurred in case of an accident may be small, medium or large, and that the loading factor is so large that only large losses are covered in an optimal contract (i.e. the deductible is larger than medium-sized losses). Then, by a standard price effect, a decrease in the loading factor leads the individual to purchase more insurance, i.e., to choose a contract with a lower deductible: he receives a larger compensation for large losses, while, if the change in the loading remains small, there is still no compensation for small or medium-sized losses. Assume furthermore that self-insurance shifts probability weights from medium-sized losses to low-sized losses, without significant effect on the probability of large losses. In that case, the improvement in the coverage of large losses makes the individual more willing to self-insure. This is because these expenses are paid in all possible states, including the worst case scenario with large losses, which is now less painful because of the better insurance coverage. We will show that the complementarity between market insurance and self-insurance emerges in such a case: a decrease in the loading factor (i.e. a cheaper insurance coverage) leads the individual to purchase more market insurance, by a standard price effect, and this improvement in the insurance coverage (that boils down to increasing the worst-case bound of final wealth in the above three-state example) stimulates the individual's willingness to pay for self-insurance. A decrease in the loading factor also reduces the savings in insurance premium generated by self-insurance. This countervailing insurance premium effect is dominated by the insurance coverage effect when self-insurance does not strongly affect the probability of incurring the largest losses and the loading factor is large. On the contrary, market insurance and self-insurance may be substitutable when insurance is available at a lower price, with coverage of a wide spectrum of possible losses, be they small or large, because the insurance premium effect dominates the insurance coverage effect in that case.

One of the main advantages of the two-state setup of Ehrlich and Becker (1972) is the clear distinction between self-protection and self-insurance. In doing so, the specificity of the interaction between market insurance and each of these risk-management behaviors is highlighted in a very useful way. In more general terms, we may consider that these risk-management expenses are translated into a first-order dominance negative shift in the probability distribution of possible losses. This is pure self-insurance if the probability mass of the no-loss state is unaffected, while pure self-protection corresponds to the case where this probability mass is increased, with an unchanged probability distribution of losses conditionally on a loss occurring. In practice, both forms of risk prevention frequently interact, and we will show that our conclusion about

the complementarity between self-insurance and market insurance is robust to the co-existence of self-insurance and self-protection. We will also show that the risk exposure with self-insurance (possibly coexisting with self-protection) can be deduced from applying a deterministic function (called the *residual loss function*) to the random losses that would be incurred in the absence of self-insurance. A simple condition on the residual loss function will allow us to characterize the case where self-insurance and market insurance are complements.

Very few articles have studied self-insurance with multiple loss levels. They show that the comparative static results obtained in the two-state case are not necessarily valid in the multi-state case. Indeed, in the two-state case, self-insurance shares the same properties as market insurance because loss reduction efforts and insurance coverage produce similar hedging benefits by substituting wealth from the no-accident state to the accident state. In the multi-state case, self-insurance measures may affect the probability distribution of losses very differently, depending on the technology used and the nature of the risk. Hence, the protection offered by self-insurance can be very different from the coverage brought about by the optimal insurance contract. The impacts of these differences on self-insurance choices are illustrated by Hiebert (1989), who considers variations in risk aversion, and by Lee (2010) who considers wealth effects. Other studies kept the assumption of a single-valued self-insurable loss, and focused attention on the effect of an additional source of uncertainty, either under the form of non-reliable self-insurance (Briys et al., 1991) or by adding an additional independent background risk to the self-insurance problem (Mahul, 1997). For our part, we consider a general setup with a continuum of losses and we show that self-insurance and market insurance may be complementary for high loading rates, and substitutable for low loading rates, due to the balance between the insurance-coverage effect and the insurance-premium effect that tips in different directions in each case.

The economic policy dimension of this issue is particularly relevant in several areas, and it is worth emphasizing it. Of particular importance is the case of agriculture, since farmers and other agricultural businesses face significant risks from natural hazards, pests and other unforeseen events. The high transaction costs, strengthened by the insurers' capital costs in the climate change context, make crop insurance expensive, and there is intense debate about whether and to what extent crop insurance should be subsidized.¹ Some of these debates, starting in the 1990's, centered on whether crop insurance could be a substitute to pesticides, fertilizers, crop specialization or

¹See for instance Gao (2023) in the case of the US crop insurance program.

irrigation, viewed as alternative ways to protect farmers against risks affecting their crop.² This line of research, particularly Weber et al. (2016), suggests that, in practice, expanded insurance coverage has little effect on technologies used by farmers. Our approach is coherent with these findings: pesticides and other risk-reducing devices may be considered as self-insurance, and there is no reason to think that they should be substitutable to insurance. In a more theory-oriented perspective and in line with recent works in ecology, Quaas and Baumgärtner (2008) consider the contribution of farmers to ecosystem quality and biodiversity as self-insurance.³ They develop a model inspired from Ehrlich and Becker (1972) in which this form of self-insurance (which they call “natural insurance”) and market insurance are substitutes, with the ultimate conclusion that improving access to market insurance leads to a lower level of ecosystem quality. Our model takes a different view on this issue, by putting into question the alleged substitutability between these two forms of insurance.

Beyond the case of crop insurance, the insurability of large risks is increasingly becoming a major concern due to the high capital costs required to cover risks with a systemic dimension (see Hartwig et al., 2020, in the case of economic losses caused by pandemics). The question of whether or not the government should provide funding for individuals and firms to buy insurance in those cases, along with implementing policies such as insurance mandates and post-disaster relief programs, is becoming increasingly important (see Boomhower et al., 2023, for an analysis of these issues in the case of wildfire insurance). In such cases, understanding how government funding affects self-insurance choices is of major importance.

The rest of the paper is organized as follows: in Section 2, we describe our setup. Section 3 studies the interaction between market insurance and self-insurance in the constant absolute risk aversion (CARA) case. In Section 4, we extend our results to the case where the individual is prudent, without postulating constant absolute risk aversion. The last section wraps up. All proofs are in the Appendix.

²See Horowitz and Lichtenberg (1993), Smith and Goodwin (1996), Wu (1999), and more recently Weber et al. (2016) for a study with US data, and Chakir and Hardelin (2014) and Enjolras and Aubert (2020) for studies with French data.

³The effect of biodiversity on the productivity of ecosystems is a major field of research in ecology and theoretical biology, the main conclusion being that species richness favors the intertemporal stability of ecosystems. Biodiversity is viewed as a “natural insurance” device that acts either by guaranteeing that some species maintains when other fails as in portfolio management, or by creating asynchronicity between fluctuations in the development of species when the ecosystem may be affected by environmental shocks such as heatwaves or rain deficit. See Yachi and Loreau (1999), Cottingham et al. (2001) and Shanafelt et al. (2015) for illustrative examples of this line of research.

2 The model

Consider a risk-averse individual (household or business) who has initial wealth w and may incur a loss $x \in (0, \bar{x}]$ in the case of an accident. Should an accident occur, the extent of the loss depends on the individual's behavior: he can reduce it by taking appropriate self-insurance measures. We denote by $a \geq 0$ the corresponding expenditures and by $F(x|a)$ the resulting cumulative probability distribution function of the loss x . There is a (possibly zero) mass of probability $F(0|a) = \bar{F}(a)$ for the no-accident state, and $F(x|a)$ is differentiable with respect to x when $x > 0$. Here and in what follows subscripts refer to partial derivatives, and we denote $f(x|a) = F_x(x|a)$. Hence, $f(x|a)/[1 - \bar{F}(a)]$ is the density of the loss, conditionally on an accident occurring. We assume that increasing a improves this distribution in the sense of the First-Order Stochastic Dominance (FOSD), i.e. that $F_a(x|a) \geq 0$ with strict inequalities in a subset of $(0, \bar{x})$ of positive measure. We also assume $F_{aa}(x|a) \leq 0$, which corresponds to the decreasing marginal efficiency of self-insurance. The case $F(0|a) = \bar{F} \geq 0$ for all a , a constant, corresponds to the “pure” self-insurance case as defined by Ehrlich and Becker (1972). It means that the probability of a loss occurring, $1 - \bar{F}$, does not depend on a . In what follows, we assume that $F(0|a)$ may depend on a (i.e. $\bar{F}(a)$ with $\bar{F}'(a) > 0$ for some a), a setting in which self-insurance measures also affect self-protection.⁴

We assume $d\mathbb{E}[x|a]/da < -1$ at $a = 0$, and furthermore we have $d^2\mathbb{E}[x|a]/da^2 = -\int_0^{\bar{x}} F_{aa}(x|a)dx \geq 0$: hence, an increase in self-insurance expenses reduces the expected loss by an amount larger than its expected cost (at least when a is small), but with a decreasing marginal efficiency. We also assume that a is verifiable, i.e., there is no moral hazard impeding the contractual relationship between the insurer and the insured.

The policyholder's preferences are characterized by Von Neumann-Morgenstern utility function $u(w_f)$ that satisfies $u' > 0$, $u'' < 0$ at any final wealth level w_f . The market insurance contract specifies a premium P paid at the outset, and an indemnity $I(x)$ for all $x \in [0, \bar{x}]$, and thus $w_f = w - x - P - a + I(x)$. Assuming that the insurance market

⁴For the sake of notational simplicity, we assume that the loss upper limit \bar{x} does not depend on a , and that $f(x|a)$ is twice-differentiable almost everywhere in \mathbb{R}_+^2 with $f(\bar{x}|a) = 0$. We consider the case where \bar{x} may depend on a in Section 3.4. We also assume that $x \mapsto F_a(x|a)$ is continuous in \mathbb{R}_+^* (and right continuous at $x = 0$), for all $a \geq 0$.

is competitive, the optimal insurance contract and self-insurance level are solution of

$$\begin{aligned} \max_{I(\cdot), P, a} \mathbb{E}u &= \int_0^{\bar{x}} u(w - x - P - a + I(x))dF(x|a) \\ \text{s.t.} & \\ P &\geq (1 + \sigma) \int_0^{\bar{x}} I(x)dF(x|a), \\ I(x) &\geq 0 \text{ for all } x \in [0, \bar{x}], \end{aligned}$$

where $\sigma > 0$ is the loading factor. We know from Arrow (1971) that, in that case, the optimal insurance contract entails full coverage over a deductible $D > 0$, leading to $I(x) = \max\{0, x - D\}$. The corresponding insurance premium is given by

$$P = \hat{P}(D, a, \sigma) \equiv (1 + \sigma) \int_D^{\bar{x}} (x - D)dF(x|a),$$

and the individual's program can be rewritten as

$$\max_{a, D} \mathbb{E}u \equiv \int_0^D u(w - x - \hat{P}(D, a, \sigma) - a)dF(x|a) + [1 - F(D|a)]u(w - D - \hat{P}(D, a, \sigma) - a). \quad (1)$$

Let $a^*(\sigma)$ and $D^*(\sigma)$ be the optimal solution to this problem when the loading factor is σ .⁵ An increase in σ makes insurance more expensive and this reduces the demand for market insurance under the form of a higher deductible. This direct price effect may be compensated by a wealth effect when absolute risk aversion changes with wealth, hence an ambiguous global effect in that case. Furthermore, the change in the optimal self-insurance expenses $a^*(\sigma)$ induced by the increase in σ modifies the individual's risk exposure, which also affects the optimal deductible $D^*(\sigma)$. If the direct price effect dominates the indirect consequences of the wealth effect and of the change in risk exposure, then $D^*(\sigma)$ is an increasing function. In that case, market insurance and self-insurance are substitutable risk management tools when $a^*(\sigma)$ is increasing, and they are complementary when $a^*(\sigma)$ is decreasing.

⁵In what follows, it is assumed that $a^*(\cdot)$ and $D^*(\cdot)$ are continuous functions from \mathbb{R}_+ to \mathbb{R}_+ . In full generality, we cannot exclude that $a^*(\sigma)$ and/or $D^*(\sigma)$ are multivalued for some σ . By a standard continuity argument of consumer demand theory (see Proposition 3.AA.1 in Mas-Colell et al., 1995), $(a^*(\cdot), D^*(\cdot))$ would be an upper hemicontinuous correspondence in that case. However, since $a^*(\sigma)$ and $D^*(\sigma)$ are uniquely defined when σ is either equal to 0 or large enough, we can conclude that $a^*(\cdot)$ and $D^*(\cdot)$ are continuous functions at least when σ is small or large enough.

3 Interaction between market insurance and self-insurance

We will start by focusing attention on the two polar cases that correspond to large and small values of the loading factor σ , with large and small values of the optimal deductible $D^*(\sigma)$, respectively. The case of constant absolute risk aversion will allow us to show that market insurance and self-insurance may be complementary or substitutable, according to the value of σ . We then examine the full scope of the loading factor, i.e. from 0 to the largest value beyond which the individual does not purchase market insurance, and present conditions that allow for a monotonic deductible and a bell-shaped (i.e. first increasing and then decreasing) self-insurance expenditure. We illustrate these theoretical results through simulations before examining a case where self-insurance and market insurance are substitutable for all values of σ . Finally, our results will be interpreted in a very intuitive way through the notion of residual loss function.

3.1 Two polar cases

Let us focus attention on the case where more self-insurance (i.e. an increase in a) does not affect significantly the probability of incurring the largest losses.⁶ This is obviously the case if, for any a , there exists $\hat{x}(a)$ smaller than \bar{x} such that an increase in a does not affect $F(x|a)$ when x is in the interval $[\hat{x}(a), \bar{x}]$. Fig. 1 illustrates this extreme case: an increase in self-insurance only reduces the probability of incurring medium-sized losses: an increase in a shifts $F(x|a)$ upwards when x is smaller than $\hat{x}(a)$, without change for loss levels larger than $\hat{x}(a)$. In what follows, more generally, the relative weakness of self-insurance in the prevention of the largest risks corresponds to the case where $F_a(x|a)$ is low (more precisely, smaller than $f(x|a)$) when x is close to \bar{x} .

Market insurance is attractive when the loading factor is not too large, and we denote $\bar{\sigma}$ its upward limit, above which the individual does not purchase insurance, that is $\bar{\sigma} = \sup\{\sigma \in \mathbb{R}_+ \mid D^*(\sigma) < \bar{x}\} > 0$, with $D^*(\sigma)$ increasing when σ is close

⁶This case is illustrated by Lee (2010) in his study of wealth effects in self-insurance: as he writes, the construction code may significantly reduce losses resulting from a small earthquake, because design constraints enable buildings to withstand earthquake forces up to some limit, but a strong earthquake may destroy all buildings and make these constraints almost useless. Similarly, in ecosystem management, preserving biodiversity and adopting less intensive agricultural practices may protect farmers from short-term droughts, but this would not protect them from the consequences of a catastrophic long-term drought.

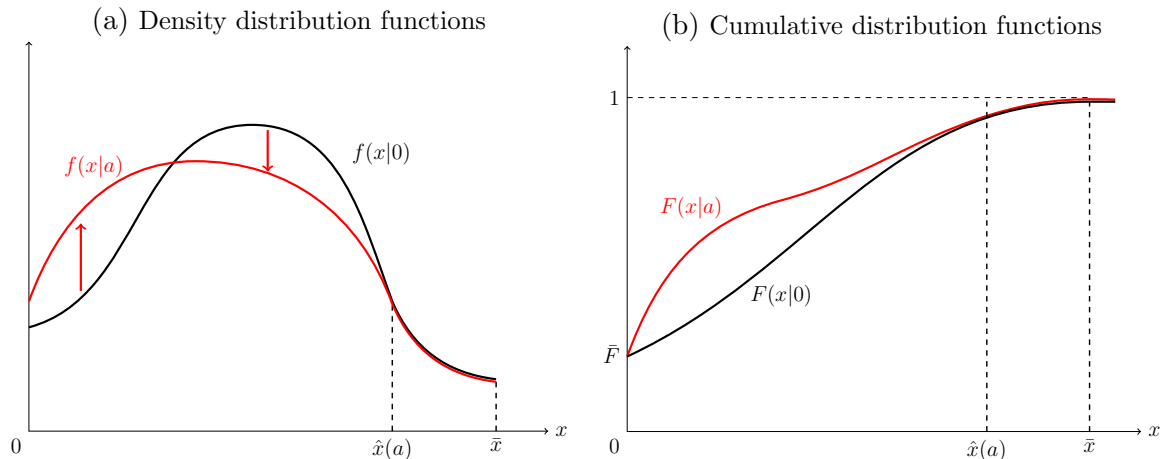


Figure 1: Shifts in the p.d.f. and in the c.d.f. when self-insurance increases from 0 to $a > 0$.

to $\bar{\sigma}$ and $D^*(\bar{\sigma}) = \bar{x}$. We also have $D^*(0) = 0$ because full coverage is optimal when insurance is purchased at actuarial price, and $D^*(\sigma) > 0$ when $\sigma > 0$, which implies that $D^*(\sigma)$ is increasing when σ is small. In other words, at least when σ is close to 0 and $\bar{\sigma}$, the direct price effect on $D^*(\sigma)$ dominates the indirect consequences of changes in wealth and risk exposure when σ changes. For intermediate values of σ , $D^*(\sigma)$ may not be monotonic, although we may find assumptions under which this will be the case (see Section 3.2). This section restricts attention to the case of constant absolute risk aversion (i.e. u is a CARA function) where the wealth effect on optimal choices vanishes. The more general case of a prudent individual will be considered in Section 4. Proposition 1 shows that market insurance and self-insurance are complementary when self-insurance does not affect significantly the probability of incurring the largest-size losses and the loading factor σ is large enough.

Proposition 1 *Assume that, for all $a > 0$, there exists $\hat{x}(a)$ smaller than \bar{x} such that $F_a(x|a) < f(x|a)$ if $x \in [\hat{x}(a), \bar{x}]$. If the individual displays constant absolute risk aversion, then $D^*(\sigma)$ is increasing and $a^*(\sigma)$ is decreasing when σ is close to $\bar{\sigma}$.*

The condition on the distribution of losses in Proposition 1 stipulates that self-insurance does not strongly affect the likelihood of the largest possible loss values. More precisely, let us define the “global cost of risk” as the sum of the loss and the self-insurance expenses, $z \equiv x + a$, which is distributed according to c.d.f. $G(z|a) = F(z - a|a)$, with $G(a|a) = \bar{F}$, over the support $[a, a + \bar{x}]$. We thus have $G_a(z|a) = F_a(z - a|a) - f(z - a|a)$ and we will say that self-insurance expenses a are locally weak

(resp. strong) at loss level $x = x_0$ when $G_a(x_0 + a|a)$ is negative (positive). The weak or strong power of self-insurance thus characterizes its marginal efficiency depending on the level of loss considered. In words, self-insurance is strong at loss level $x_0 \in (0, \bar{x}]$ when increasing a reduces the probability that the global cost of risk will be larger than $x_0 + a$, and it is locally weak otherwise. Clearly self-insurance is locally weak in the neighborhood of \bar{x} in the extreme case described in Fig. 1, but of course, the validity of Proposition 1 is broader.

In order to intuitively interpret Proposition 1, consider how market insurance affects the individual's willingness to pay for self-insurance. Denote by \bar{V} his expected utility in a status quo situation where he self-insures for an amount a while the deductible is D , and consider the supplementary self-insurance spending δ that the individual would be willing to make (for D unchanged) to reduce his risk exposure from $F(\cdot|a)$ to $F(\cdot|a + \varepsilon)$, $\varepsilon > 0$. In other words, $\delta \equiv \delta(\varepsilon, D)$ is defined by

$$\begin{aligned} \bar{V} = & \int_0^D u(w - x - \hat{P}(D, a + \varepsilon, \sigma) - a - \delta) dF(x|a + \varepsilon) \\ & + [1 - F(D, a + \varepsilon)]u(w - D - \hat{P}(D, a + \varepsilon, \sigma) - a - \delta), \end{aligned}$$

with $\delta(0, D) = 0$. Differentiating \bar{V} with respect to ε and δ yields the marginal willingness to pay for an improvement in the risk exposure:

$$\begin{aligned} \left. \frac{\partial \delta(\varepsilon, D)}{\partial \varepsilon} \right|_{\varepsilon=0} = & \frac{\int_0^D [u(w - x - P - a) - u(w - D - P - a)] dF_a(x|a)}{\partial \mathbb{E}u / \partial w} \\ & - \frac{\partial \hat{P}(D, a, \sigma)}{\partial a}, \end{aligned} \quad (2)$$

where $\partial \mathbb{E}u / \partial w > 0$ and $P = \hat{P}(D, a, \sigma)$ with

$$\frac{\partial \hat{P}(D, a, \sigma)}{\partial a} = -(1 + \sigma) \int_D^{\bar{x}} F_a(x|a) dx \leq 0, \quad (3)$$

since $F_a(x|a) \geq 0$. Equation (2) is the sum of two positive terms that correspond to the benefits of an increase in self-insurance for the insured. The first term is the monetary equivalent of the expected utility gain that results from the downward shift in losses, for a given insurance premium. It corresponds to the previously mentioned insurance coverage effect. This first term cancels when $D = 0$ and using $u' > 0$ and $F_a(x|a) \geq 0$ shows that it is positive when $D > 0$. The second term in (3) is the decrease in the insurance premium due to the decreased probability of losses above the deductible: it

is the insurance premium effect.

Under CARA preferences, the marginal willingness to pay is not affected by wealth effects. This makes the analysis of the insurance coverage effect much easier. To start with, consider the case illustrated in Fig. 1 where $F_a(D|a) = 0$ when D is large enough but lower than \bar{x} , a circumstance under which the insurance premium effect fully vanishes. Using (2) and integrating by parts give

$$\left. \frac{\partial^2 \delta(\varepsilon, D)}{\partial \varepsilon \partial D} \right|_{\varepsilon=0} = - \frac{[1 - F(D|a)] \left[e^{-\alpha D} F_a(0|a) + \int_0^D e^{-\alpha(D-x)} F_a(x|a) dx \right]}{\left[\int_0^D e^{-\alpha(D-x)} dF(x|a) + 1 - F(D|a) \right]^2} < 0,$$

where α is the index of absolute risk aversion. In a nutshell, in the case of Fig. 1, a smaller deductible increases the willingness to pay for a favorable shift of $F(\cdot|a)$ corresponding to a unit increase in a . In other words, in this case, because of the insurance coverage effect, the larger the deductible, the lower the marginal willingness to pay for self-insurance. This remains true if $F_a(x|a)$ is positive but not too large when x is close to \bar{x} , a case in which the insurance premium effect is positive but small.

The next Proposition considers the other polar case and shows that market-insurance and self-insurance are substitutable when the loading factor is small.

Proposition 2 *If the individual displays constant absolute risk aversion, $D^*(\sigma)$ and $a^*(\sigma)$ are increasing when σ is close to zero.*

The intuition of Proposition 2 is the following. We have $D^*(0) = 0$ and $D^*(\sigma) > 0$ when $\sigma > 0$, hence $D^*(\sigma)$ is locally increasing when σ is small. Since the policyholder faces no risk when $\sigma = 0$, his self-insurance level is the same as if he were risk-neutral, i.e. $a^*(0) = a_N$, where a_N solves $\partial \mathbb{E}[x|a_N]/\partial a = -1$. In other words, a_N equalizes the marginal decrease in expected loss with the marginal cost of self-insurance. When σ is close to 0, $D^*(\sigma)$ is also close to zero. In that case, the individual is almost fully covered and reducing the insurance premium is the main reason for increasing self-insurance. In other words, the insurance premium effect dominates the insurance coverage effect, and consequently an increase in σ triggers an increase in $a^*(\sigma)$ above a_N because of this cost-cutting objective. This mechanism may be related to the marginal willingness to pay for self insurance, since

$$\left. \frac{\partial \delta(\varepsilon, D)}{\partial \varepsilon} \right|_{\varepsilon=0} \simeq - \left. \frac{\partial \hat{P}(D, a, \sigma)}{\partial a} \right|_{D=0, a=a_N} = 1 + \sigma,$$

when D is close to 0. Hence, at the first order, an increase in σ in the neighborhood of 0 leads to an increase in the marginal willingness to pay for self-insurance from 1 (its equilibrium value when $D = 0$ and $a = a_N$) to $1 + \sigma$. This leads to simultaneous increases in $D^*(\sigma)$ and $a^*(\sigma)$ when σ increases, which corresponds to the substitutability between market insurance and self-insurance established in Proposition 2.

3.2 Full scope for the loading factor

The next two propositions provide sufficient conditions on the distribution of losses for $D^*(\sigma)$ to be increasing and $a^*(\sigma)$ to be bell-shaped when σ goes from 0 to $\bar{\sigma}$.

Proposition 3 *Assume $f_{aa}(x|a) - f_{ax}(x|a) \geq 0$ for all x such that $f_a(x|a) < 0$. Then, if the individual displays constant absolute risk aversion, $D^*(\sigma)$ is increasing for all $\sigma \in [0, \bar{\sigma}]$.*

Proposition 3 gives a sufficient condition for the deductible to increase with the loading factor σ in the CARA case. We know that this property is true in the pure market insurance problem without self-insurance. However, it is also known that, in this problem, a change in the probability distribution of losses affects the optimal deductible in an ambiguous way (see, e.g., Briys and Schlesinger, 1990). Here, when σ changes, in addition to the standard price effect on insurance demand, the optimal deductible $D^*(\sigma)$ is also affected by the change in the risk exposure induced by the variation in the optimal self-insurance level $a^*(\sigma)$. It is shown in the proof of the proposition that the price effect dominates, i.e. that $D^*(\sigma)$ is increasing for all $\sigma \in [0, \bar{\sigma}]$, when $f_{aa}(x|a) > f_{ax}(x|a)$ for all x such that $f_a(x|a) < 0$. This condition corresponds to a limit on the marginal efficiency of self-insurance, in the sense that it is more and more costly to reduce the probability of incurring large losses (i.e., loss values x such that $f_a(x|a) < 0$) when a is increasing.⁷ If this condition were not satisfied, then an increase in $a^*(\sigma)$ induced by a decrease in σ could reduce the risk exposure so strongly

⁷To be more specific, consider a status quo situation where $a = a_0$ and $x = x_0$ with $f_a(x_0|a_0) < 0$. Hence, $-f_a(x_0|a_0)$ is a measure of the marginal efficiency of self-insurance on losses around x_0 (whose probability decreases when self-insurance increases). Starting from this status quo, consider a small increase in self-insurance $da > 0$ compensated by a decrease in the loss level $dx = -da$ so that the global cost $x + a = x_0 + a_0$ is unchanged. This compensated increase in self-insurance yields

$$df_a = f_{ax}(x_0|a_0)dx + f_{aa}(x_0|a_0)da = [f_{aa}(x_0|a_0) - f_{ax}(x_0|a_0)]da > 0,$$

which means that, for a global cost $x + a = x_0 + a_0$, the marginal efficiency of self-insurance is decreasing. One may check that the condition is satisfied in the simulation presented in section 3.3.

that the insured would choose a higher deductible. The condition prevents to have such a Giffen effect on market-insurance.

Proposition 4 *Assume that (i) $\mathcal{C}(a) \equiv \{x|F_a(x|a) - f(x|a) \geq 0\}$ is a convex subset of $[0, \bar{x})$, and (ii) $\mathcal{C}(a') \subseteq \mathcal{C}(a)$ for all relevant $(a, a'), a < a'$. Then, if $D^*(\sigma)$ is an increasing function and if the individual displays constant absolute risk aversion, $a^*(\sigma)$ is a unimodal function when $\sigma \in [0, \bar{\sigma}]$.*

As in the previous propositions, the results are obtained under constant absolute risk aversion. Assumption (i) introduces $\mathcal{C}(a)$, the subset of $[0, \bar{x})$ in which $F_a(x|a)$ is larger or equal to $f(x|a)$, i.e. the subset of losses where self-insurance is locally strong. As in Proposition 1, $\mathcal{C}(a)$ does not include \bar{x} for all relevant values of a . More restrictively, according to assumptions (i) and (ii), $\mathcal{C}(a)$ is a sub-interval of $[0, \bar{x})$ that shrinks when a increases. In particular, $\mathcal{C}(a') \subseteq \mathcal{C}(a)$ when $a' > a$ conveys the idea that the marginal efficiency of self-insurance is decreasing when a is increasing.⁸

We already know from Propositions 1 and 2 that $a^*(\sigma)$ is increasing when σ is close to 0, and that it is decreasing when σ is close to $\bar{\sigma}$ if $F_a(x|a) < f(x|a)$ when x is close to its upper limit \bar{x} . Under the conditions postulated in Propositions 3 and 4, $a^*(\sigma)$ is bell-shaped over the whole interval $[0, \bar{\sigma}]$. In other words, because of the interaction between the insurance premium effect and the insurance coverage effect, there exists $\hat{\sigma} \in (0, \bar{\sigma})$ such that $a^*(\sigma)$ is monotonically increasing when $\sigma < \hat{\sigma}$ and monotonically decreasing when $\sigma > \hat{\sigma}$. Firstly, for a given deductible, the larger the loading factor σ , the larger the decrease in the insurance premium induced by an increase in self-insurance. This positive effect of an increase in σ on $a^*(\sigma)$ corresponds to the risk premium effect. However, and this is the channel of the insurance coverage effect, the increase in σ also triggers an increase in the optimal deductible $D^*(\sigma)$. As we have seen above, the increase in the deductible reduces the marginal willingness to pay for self-insurance, hence a countervailing effect of the increase of the deductible on the optimal self-insurance. This increase simultaneously reduces the intensity of the insurance premium effect and triggers the insurance coverage effect. When σ is larger than $\hat{\sigma}$, the balance tips in favor of reducing self-insurance expenses when the deductible increases.

⁸ $\mathcal{C}(a)$ is a sub-interval of $[0, \bar{x})$ if $F_a(x|a) - f(x|a)$ is quasi-convex in x . As shown in the proof of Proposition 3, the condition postulated in this proposition ensures that this function is non-increasing w.r.t. a , and thus $\mathcal{C}(a') \subseteq \mathcal{C}(a)$ if $a' > a$.

3.3 Illustration

For illustrative purposes, consider the case where $f(x|a) = (1 - \bar{F})h(x|a)$ where $\bar{F} > 0$ is a constant (pure self-insurance), and

$$h(x|a) = \frac{1}{\bar{x}} \left(1 - \sqrt{\frac{a}{\bar{a}}} \frac{3\sqrt{3}}{2} (2x/\hat{x} - 1)(1 - (2x/\hat{x} - 1)^2) \mathbf{1}_{\{x \in [0, \hat{x}]\}} \right),$$

with $a \in [0, \bar{a}]$ and $\hat{x} \leq \bar{x}$. Hence, $h(x|a)$ is the conditional density of the loss, with a sinusoidal shape when $x \leq \hat{x}$, and it is flat above \hat{x} , with $h(x|a) = 1/\bar{x}$ if $x \in [\hat{x}, \bar{x}]$, $h(x|a) > 1/\bar{x}$ if $x \in (0, \hat{x}/2)$ and $h(x|a) < 1/\bar{x}$ if $x \in (\hat{x}/2, \hat{x})$. Function $h(x|a)$ is illustrated by Fig. 2a in the cases $a = \bar{a}$ (the blue curve), and $a = \bar{a}/2$ (the red curve), with $\bar{a} = 10$, $\hat{x} = 70$ and $\bar{x} = 100$.⁹ Fig. 2b depicts the c.d.f. of the loss when a is at its maximum \bar{a} (the blue curve) and at its minimum $a = 0$ (the red line).

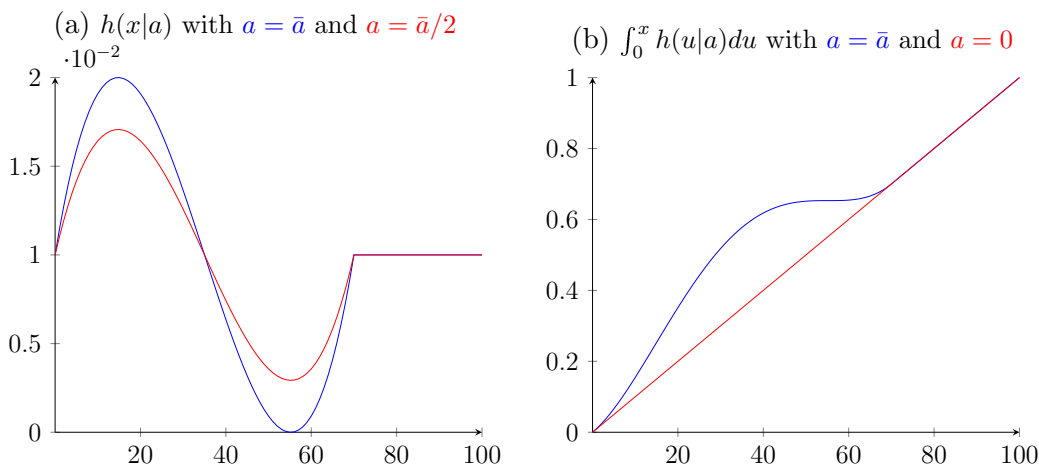


Figure 2: Conditional density and probability distributions of the loss ($\bar{x} = 100$, $\bar{a} = 10$ and $\hat{x} = 70$).

The parameter values used in the simulations are presented in Table 1,¹⁰ and under these assumptions, we have computed the optimal deductible and self-insurance levels when the loading factor σ goes from 0 to 150%.

The results are summarized in Table 2 and illustrated in Fig. 3, 4 and 5, with

⁹Conditional density $h(x|a)$ reaches a maximum when $x = \hat{x}(\sqrt{3} - 1)/2\sqrt{3}$ and a minimum when $x = \hat{x}(\sqrt{3} + 1)/2\sqrt{3}$.

¹⁰Normalization $\bar{a} = 10$ yields $a_N = 1.152$. Furthermore, assuming $\bar{F} = 0.20$ means that there is a 80% chance of incurring a loss, which would be uniformly distributed over $[0, \bar{x}]$ if the individual does not take any self-insurance measure. The value of the initial wealth level w is only relevant to derive the expected utility level reached in each case.

Table 1: Calibration values

α	w	\bar{F}	\bar{x}	\bar{a}	\hat{x}	a_N
0.05	20	0.2	100	10	70	1.152

Parameters: α is the ARA coefficient, w the initial wealth, \bar{F} the probability that no accident occurs, \bar{x} the maximal loss, \bar{a} the maximum self-insurance level and \hat{x} the maximum loss affected by self-insurance measures. a_N is the self-insurance level under risk neutrality.

Table 2: Simulation results

σ (%)	D^*	a^*	\hat{P}	$\mathbb{E}u$	$\mathbb{E}[x a^*]$	$\sigma_x(a^*)$
0	0.000	1.152	37.695	0.000	47.119	30.191
10	9.249	1.389	33.506	-0.474	46.837	30.302
20	16.734	1.601	30.321	-0.926	46.604	30.392
30	22.816	1.753	28.018	-1.358	46.446	30.452
40	27.976	1.836	26.267	-1.772	46.363	30.483
50	32.553	1.850	24.819	-2.172	46.350	30.488
60	36.774	1.797	23.500	-2.556	46.403	30.468
70	40.801	1.682	22.186	-2.923	46.520	30.424
80	44.742	1.514	20.786	-3.271	46.698	30.356
90	48.650	1.308	19.248	-3.596	46.931	30.265
100	52.508	1.088	17.581	-3.894	47.200	30.158
110	56.227	0.882	15.860	-4.163	47.480	30.043
120	59.694	0.709	14.196	-4.403	47.741	29.933
130	62.840	0.574	12.669	-4.614	47.966	29.836
140	65.655	0.474	11.316	-4.799	48.152	29.755
150	68.163	0.401	10.135	-4.962	48.301	29.689

Simulation results under parameter values given Table 1. $\mathbb{E}u$ is normalized to 0 for $\sigma = 0$. The last two columns $\mathbb{E}[x|a^*]$ and $\sigma_x(a^*)$ correspond to the expected loss and the standard deviation of the loss, respectively, conditional to the occurrence of an accident.

the loading factor (in percentage) on the horizontal axis.¹¹ A fair premium ($\sigma = 0$) leads to a complete insurance coverage (i.e. $D = 0$) with self-insurance expenses $a^*(0) = a_N = 1.152$. As the loading factor increases, the optimal deductible $D^*(\sigma)$ increases, while the insurance premium and the individual's expected utility decrease.¹²

Their paths are given Fig. 3a, 4a and 4b, respectively. Optimal self-insurance $a^*(\sigma)$ rises at first to reach 1.850 when $\sigma = 50\%$, and then decreases.¹³ Hence market insurance and self-insurance are substitutable when σ is smaller than 50%, and they are complementary for larger values of the loading factor.

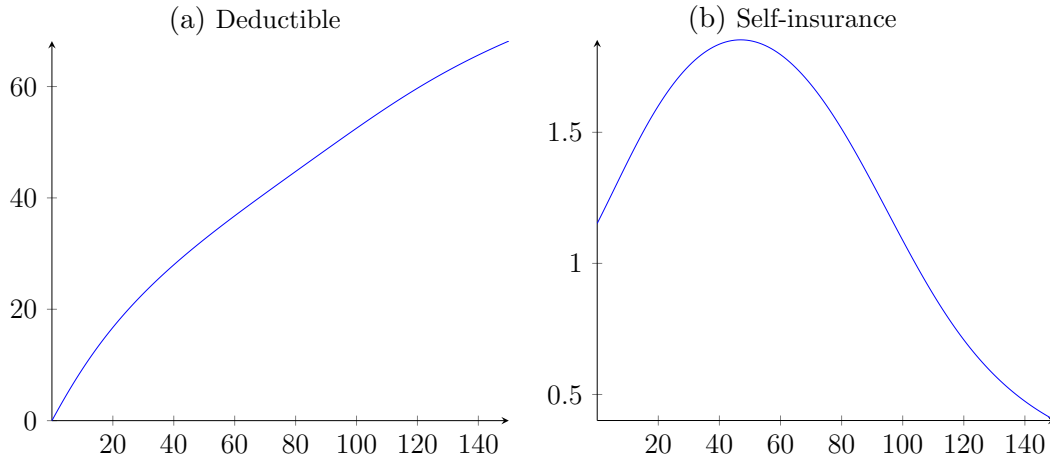


Figure 3: Deductible and self-insurance as functions of the loading factor.

The non-monotonic variations of the self-insurance expenses result in non-monotonic (and opposite) variations of the expected value $\mathbb{E}[x|a^*(\sigma)]$ and standard deviation $\sigma_x(a^*(\sigma))$ of the loss as illustrated in Fig. 5a and 5b, respectively.

3.4 Large-risk self-insurance

So far, we have presented conditions on the probability distribution of the loss that lead to the complementarity between self-insurance and market insurance, at least for high values of the loading factor. We believe that these conditions are representative of many real situations, but this obviously does not mean that strict substitutability

¹¹The simulations were performed with Mathematica.

¹²The monotonic decrease in the insurance premium P when σ is increasing is a consequence of the high value of the price-elasticity of insurance demand when $\alpha = 0.05$. A larger index of absolute risk aversion would correspond to a lower price-elasticity, and the relationship between P and σ would not be monotonic in that case.

¹³One can easily verify that the conditions of Proposition 4 hold, i.e. that $\mathcal{C}(a) \equiv \{x|F_a(x|a) - f(x|a) \geq 0\}$ is an interval that shrinks when a varies from 0.4 to 2.

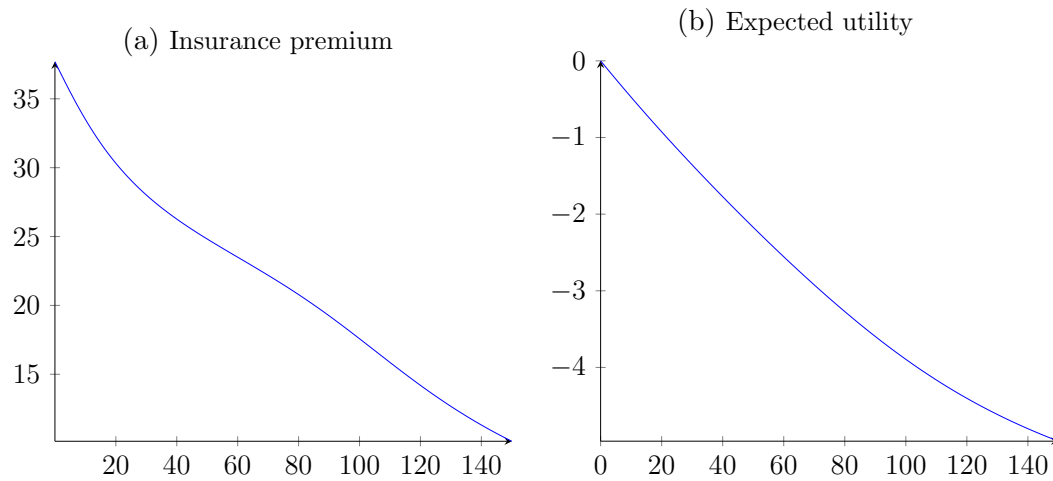


Figure 4: Premium and expected utility as functions of the loading factor.

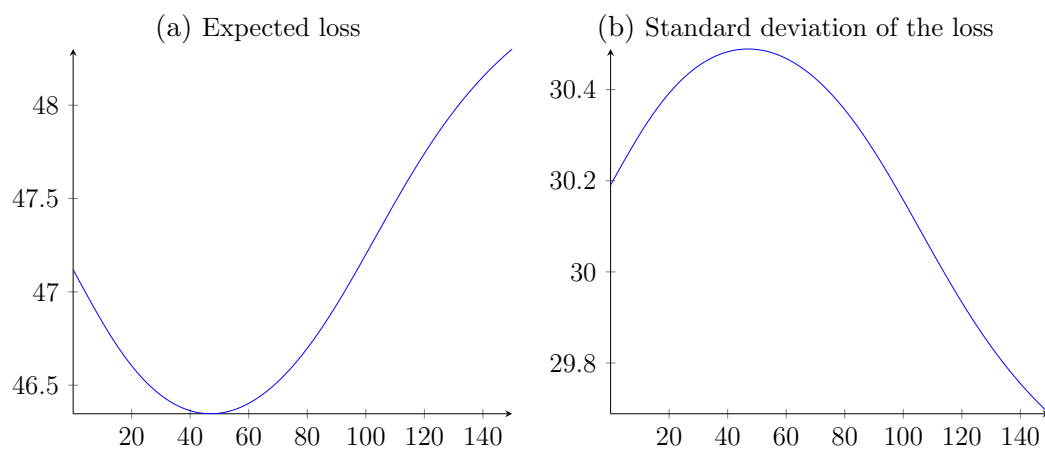


Figure 5: Expected value and standard deviation of the loss as functions of the loading factor.

is not possible in particular situations. Hence, for the sake of completeness, we now present conditions that make Ehrlich and Becker (1972)'s substitutability result between market and self-insurance valid in a continuum-of-loss framework. As shown in Proposition 5, this is the case if self-insurance is locally strong in the upper tier of losses.

Proposition 5 *Assume that $\mathcal{C}(a) \equiv \{x|F_a(x|a) - f(x|a) \geq 0\} = [\tilde{x}(a), \bar{x}]$ with $\tilde{x}'(a) \geq 0$. Then, if $D^*(\sigma)$ is an increasing function and if the individual displays constant absolute risk aversion, market insurance and self-insurance are substitutes for all $\sigma \in [0, \bar{\sigma}]$.*

Under the conditions of Proposition 5, the set of loss levels for which self-insurance is locally strong, $\mathcal{C}(a)$, is an interval that includes the largest losses for all a , i.e. losses close to \bar{x} , contrary to Propositions 1 and 4. The condition $\tilde{x}'(a) \geq 0$ indicates that $\mathcal{C}(a)$ shrinks when a increases, due to a decreasing marginal efficiency of self-insurance. An example of such a case is given by proportional self-insurance, i.e. when loss x is the realization of the random variable $X(a) = X_0\gamma(a)$, with $\gamma(0) = 1$ and $\gamma'(a) < 0$, where X_0 is the random variable of the loss without any self-insurance (hence with c.d.f. $F(x|0)$) and $\gamma(a)$ is a loss reduction factor that is the same for all potential losses that would occur if there were no self-insurance.¹⁴ We then have $F(x|a) \equiv \Pr\{X(a) \leq x\} = \Pr\{X_0 \leq x/\gamma(a)\} = F(x/\gamma(a)|0)$ for all $x \in (0, \bar{x}\gamma(a)]$, which gives $\tilde{x}(a) = -\gamma(a)/\gamma'(a)$ with $\tilde{x}'(a) \geq 0$ if $\gamma''(a) \geq \gamma'(a)^2/\gamma(a)$.¹⁵

3.5 Self-insurance as a residual loss function

We have assumed that the random loss x is distributed in $[0, \bar{x}]$ according to c.d.f. $F(x|a)$ for all $a \geq 0$. Interestingly, the effect of self-insurance on the individual's risk exposure may also be characterized by a function $L(\cdot|a)$ from $[0, \bar{x}]$ to $[0, \bar{x}]$ called *residual loss function* (RLF) such that $L(x|a) \leq x$ and $L(x|0) = x$. By definition of the RLF, the random loss under self-insurance expenses a has the same probability distribution as $L(X_0|a)$ where X_0 is the random variable that corresponds to the no-self-insurance case, i.e. $a = 0$. In other words, function $L(\cdot|a)$ is defined by

$$F(L(x|a)|a) = F(x|0) \text{ for all } x \text{ in } [0, \bar{x}],$$

¹⁴This is the case considered by Quaas and Baumgärtner (2008) in a model with mean-variance preferences.

¹⁵Under proportional self-insurance, the upper bound of possible losses depends on self-insurance expenses. The proof of Proposition 5 covers this case.

with $L_x > 0, L_a \leq 0$.¹⁶ Fig. 6 illustrates this definition in the pure self-insurance case (i.e., $\bar{F}'(a) = 0$ for all a), and for simplicity, we restrict ourselves to this case in what follows. The black curve depicts the c.d.f. of the loss without self-insurance $F(x|0)$, and the blue curve the distribution of the loss with $a > 0$, $F(x|a)$. There is the same probability that the loss is lower than x_0 in the absence of self-insurance and lower than $L(x_0|a)$ under self-insurance at level $a > 0$.¹⁷

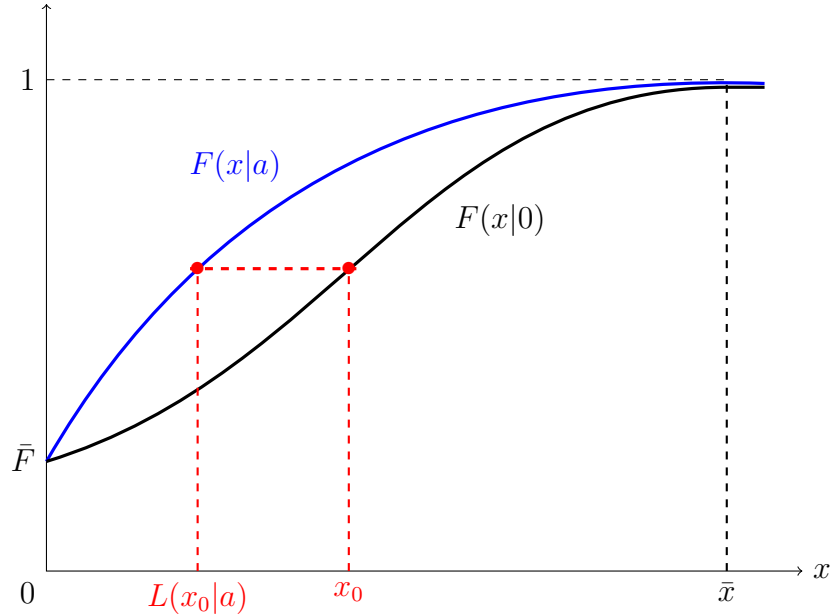


Figure 6: Derivation of the residual loss function.

Proposition 1 has shown that self-insurance and market insurance are complementary when σ is large, under the assumption $F_a(x|a) - f(x|a) < 0$ when x is close to \bar{x} , a case that we have called “locally weak self-insurance”. Furthermore, Proposition 4 has extended this result under the assumption that $F_a(x|a) - f(x|a)$ is positive when x is in an interval $\mathcal{C}(a)$ whose upper and lower bounds are lower than \bar{x} and possibly larger than 0, respectively. The following proposition highlights the close link between this characterization and the RLF.

Proposition 6 *For all (x, a) , we have $F_a(x|a) - f(x|a) < 0$ iff $L_a(L^{-1}(x|a)|a) > -1$.*

¹⁶Proportional self-insurance corresponds to the case where $L(x|a) = \gamma(a)x$, with $\gamma(a) \in (0, 1)$, $\gamma'(a) < 0$.

¹⁷The definition of RLF does not require that the random loss under self-insurance is a deterministic function of the random loss that would occur in the absence of self-insurance. For the sake of clarity, consider Ω the underlying set of states of nature (associated with a set of events and a probability measure) and let $X_a : \Omega \rightarrow [0, \bar{x}]$ be the random variable such that $X_a(\omega)$ is the loss in state $\omega \in \Omega$ under self-insurance expenses a . According to the definition of RLF, X_a and $L(X_0|a)$ are identically distributed in $[0, \bar{x}]$, which does not imply that X_a is a deterministic function of X_0 .

Using Proposition 6 allows us to reformulate Proposition 1 in an intuitive way. Since $L^{-1}(x|a)$ goes to \bar{x} when x goes to \bar{x} , the condition in Proposition 1 means that a marginal increase in self-insurance expenses da is equivalent to a residual loss decrease $-L_a(x|a)da$ smaller than da . This is why self-insurance was said to be locally weak in that case, although Proposition 6 now shows that “ex post inefficient” would have been a better terminology. Hence, the condition in Proposition 1 means that increasing self-insurance would be ex post inefficient if only the largest losses were considered. The condition $\mathcal{C}(a') \subseteq \mathcal{C}(a)$ if $a' > a$ in Proposition 4 may also be reinterpreted in the same perspective: when self-insurance expenses increase, there is a widening of the lowest and highest parts of interval $[0, \bar{x}]$ where increasing self-insurance is ex post inefficient, which corresponds to a decreasing marginal efficiency of self-insurance.

A possible interpretation of condition $L_a(L^{-1}(x|a)|a) > -1$ is in terms of regret the individual may feel when he suffers a loss x after spending a on self-insurance. Indeed, assume that the individual’s rationality reflects an underlying causal model in which he faces an exogenous loss (i.e. independent of his decisions), distributed according to c.d.f. $F(\cdot|0)$, which can be reduced through self-insurance, the final loss being distributed according to c.d.f. $F(\cdot|a)$. The condition $L_a(L^{-1}(x|a)|a) > -1$ then means that after suffering a loss x and realizing that this loss would have been $L^{-1}(x|a)$ without self-insurance, this individual regrets in retrospect having spent too much money on self-insurance.¹⁸

4 Beyond the CARA case

When absolute risk aversion is not constant, the optimal choices about market insurance and self insurance are affected by wealth effects that have been investigated for each instrument separately in the literature. In the following, we untangle their interplay within our framework. In section 4.1, we highlight sufficient conditions under which the complementarity between market insurance and self-insurance remains valid when the loading factor is high. One key condition is that self-insurance is an inferior good. Section 4.2 describes the conditions on the distribution of the global cost of risk that enable this property to hold for individuals who are risk-averse and prudent. Section 4.3 details the impact of self-insurance on this distribution, which can be decomposed into an expected value effect and a volatility effect, and then returns to the

¹⁸The condition in Proposition 1 then means that he experiences such regret after suffering large losses. Similarly, when loss x is in $\mathcal{C}(a)$ introduced in proposition 4, he does not regret his self-insurance expenditures.

notion of local weakness or strength of self-insurance and its interpretation in terms of regret.

4.1 Complementarity of self-insurance and market insurance

Wealth effects are at work when we extend our problem beyond the CARA case. In particular, a decrease in the loading factor σ under its upper bound $\bar{\sigma}$ creates a negative wealth effect, because of the increase in the insurance premium from 0 to $\hat{P}(D^*(\sigma), a^*(\sigma), \sigma) > 0$. When self-insurance is an inferior good (as market insurance under DARA preferences), this wealth effect reinforces the demand for self-insurance induced by a decrease in σ , which confirms the complementarity between market insurance and self-insurance when σ is large. More formally, we obtain the following result where $A(w_f) \equiv -u''(w_f)/u'(w_f)$ denotes the index of absolute risk aversion as a function of the final wealth w_f :

Proposition 7 *If $\lim_{x \rightarrow \bar{x}} \partial \ln f(x|a)/\partial a > -A(w - \bar{x} - a^*(\bar{\sigma}))$ and self-insurance is an inferior good in the absence of market insurance, then self-insurance and market insurance are complementary when σ is close to $\bar{\sigma}$.*

The first assumption in Proposition 7 differs from that in Proposition 1, although both assumptions have in common that an increase in a should not strongly affect the probability distribution of the largest losses.¹⁹ The other assumption stipulates that self-insurance is an inferior good in the absence of market insurance, and we now examine why this may be the case.

4.2 Self-insurance as an inferior good

In the Ehrlich and Becker (1972)'s two-state model without market insurance, Dionne and Eeckhoudt (1985) and Briys and Schlesinger (1990) show that self-insurance is an inferior good when the individual displays decreasing absolute risk aversion (DARA). Lee (2010) shows that this result is not robust in a continuum-of-loss framework and shows that self-insurance can be either an inferior good or a normal good in the DARA case. We detail in the following the effects at work in the continuum-of-loss setup and we discuss the characteristics of loss distributions that make self-insurance an inferior

¹⁹The condition in Proposition 7 is a sufficient one that is more restrictive than the one in Proposition 1. It is compatible with $f_a(x|a) < 0$ when x is close to \bar{x} . In other words, similarly to the condition in Proposition 1, an increase in self-insurance may reduce the probability of the largest losses, but “not too much”.

good for individuals who are risk-averse and prudent (i.e. assuming $u' > 0, u'' < 0$ and $u''' > 0$).

Considering the global cost of risk $z = x + a$, the individual's program in the absence of market insurance is written as

$$\max_{a \geq 0} \int_a^{a+\bar{x}} u(w - z) dG(z|a) \equiv U(w, a),$$

where $G(z|a) = F(z - a|a)$. Define $a^*(w)$ the optimal self-insurance level as a function of initial wealth w in this problem without market insurance, with $\partial U(w, a^*)/\partial a = 0$ at an interior solution $a^*(w) > 0$. The implicit function theorem yields

$$\frac{da^*}{dw} = - \frac{\partial^2 U(w, a)/\partial w \partial a}{\partial^2 U(w, a)/\partial a^2}$$

where $a = a^*(w)$ and $\partial^2 U(w, a)/\partial a^2 < 0$. Hence da^*/dw and $\partial^2 U(w, a)/\partial w \partial a$ have the same sign, and self-insurance is an inferior good in the absence of market insurance if $\partial^2 U(w, a)/\partial w \partial a < 0$. We have

$$\frac{\partial^2 U(w, a)}{\partial w \partial a} = \frac{\partial}{\partial a} \left[\int_a^{a+\bar{x}} u'(w - z) dG(z|a) \right] = \frac{\partial}{\partial a} \left[\int_a^{a+\bar{x}} v(z - w) dG(z|a) \right]$$

where $v(y) \equiv u'(-y)$, and thus $v'(y) = -u''(-y) > 0$ and $v''(y) = u'''(-y)$. Consequently, if $u''' > 0$ (i.e., the individual is prudent), then $\partial^2 U(w, a)/\partial w \partial a < 0$ if an increase in a *reduces* the expected utility of an individual with wealth z distributed according to c.d.f. $G(z|a)$ and with utility function v such that $v' > 0$ and $v'' > 0$ (i.e. a risk-lover individual).

As detailed in the following section, self-insurance produces a change in the distribution of the loss that can be described as the sum of a change in the expected value of the global cost of risk and a mean-preserving change in this global cost (hereafter called the mean-value effect and the volatility effect, respectively). The mean-value effect is detrimental to the individual with utility $v(z - w)$ if the expected value of his final wealth decreases with a because $v' > 0$. As regard to the volatility effect, it is also detrimental to this risk-loving individual if the increase in a reduces the risk of his wealth (in the sense of Rothschild and Stiglitz, 1970). Self-insurance is an inferior good under the cumulated actions of these two effects as shown in the following section.

4.3 Self-insurance and the distribution of risk

Consider two self-insurance levels a_0 and $a_1 > a_0$, and let Z_i be the random variable corresponding to the global cost z when $a = a_i$, for $i = 0$ and 1 , with c.d.f. $G_i(z) = G(z|a_i)$ and expected value $m_i = \mathbb{E}Z_i$. The expected value effect associated with the change from Z_0 to Z_1 is the increase in expected cost given by

$$m_1 - m_0 = a_1 - a_0 - (\mathbb{E}X_0 - \mathbb{E}X_1),$$

where X_i is the random loss under self-insurance a_i , for $i \in \{0, 1\}$. When $m_1 > m_0$, by analogy with market insurance, this expected value effect may be interpreted as the *implicit loading of self-insurance*. In other words, this is the decrease in expected wealth that results from additional self-insurance expenses.

Let us define random variable $Z_2 = Z_0 + m_1 - m_0$ and G_2 and $m_2 = \mathbb{E}Z_2$, its c.d.f. and expected value. Hence, Z_2 is the residual global cost of risk when $a = a_0$ after compensating for the decrease in expected wealth when a changes from a_0 to a_1 . By construction, $m_2 = m_1$. The change from Z_0 to Z_1 when a increases from a_0 to a_1 may then be decomposed into two steps as follows:

$$Z_0 \xrightarrow[\text{step 1}]{+m_1-m_0} Z_2 \xrightarrow[\text{step 2}]{+Z_1-Z_2} Z_1,$$

where steps 1 and 2 correspond to the expected value effect and to the volatility effect, respectively.

Fig. 7 illustrates this two-step decomposition when $m_0 < m_1$ (i.e., when there is a positive implicit loading of self-insurance). $G_2(z)$ corresponds to a rightward shift of $G_0(z)$. Fig. 7 is drawn in the case where Z_1 is a mean-preserving contraction of Z_2 , with only one crossing point between G_1 and G_2 . In this case, the expected value effect and the volatility effect cumulate, and self-insurance is an inferior good.

In Fig. 7, $G_1(z) > G_0(z)$ if $z \in (\hat{z}_0, \hat{z}_1)$.²⁰ We thus have $F(z - a_1|a_1) > F(z - a_0|a_0)$ for all $z \in (\hat{z}_0, \hat{z}_1)$, or equivalently, using the RLF and defining $\Delta a \equiv a_1 - a_0$, $\hat{x}_0 \equiv \hat{z}_0 - a_1$ and $\hat{x}_1 \equiv \hat{z}_1 - a_1$,

$$F(x|a_1) = F(L^{-1}(x|a_1)|0) > F(x + \Delta a|a_0) = F(L^{-1}(x + \Delta a|a_0)|0)$$

and thus $L^{-1}(x|a_1) > L^{-1}(x + \Delta a|a_0)$ implying $L(L^{-1}(x|a_1)|a_0) + a_0 > x + a_1$ for

²⁰If $G_1(z) < G_0(z)$ for all z , then a_0 would be obviously a better choice than a_1 whatever the individual's preferences.

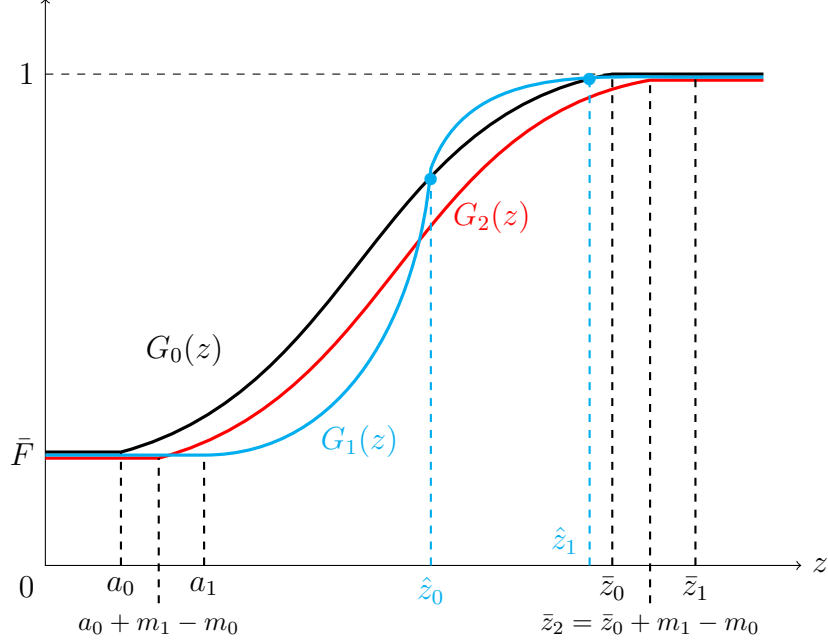


Figure 7: Two-step decomposition of an increase in a from a_0 to a_1 . Curve $G_2(z)$ depicts the c.d.f. of Z_2 and corresponds to a rightward shift of the c.d.f. of Z_0 , $G_0(z)$. Curve $G_1(z)$ depicts the c.d.f. of Z_1 which is a mean-preserving contraction of Z_2 .

all $x \in (\hat{x}_0, \hat{x}_1)$. Following the interpretation in terms of regret introduced in Section 3.5, the individual who suffers a loss $x \in (\hat{x}_0, \hat{x}_1)$ after spending a_1 on self-insurance, considers that he would have suffered a loss larger than $x + a_1 - a_0$ with only a_0 spent on self-insurance, and thus doesn't regret in retrospect having spent a_1 rather than a_0 .

5 Conclusion

When the price of market insurance is high because of transaction costs, taxes or solvency capital costs, subsidizing insurance may be a matter of governmental decision in order to improve risk-sharing in the economy. The resulting change in insurance price is likely to affect self-insurance, and this may be a concern for public decision-making. This is particularly the case in agriculture, when self-insurance creates detrimental externalities, for instance when farmers use more pesticides to protect their crops against pests attacks or when they replace rain-fed farming by irrigation and water resources are scarce. Conversely, externalities would be favorable when self-insurance takes the form of biodiversity preservation and ecosystem protection. When such important matters are at stake, understanding how self-insurance choices react to changes in the insurance price becomes particularly important.

In their seminal paper, Ehrlich and Becker (1972) conclude that self-insurance and market insurance are substitutes, and consequently, subsidizing market insurance would tend to discourage policyholders from exerting self-insurance efforts. In the case of agriculture, this means that subsidizing crop insurance would encourage farmers to deviate from less profitable practices based on biodiversity preservation and ecosystem protection, insofar as these technologies play a self-insurance role. In other words, this suggests that favoring farmers' access to market insurance could undermine the consideration of environmental issues in public policy-making.

As shown in this paper, this would be an over-hasty conclusion. When the cost of accidents is random, contrary to the model of Ehrlich and Becker (1972), then the combination of market insurance and self-insurance goes through the design of the optimal insurance policy. We have considered a model in which this optimal policy corresponds to a straight-deductible contract, which opens the way to an adequate combination of market insurance and self-insurance. Beyond the mere fact that self-insurance shifts the probability distribution of losses leftwards, it has appeared that the substitutability or complementarity between self-insurance and market insurance depends on which levels of loss are less likely to occur when self-insurance expenses increase. When the largest gains from self-insurance correspond to the upper tier of losses, as in the case of proportional self-insurance, then self-insurance exerts risk coverage effects similar to those of an insurance contract covering losses above a deductible, and it is therefore substitutable to market insurance in that case. On the contrary, when self-insurance is more effective in decreasing the probability of medium-sized losses and turning them into small losses, then it is a natural complement to market insurance. In that case, when the high price of insurance leads to a large deductible, reducing the deductible by lowering the loading factor motivates the policyholder to self-insure more intensely. More insurance then goes hand in hand with more self-insurance.

References

- Arrow, K. (1971). *Essays in the Theory of Risk Bearing*. North-Holland, Amsterdam.
- Boomhower, J., Fowlie, M., and Plantinga, A. J. (2023). Wildfire insurance, information, and self-protection. *AEA Papers and Proceedings*, 113:310–15.
- Briys, E. and Schlesinger, H. (1990). Risk aversion and the propensities for self-insurance and self-protection. *Southern Economic Journal*, pages 458–467.
- Briys, E., Schlesinger, H., and Schulenburg, J.-M. G. v. (1991). Reliability of risk management: market insurance, self-insurance and self-protection reconsidered. *The Geneva Papers on Risk and Insurance Theory*, 16:45–58.
- Chakir, R. and Hardelin, J. (2014). Crop insurance and pesticide use in french agriculture: an empirical analysis. *Revue d'Études en Agriculture et Environnement*, 95(01):25–50.
- Cottingham, K., Brown, B., and Lennon, J. (2001). Biodiversity may regulate the temporal variability of ecological systems. *Ecology Letters*, 4(1):72–85.
- Dionne, G. and Eeckhoudt, L. (1985). Self-insurance, self-protection and increased risk aversion. *Economics Letters*, 17(1-2):39–42.
- Ehrlich, I. and Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. *Journal of political Economy*, 80(4):623–648.
- Enjolras, G. and Aubert, M. (2020). How does crop insurance influence pesticide use? evidence from french farms. *Review of Agricultural, Food and Environmental Studies*, 101(4):461–485.
- Gao (2023). Farm bill: Reducing crop insurance costs could fund other priorities. US Government Accountability Office.
- Hartwig, R., Niehaus, G., and Qiu, J. (2020). Insurance for economic losses caused by pandemics. *The Geneva Risk and Insurance Review*, 45:134–170.
- Hiebert, L. D. (1989). Optimal loss reduction and increases in risk aversion. *Journal of Risk and Insurance*, pages 300–305.
- Horowitz, J. and Lichtenberg, E. (1993). Insurance, moral hazard, and chemical use in agriculture. *American Journal of Agricultural Economics*, 75(4):926–935.
- Lee, K. (2010). Wealth effects on self-insurance. *The Geneva Risk and Insurance Review*, 35(2):160–171.
- Mahul, O. (1997). Self-insurance in the presence of background risk. In Huirne, R.B.M., J. H. and Dijkhuizen, A., editors, *Risk Management Strategies in Agriculture: State*

- of the Art and Future Perspectives*, chapter 11, pages 139–150. Backhuys Publishers.
- Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic theory*. Oxford University Press New York.
- Quaas, M. F. and Baumgärtner, S. (2008). Natural vs. financial insurance in the management of public-good ecosystems. *Ecological Economics*, 65(2):397–406.
- Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3):225–243.
- Shanafelt, D. W., Dieckmann, U., Jonas, M., Franklin, O., Loreau, M., and Perrings, C. (2015). Biodiversity, productivity, and the spatial insurance hypothesis revisited. *Journal of Theoretical Biology*, 380:426–435.
- Smith, V. and Goodwin, B. (1996). Crop insurance, moral hazard, and agricultural chemical use. *American Journal of Agricultural Economics*, 78(2):428–438.
- Weber, J., Key, N., and O’donoghue, E. (2016). Does federal crop insurance make environmental externalities from agriculture worse? *Journal of the Association of Environmental and Resource Economists*, 3(3):707–742.
- Wu, J. (1999). Crop insurance, acreage decisions, and nonpoint-source pollution. *American Journal of Agricultural Economics*, 81(2):305–320.
- Yachi, S. and Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proceedings of the National Academy of Sciences*, 96(4):1463–1468.

Appendix

A Lemma 1

Lemma 1 *The optimal solution to the individual's problem a^* , D^* satisfies*

$$\int_0^D u'(w - x - \hat{P}(D, a, \sigma) - a) dF(x|a) = H(D, a, \sigma) u'(w - D - \hat{P}(D, a, \sigma) - a) \quad (4)$$

$$\int_{0+}^D u'(w - x - \hat{P}(D, a, \sigma) - a) F_a(x|a) dx = K(D, a, \sigma) u'(w - D - \hat{P}(D, a, \sigma) - a). \quad (5)$$

where

$$K(D, a, \sigma) \equiv \frac{1}{1 + \sigma} - \int_D^{\bar{x}} F_a(x|a) dx, \quad (6)$$

$$H(D, a, \sigma) \equiv \frac{1}{1 + \sigma} - [1 - F(D|a)]. \quad (7)$$

Proof

Maximizing $\mathbb{E}u$ given by (1) w.r.t. D and a leads to the following first-order optimality conditions

$$-[1 - F(D|a)]u'(w - D - \hat{P} - a) - \frac{\partial \hat{P}}{\partial D} \frac{\partial \mathbb{E}u}{\partial w} = 0 \quad (8)$$

and

$$\begin{aligned} & \bar{F}'(a)u(w - \hat{P}(D, a, \sigma) - a) + \int_{0+}^D u(w - x - \hat{P} - a) f_a(x|a) dx \\ & - F_a(D|a)u(w - D - \hat{P} - a) - \left(1 + \frac{\partial \hat{P}}{\partial a}\right) \frac{\partial \mathbb{E}u}{\partial w} = 0 \end{aligned} \quad (9)$$

respectively, where

$$\begin{aligned} \frac{\partial \mathbb{E}u}{\partial w} = & \bar{F}(a)u'(w - \hat{P}(D, a, \sigma) - a) + \int_{0+}^D u'(w - x - \hat{P} - a) dF(x|a) \\ & + [1 - F(D|a)]u'(w - D - \hat{P} - a), \end{aligned} \quad (10)$$

$$\frac{\partial \hat{P}(D, a, \sigma)}{\partial D} = -(1 + \sigma)[1 - F(D|a)]. \quad (11)$$

Integrating by parts and observing that $F(\bar{x}|a) = 1$ for all a gives $F_a(\bar{x}|a) = 0$, it comes

$$\frac{\partial \hat{P}(D, a, \sigma)}{\partial a} = (1 + \sigma) \int_D^{\bar{x}} (x - D) f_a(x|a) dx = -(1 + \sigma) \int_D^{\bar{x}} F_a(x|a) dx. \quad (12)$$

Using (8), (10) and (11) gives

$$(1 + \sigma) \int_0^D u'(w - x - \hat{P} - a) dF(x|a) = u'(w - D - \hat{P} - a) (1 - (1 + \sigma)[1 - F(D|a)]). \quad (13)$$

Integrating by parts the integral in (9) gives

$$\begin{aligned} \int_{0_+}^D u(w - x - \hat{P} - a) f_a(x|a) dx &= u(w - D - \hat{P} - a) F_a(D|a) - u(w - \hat{P} - a) \bar{F}'(a) \\ &\quad + \int_{0_+}^D u'(w - x - \hat{P} - a) F_a(x|a) dx. \end{aligned}$$

Substituting in (9) and using (12) yields

$$\int_{0_+}^D u'(w - x - \hat{P} - a) F_a(x|a) dx = \left[1 - (1 + \sigma) \int_D^{\bar{x}} F_a(x|a) dx \right] \frac{\partial \mathbb{E}u}{\partial w}. \quad (14)$$

Furthermore, (8) and (11) give

$$\frac{\partial \mathbb{E}u}{\partial w} = \frac{u'(w - D - \hat{P} - a)}{1 + \sigma}, \quad (15)$$

and (14) and (15) yield

$$(1 + \sigma) \int_{0_+}^D u'(w - x - \hat{P} - a) F_a(x|a) dx = u'(w - D - \hat{P} - a) \left[1 - (1 + \sigma) \int_D^{\bar{x}} F_a(x|a) dx \right]. \quad (16)$$

B Proof of Proposition 1

In the CARA case, we may write $u(w) = -e^{-\alpha w}$ where $\alpha > 0$ is the index of absolute risk aversion. Using $u'(w) = -\alpha u(w)$ and factorizing out $u(w - D - \hat{P}(D, a, \sigma) - a)$,

equations (4) and (5) can be rewritten as

$$\int_0^D e^{-\alpha(D-x)} dF(x|a) = H(D, a, \sigma), \quad (17)$$

$$\int_0^D e^{-\alpha(D-x)} F_a(x|a) dx = K(D, a, \sigma), \quad (18)$$

respectively. Totally differentiating these equations, using $H_D = f(D|a)$, $H_a = F_a(D|a) = K_D$, $K_a = -\int_D^{\bar{x}} F_{aa}(x|a) dx$, $H_\sigma = K_\sigma = 1/(1 + \sigma)^2$, and collecting terms give

$$\begin{aligned} \alpha H \frac{dD}{d\sigma} - \left[\int_0^D e^{-\alpha(D-x)} f_a(x|a) dx - F_a(D|a) \right] \frac{da}{d\sigma} &= \frac{1}{(1 + \sigma)^2}, \\ \alpha K \frac{dD}{d\sigma} - \left[\int_0^D e^{-\alpha(D-x)} F_{aa}(x|a) dx + \int_D^{\bar{x}} F_{aa}(x|a) dx \right] \frac{da}{d\sigma} &= \frac{1}{(1 + \sigma)^2}, \end{aligned}$$

which can be written as

$$\begin{bmatrix} \alpha H & -A \\ \alpha K & -B \end{bmatrix} \begin{bmatrix} dD/d\sigma \\ da/d\sigma \end{bmatrix} = \frac{1}{(1 + \sigma)^2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (19)$$

where

$$\begin{aligned} A &= \int_0^D e^{-\alpha(D-x)} f_a(x|a) dx - F_a(D|a), \\ B &= \int_0^D e^{-\alpha(D-x)} F_{aa}(x|a) dx + \int_D^{\bar{x}} F_{aa}(x|a) dx. \end{aligned}$$

We have

$$\begin{vmatrix} \alpha H & -A \\ \alpha K & -B \end{vmatrix} = \alpha(AK - HB) = \alpha\Delta$$

where

$$\Delta \equiv K \left[\int_0^D e^{-\alpha(D-x)} f_a(x|a) dx - F_a(D|a) \right] - H \left[\int_0^D e^{-\alpha(D-x)} F_{aa}(x|a) dx + \int_D^{\bar{x}} F_{aa}(x|a) dx \right] \quad (20)$$

is the determinant of the Hessian matrix of $\mathbb{E}u$ evaluated at a maximum of this function

(i.e. at $D = D^*(\sigma)$ and $a = a^*(\sigma)$). We thus have $\Delta > 0$ and solving (19) yields

$$\begin{aligned} \begin{bmatrix} dD/d\sigma \\ da/d\sigma \end{bmatrix} &= \frac{1}{\alpha\Delta(1+\sigma)^2} \begin{bmatrix} -B & A \\ -\alpha K & \alpha H \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{\alpha\Delta(1+\sigma)^2} \begin{bmatrix} A - B \\ \alpha(H - K) \end{bmatrix}, \end{aligned}$$

hence

$$\frac{dD^*}{d\sigma} = \frac{\int_0^D e^{-\alpha(D-x)} f_a(x|a) dx - F_a(D|a) - \left[\int_0^D e^{-\alpha(D-x)} F_{aa}(x|a) dx + \int_D^{\bar{x}} F_{aa}(x|a) dx \right]}{\alpha\Delta(1+\sigma)^2} \quad (21)$$

and

$$\frac{da^*}{d\sigma} = \frac{\int_0^D e^{-\alpha(D-x)} [f(x|a) - F_a(x|a)] dx}{(1+\sigma)^2\Delta} = \frac{\int_D^{\bar{x}} [F_a(x|a) - f(x|a)] dx}{(1+\sigma)^2\Delta} \quad (22)$$

in which $a = a^*(\sigma)$ and $D = D^*(\sigma)$. The first expression of $da^*/d\sigma$ in (22) is obtained by replacing H and K by the optimal values given by (17) and (18) respectively, while the second expression is obtained replacing H and K by the definitions of functions $H(\cdot)$ and $K(\cdot)$ given by (7) and (6) respectively. From this latter expression, and since we have $D^*(\sigma) > \hat{x}(a)$ when σ is large enough which implies $F_a(x|a) < f(x|a)$ for all $x \in [D, \bar{x}]$, it comes

$$\int_{D^*(\sigma)}^{\bar{x}} [F_a(x|a^*(\sigma)) - f(x|a^*(\sigma))] dx < 0$$

and thus $da^*/d\sigma < 0$ when $\sigma < \bar{\sigma}$, σ close to $\bar{\sigma}$. We also have $dD^*/d\sigma > 0$ when σ is close to $\bar{\sigma}$ since $D^*(\bar{\sigma}) = \bar{x}$ and $D^*(\sigma) < \bar{x}$ when $\sigma < \bar{\sigma}$.

C Proof of Proposition 2

We have $D^*(\sigma) > 0$ when σ is close to 0 since $D^*(0) = 0$ and $D^*(\sigma) > 0$ when $\sigma > 0$. We may write $F(x|a) = \bar{F}(a) + [1 - \bar{F}(a)]H(x|a)$, where $H(x|a)$ is the c.d.f. of the loss conditionally on a loss occurring, with $H(0|a) = 0$ and $H(\bar{x}|a) = 1$. Hence, we have $F_a(x|a) = \bar{F}'(a)[1 - H(x|a)] + [1 - \bar{F}(a)]H_a(x|a)$ and $f(x|a) = [1 - \bar{F}(a)]h(x|a)$

where $h(x|a) = H_x(x|a)$. We may write

$$\begin{aligned} \int_{0+}^{\bar{x}} [F_a(x|a) - f(x|a)] dx &= \bar{F}'(a) \int_0^{\bar{x}} [1 - H(x|a)] dx \\ &+ [1 - \bar{F}(a)] \int_0^{\bar{x}} H_a(x|a) dx - [1 - \bar{F}(a)], \end{aligned} \quad (23)$$

using $\int_0^{\bar{x}} h(x|a) dx = 1$, and

$$\begin{aligned} a^*(0) &\in \arg \min_a [1 - \bar{F}(a)] \int_0^{\bar{x}} x h(x|a) dx + a \\ &= \arg \min_a [1 - \bar{F}(a)] \int_0^{\bar{x}} [1 - H(x|a)] dx + a. \end{aligned}$$

The corresponding FOC gives

$$-\bar{F}'(a^*(0)) \int_0^{\bar{x}} [1 - H(x|a^*(0))] dx - [1 - \bar{F}(a^*(0))] \int_0^{\bar{x}} H_a(x|a^*(0)) dx + 1 = 0. \quad (24)$$

When $\sigma \rightarrow 0$, $D^*(\sigma) \rightarrow 0$ and thus, using (23) and (24),

$$\int_{D^*(\sigma)}^{\bar{x}} [F_a(x|a^*(\sigma)) - f(x, a^*(\sigma))] dx \rightarrow \int_{0+}^{\bar{x}} [F_a(x|a^*(0)) - f(x|a^*(0))] dx = \bar{F}(a^*(0)).$$

When $\bar{F}(a^*(0)) > 0$, using (22) and $\Delta > 0$ then gives $a^{*\prime}(\sigma) > 0$ when σ is close to 0.

When $\bar{F}(a) = 0$ for all a , we also obtain $a^{*\prime}(\sigma) = 0$ when $\sigma \rightarrow 0$. Indeed, suppose the contrary, i.e. $a^{*\prime}(\sigma) \leq 0$ when $\sigma \rightarrow 0$, $\sigma > 0$. Let $\psi^*(\sigma) \equiv \psi(D^*(\sigma), a^*(\sigma))$ where

$$\psi(D, a) \equiv \int_D^{\bar{x}} [F_a(x|a) - f(x|a)] dx. \quad (25)$$

We know from (22) in the proof of Proposition 1 that $a^{*\prime}(\sigma)$ and $\psi^*(\sigma)$ have the same sign. We have $\psi^*(0) = \bar{F}(a^*(0)) = 0$ and

$$\begin{aligned} \psi^{*\prime}(\sigma) &= -[F_a(D^*(\sigma)|a^*(\sigma)) - f(D^*(\sigma)|a^*(\sigma))] D^{*\prime}(\sigma) \\ &+ a^{*\prime}(\sigma) \int_{D^*(\sigma)}^{\bar{x}} [F_{aa}(x|a^*(\sigma)) - f_a(x|a^*(\sigma))] dx. \end{aligned} \quad (26)$$

$\bar{F}(a) = F(0|a) = 0$ for all a gives $F_a(0|a) = 0$ implying $F_a(x|a) \rightarrow 0$ when $x \rightarrow 0$ since $x \mapsto F_a(x|a)$ is right-continuous. Also, as $F(\bar{x}|a) = 1$ for all a , we have $F_a(\bar{x}|a) = 0$ for

all a and thus $\int_0^{\bar{x}} f_a(x|a)dx = F_a(\bar{x}|a) - F_a(0|a) = 0$. Consequently

$$\int_{D^*(\sigma)}^{\bar{x}} [F_{aa}(x|a^*(\sigma)) - f_a(x|a^*(\sigma))]dx \rightarrow \int_0^{\bar{x}} F_{aa}(x|a^*(0))dx \leq 0$$

when $\sigma \rightarrow 0$ since $F_{aa}(x|a) \leq 0$. As $a^{*\prime}(\sigma) \leq 0$ when $\sigma \rightarrow 0$, the second term of (26) is thus non-negative. Furthermore, $D^{*\prime}(\sigma) \rightarrow +\infty$ when $\sigma \rightarrow 0$ because the numerator in (21) goes to a positive limit whereas the denominator goes to zero. We thus have

$$-[F_a(D^*(\sigma)|a^*(\sigma)) - f(D^*(\sigma)|a^*(\sigma))]D^{*\prime}(\sigma) \rightarrow f(0|a^*(0))D^{*\prime}(0)$$

when $\sigma \rightarrow 0$, and thus $\psi^{*\prime}(\sigma) \rightarrow +\infty$ when $\sigma \rightarrow 0$. Using $\psi^*(0) = 0$ yields $\psi^*(\sigma) > 0$ when $\sigma > 0$, σ close to 0, hence $a^{*\prime}(\sigma) > 0$ when $\sigma \rightarrow 0$, $\sigma > 0$, a contradiction.

D Proof of Proposition 3

(21) can be written as

$$\frac{dD^*}{d\sigma} = \frac{A(D, a) - B(D, a)}{\alpha\Delta(1 + \sigma)^2}$$

where $D = D^*(\sigma)$, $a = a^*(\sigma)$, and

$$A(D, a) \equiv \int_0^D e^{-\alpha(D-x)} [f_a(x|a) - F_{aa}(x|a)]dx,$$

$$B(D, a) \equiv F_a(D|a) + \int_D^{\bar{x}} F_{aa}(x|a)dx.$$

Let

$$\phi(x|a) \equiv f_a(x|a) - F_{aa}(x|a). \quad (27)$$

We have $f(x|a) = 0$ for all a when $x \geq \bar{x}$ and thus $f_a(\bar{x}|a) = 0$ because $f(x|a)$ is twice-differentiable at $x = \bar{x}$. Furthermore, we have $F(x|a) = 1$ for all a and $x \geq \bar{x}$, which implies $F_{aa}(\bar{x}|a) = 0$. Hence, we have $\phi(\bar{x}|a) = 0$ for all a . For any x, a such that $\phi(x|a) < 0$, we have $f_a(x|a) < 0$ (because $F_{aa}(x|a) \leq 0$ for all x), and thus $\phi_x(x|a) = f_{ax}(x|a) - f_{aa}(x|a) \leq 0$ under the condition postulated in the Proposition. This implies $\phi(x'|a) < 0$ for all $x' > x$, which contradicts $\phi(\bar{x}|a) = 0$. Hence $\phi(x|a) \geq 0$ for all $x \in (0, \bar{x})$, which gives

$$A(D, a) \equiv \int_0^D e^{-\alpha(D-x)} \phi(x|a)dx \geq 0.$$

Furthermore, we have $B(0, a) = \int_0^{\bar{x}} F_{aa}(x|a)dx < 0$, $B(\bar{x}, a) = 0$ and $\partial B(D, a)/\partial D = f_a(D|a) - F_{aa}(D|a) = \phi(D|a) \geq 0$, which implies $B(D, a) < 0$ for all $D \leq \bar{x}$. We deduce $A(D, a) - B(D, a) > 0$ and thus $dD^*/d\sigma > 0$ for all $\sigma \in (0, \bar{\sigma})$.

E Proof of Proposition 4

We know from (22) and (25) that $a^{*\prime}(\sigma)$ and $\psi^*(\sigma) \equiv \psi(D^*(\sigma), a^*(\sigma))$ have the same sign, and from Propositions 1 and 2 that $\psi^*(0_+) > 0$, $\psi^*(\bar{\sigma}) = 0$ and $\psi^*(\bar{\sigma}_-) < 0$. Let $\hat{\sigma} \equiv \inf\{\sigma : \forall \sigma' \in [\sigma, \bar{\sigma}], \psi^*(\sigma') \leq 0\}$ so that $a^*(\hat{\sigma})$ is a local maximum, and suppose that $a^*(\cdot)$ reaches another local maximum at $\tilde{\sigma} < \hat{\sigma}$. Then, it must also reach a local minimum at $\sigma_0 \in (\tilde{\sigma}, \hat{\sigma})$, with $a^*(\sigma_0) < a^*(\tilde{\sigma})$ and $a^*(\sigma_0) < a^*(\hat{\sigma})$, and we must have $\psi^*(\hat{\sigma}) = \psi^*(\tilde{\sigma}) = \psi^*(\sigma_0) = 0$ with $\psi^{*\prime}(\hat{\sigma}) < 0$, $\psi^{*\prime}(\tilde{\sigma}) < 0$ and $\psi^{*\prime}(\sigma_0) > 0$ where, using (26),

$$\psi^{*\prime}(\sigma) = -[F_a(D^*(\sigma)|a^*(\sigma)) - f(D^*(\sigma)|a^*(\sigma))]D^{*\prime}(\sigma)$$

for $\sigma \in \{\tilde{\sigma}, \sigma_0, \hat{\sigma}\}$. As $D^{*\prime}(\sigma) > 0$, $\psi^{*\prime}(\sigma)$ and $F_a(D^*(\sigma)|a^*(\sigma)) - f(D^*(\sigma)|a^*(\sigma))$ have opposite signs. From (ii), $a^*(\sigma_0) < a^*(\tilde{\sigma})$, $a^*(\sigma_0) < a^*(\hat{\sigma})$, and $\psi^{*\prime}(\hat{\sigma}) < 0$ and $\psi^{*\prime}(\tilde{\sigma}) < 0$, we have $D^*(\hat{\sigma}) \in \mathcal{C}(a^*(\hat{\sigma})) \subseteq \mathcal{C}(a^*(\sigma_0))$ and $D^*(\tilde{\sigma}) \in \mathcal{C}(a^*(\tilde{\sigma})) \subseteq \mathcal{C}(a^*(\sigma_0))$. Also, as $\tilde{\sigma} < \sigma_0 < \hat{\sigma}$ and $D^*(\sigma)$ is non-decreasing, we have $D^*(\tilde{\sigma}) \leq D^*(\sigma_0) \leq D^*(\hat{\sigma})$ and since $\mathcal{C}(a^*(\sigma_0))$ is convex, $D^*(\sigma_0) \in \mathcal{C}(a^*(\sigma_0))$, implying $\psi^{*\prime}(\sigma_0) < 0$, a contradiction.

F Proof of Proposition 5

In this section, we consider the case where the upper limit of the loss may depend on the self-insurance effort, i.e. $\bar{x}(a)$, with $\bar{x}'(a) \leq 0$. We thus have $\mathcal{C}(a) = [\tilde{x}(a), \bar{x}(a)]$, with $\tilde{x}'(a) \geq 0$, $\bar{x}'(a) \leq 0$, and $D^*(\bar{\sigma}) = \bar{x}(a^*(\bar{\sigma}))$ by definition of $\bar{\sigma}$, and $D^*(\sigma) < \bar{x}(a^*(\sigma))$ for all $\sigma < \bar{\sigma}$.

We know from (22) and (25) that $a^{*\prime}(\sigma)$ and $\psi^*(\sigma) \equiv \psi(D^*(\sigma), a^*(\sigma))$ have the same sign, and from Proposition 2 that $\psi^*(0_+) > 0$. Suppose that $a^*(\cdot)$ is not monotonic, i.e. that it reaches a local maximum at $\tilde{\sigma} < \bar{\sigma}$. Then, either it reaches a local minimum at $\sigma_0 \in (\tilde{\sigma}, \bar{\sigma})$, or it is monotonically decreasing over $[\tilde{\sigma}, \bar{\sigma}]$. In the first case, we must have $\psi^*(\tilde{\sigma}) = \psi^*(\sigma_0) = 0$ with $\psi^{*\prime}(\tilde{\sigma}) < 0$ and $\psi^{*\prime}(\sigma_0) > 0$ where, using (26),

$$\psi^{*\prime}(\sigma) = -[F_a(D^*(\sigma)|a^*(\sigma)) - f(D^*(\sigma)|a^*(\sigma))]D^{*\prime}(\sigma)$$

for $\sigma \in \{\tilde{\sigma}, \sigma_0\}$. As $D^{*\prime}(\sigma) > 0$, $\psi^{*\prime}(\sigma)$ and $F_a(D^*(\sigma)|a^*(\sigma)) - f(D^*(\sigma)|a^*(\sigma))$ have

opposite signs. As $\psi^{*\prime}(\tilde{\sigma}) < 0$ and $\psi^{*\prime}(\sigma_0) > 0$, we have $D^*(\tilde{\sigma}) \in \mathcal{C}(a^*(\tilde{\sigma}))$ and $D^*(\sigma_0) \notin \mathcal{C}(a^*(\sigma_0))$, hence $D^*(\tilde{\sigma}) \geq \tilde{x}(a^*(\tilde{\sigma}))$ and $D^*(\sigma_0) < \tilde{x}(a^*(\sigma_0))$. From $\tilde{x}'(a) \geq 0$ and $a^*(\tilde{\sigma}) > a^*(\sigma_0)$, we have $\tilde{x}(a^*(\tilde{\sigma})) \geq \tilde{x}(a^*(\sigma_0))$, and thus $D^*(\tilde{\sigma}) > D^*(\sigma_0)$, which contradicts $\tilde{\sigma} < \sigma_0$ and $D^{*\prime}(\sigma) > 0$.

In the second case, we have $D^*(\tilde{\sigma}) \geq \tilde{x}(a^*(\tilde{\sigma}))$ as in the first case, and using $D^*(\bar{\sigma}) = \bar{x}(a^*(\bar{\sigma}))$, (17), (18) and (20), we get $H(D^*(\bar{\sigma}), a^*(\bar{\sigma}), \bar{\sigma}) = K(D^*(\bar{\sigma}), a^*(\bar{\sigma}), \bar{\sigma}) = 1/(1 + \bar{\sigma})$ and

$$\Delta \rightarrow \left[\int_0^{\bar{x}(a^*(\bar{\sigma}))} e^{-\alpha[\bar{x}(a^*(\bar{\sigma})) - x]} [f_a(x|a^*(\bar{\sigma})) - F_{aa}(x|a^*(\bar{\sigma}))] dx \right] / (1 + \bar{\sigma}) \neq 0$$

when $\sigma \rightarrow \bar{\sigma}$, and thus, from (21) and (22), $D^{*\prime}(\sigma) \rightarrow 1/\alpha(1 + \bar{\sigma}) > 0$ and $a^{*\prime}(\sigma) \rightarrow 0$ when $\sigma \rightarrow \bar{\sigma}$. As $a^*(\cdot)$ is monotonically decreasing over $[\tilde{\sigma}, \bar{\sigma}]$, we have $\psi^*(\sigma) < 0$ for all $\sigma \in (\tilde{\sigma}, \bar{\sigma})$, and from (25) and (26), $\psi^*(\bar{\sigma}) = \psi(\bar{x}(a^*(\bar{\sigma})), a^*(\bar{\sigma})) = 0$ and

$$\psi^{*\prime}(\bar{\sigma}) = -[F_a(\bar{x}(a^*(\bar{\sigma}))|a^*(\bar{\sigma})) - f(\bar{x}(a^*(\bar{\sigma}))|a^*(\bar{\sigma}))]D^{*\prime}(\bar{\sigma}) = 0.$$

Since $\psi^*(\sigma) < 0$ for σ close to $\bar{\sigma}$ and $\psi^{*\prime}(\bar{\sigma}) = \psi^*(\bar{\sigma}) = 0$, we thus must have $\psi^{*\prime\prime}(\bar{\sigma}) < 0$. Furthermore, using (26), $D^*(\bar{\sigma}) = \bar{x}(a^*(\bar{\sigma}))$ and $F_a(\bar{x}(a^*(\bar{\sigma}))|a^*(\bar{\sigma})) - f(\bar{x}(a^*(\bar{\sigma}))|a^*(\bar{\sigma})) = 0$ yields

$$\begin{aligned} \psi^{*\prime\prime}(\bar{\sigma}) &= -D^{*\prime}(\bar{\sigma})^2 \frac{\partial}{\partial x} [F_a(x|a^*(\bar{\sigma})) - f(x|a^*(\bar{\sigma}))]|_{x=\bar{x}(a^*(\bar{\sigma}))} \\ &\quad - D^{*\prime}(\bar{\sigma})a^{*\prime}(\bar{\sigma}) \frac{\partial}{\partial a} [F_a(x|a^*(\bar{\sigma})) - f(x|a^*(\bar{\sigma}))]|_{a=a^*(\bar{\sigma})}. \end{aligned}$$

Using $a^{*\prime}(\bar{\sigma}) = 0$ and $\psi^{*\prime\prime}(\bar{\sigma}) < 0$ gives $\frac{\partial}{\partial x} [F_a(x|a^*(\bar{\sigma})) - f(x|a^*(\bar{\sigma}))]|_{x=\bar{x}(a^*(\bar{\sigma}))} > 0$. As $F_a(D^*(\bar{\sigma})|a^*(\bar{\sigma})) - f(D^*(\bar{\sigma})|a^*(\bar{\sigma})) = 0$, we deduce $F_a(x|a^*(\bar{\sigma})) - f(x|a^*(\bar{\sigma})) < 0$ if $x < D^*(\bar{\sigma})$, and thus $\mathcal{C}(a^*(\bar{\sigma})) = \{D^*(\bar{\sigma})\} = \{\bar{x}(a^*(\bar{\sigma}))\}$, i.e. $\tilde{x}(a^*(\bar{\sigma})) = \bar{x}(a^*(\bar{\sigma}))$. As $D^{*\prime}(\sigma) > 0$, we have $D^*(\tilde{\sigma}) < D^*(\bar{\sigma})$. Furthermore, $\tilde{x}'(a) \geq 0$ and $a^*(\tilde{\sigma}) > a^*(\bar{\sigma})$ gives $\tilde{x}(a^*(\bar{\sigma})) \leq \tilde{x}(a^*(\tilde{\sigma}))$, hence $D^*(\tilde{\sigma}) < \tilde{x}(a^*(\tilde{\sigma}))$, a contradiction.

G Proof of Proposition 6

Using $F(L(x|a), a) = F(x|0)$ and $F_x(x|a) = f(x|a) > 0$ for all $x \in (0, \bar{x})$ shows that $L(x|a)$ is increasing, with $L(x|a) \rightarrow \bar{x}$ when $x \rightarrow \bar{x}$ and $L(x|a) \rightarrow 0$ when $x \rightarrow 0$.

Differentiating $F(x|a) = F(L^{-1}(x|a)|0)$ yields

$$\begin{aligned} F_a(x|a) &= f(L^{-1}(x|a)|0)L_a^{-1}(x|a), \\ f(x|a) &= f(L^{-1}(x|a)|0)L_x^{-1}(x|a). \end{aligned}$$

Differentiating the identity $L^{-1}(L(x|a)|a) = x$ yields

$$\begin{aligned} L_a^{-1}(L(x|a)|a) &= -L_a(x|a)/L_x(x|a), \\ L_x^{-1}(L(x|a)|a) &= -1/L_x(x|a) \end{aligned}$$

that hold for all $x \in [0, \bar{x}]$, implying

$$\begin{aligned} L_a^{-1}(x|a) &= -L_a(L^{-1}(x|a)|a)/L_x(L^{-1}(x|a)|a), \\ L_x^{-1}(x|a) &= -1/L_x(L^{-1}(x|a)|a). \end{aligned}$$

Consequently, we have

$$F_a(x|a) - f(x|a) = -\frac{f(L^{-1}(x|a)|0)}{L_x(L^{-1}(x|a)|a)}[L_a(L^{-1}(x|a)|a) + 1].$$

We have $L^{-1}(x|a) \in [x, \bar{x}]$ and thus $L^{-1}(x|a) \rightarrow \bar{x}$ when $x \rightarrow \bar{x}$. Similarly, $L^{-1}(x|a) \rightarrow 0$ when $x \rightarrow 0$. Consequently, when x is close to 0 or \bar{x} , we have $F_a(x|a) - f(x|a) < 0$ iff $L_a(x|a) > -1$.

H Proof of Proposition 7

Let $P^*(\sigma) \equiv \hat{P}(D^*(\sigma), a^*(\sigma), \sigma)$ and define $\hat{a}(\sigma)$ as the solution to

$$\max_a \int_0^{\bar{x}} u(w - x - P^*(\sigma) - a)dF(x|a).$$

In words, $\hat{a}(\sigma)$ is the optimal self-insurance expense when the individual pays $P^*(\sigma) > 0$ and is not covered by the insurance contract (with $P^*(\sigma) > 0$ if $\sigma < \bar{\sigma}$ and $P^*(\bar{\sigma}) = 0$ since $D^*(\bar{\sigma}) = \bar{x}$). Let $\sigma < \bar{\sigma}$. If self-insurance is an inferior good in the absence of market-insurance, we have $\hat{a}(\sigma) > a^*(\bar{\sigma})$. Hence, we have $a^*(\sigma) > a^*(\bar{\sigma})$ if $a^*(\sigma) > \hat{a}(\sigma)$ with $a^*(\bar{\sigma}) = \hat{a}(\bar{\sigma})$. Let

$$V(D, a, \sigma) = \int_0^D u(w - x - P^*(\sigma) - a)dF(x|a) + [1 - F(D|a)]u(w - D - P^*(\sigma) - a),$$

and let $a^{**}(D, \sigma) = \arg \max_a V(D, a, \sigma)$. We have $a^{**}(\bar{x}, \sigma) = \hat{a}(\sigma)$, and since

$$(\partial \mathbb{E}u / \partial a)|_{D=D^*(\sigma)} - (\partial V / \partial a)|_{D=D^*(\sigma)} = -\hat{P}_a(D^*(\sigma), a, \sigma)(\partial \mathbb{E}u / \partial w)|_{D=D^*(\sigma)} \geq 0$$

when $a = a^*(\sigma)$, we have $a^*(\sigma) > a^{**}(D^*(\sigma), \sigma)$. Hence, a sufficient condition for $a^*(\sigma) > \hat{a}(\sigma) = a^{**}(\bar{x}, \sigma)$ is given by $a^{**}(D^*(\sigma), \sigma) > \hat{a}(\sigma)$. This is the case if function $D \rightarrow a^{**}(D, \sigma)$ is decreasing. The F.O.C for $a^{**}(D, \sigma)$ is

$$\begin{aligned} \frac{\partial V}{\partial a} &= - \int_0^D u'(w - x - P^*(\sigma) - a) dF(x|a) - [1 - F(D|a)]u'(w - D - P^*(\sigma) - a) \\ &\quad + \int_0^D u(w - x - P^*(\sigma) - a) f_a(x|a) dx - F_a(D|a)u(w - D - P^*(\sigma) - a) \\ &= 0, \end{aligned}$$

and using the implicit function theorem yields

$$\frac{\partial a^{**}(D, \sigma)}{\partial D} = - \frac{\partial^2 V(D, a, \sigma) / \partial a \partial D}{\partial^2 V(D, a, \sigma) / \partial a^2} \Big|_{a=a^{**}(D, \sigma)}.$$

From the S.O.C, we have $\partial^2 V / \partial a^2|_{a=a^{**}(D, \sigma)} < 0$, and thus $D \rightarrow a^{**}(D, \sigma)$ is decreasing if $\partial^2 V / \partial a \partial D|_{a=a^{**}(D, \sigma)} < 0$ where

$$\partial^2 V / \partial a \partial D = [1 - F(D|a)]u''(w - D - P^*(\sigma) - a) + F_a(D|a)u'(w - D - P^*(\sigma) - a).$$

We have $\partial^2 V / \partial a \partial D < 0$ if $F_a(D|a) / [1 - F(D|a)] < A(w - D - P^*(\sigma) - a)$. Using l'Hospital rule, this condition holds when $D \rightarrow \bar{x}$ if

$$\lim_{x \rightarrow \bar{x}} - \frac{f_a(x|a)}{f(x|a)} = - \lim_{x \rightarrow \bar{x}} \frac{\partial \ln f(x|a)}{\partial a} < A(w - \bar{x} - a^*(\bar{\sigma})).$$

Under this condition, we have $a^*(\bar{\sigma}) < \hat{a}(\sigma) < a^{**}(D^*(\sigma), \sigma) < a^*(\sigma)$ if $\sigma < \bar{\sigma}$ when self-insurance is an inferior good in the absence of market insurance. In that case, self-insurance and market insurance are complementary when σ is close to $\bar{\sigma}$.