



Munich Personal RePEc Archive

Belief heterogeneity in the Arrow-Borch-Raviv insurance model

Ghossoub, Mario

University of Montreal

10 June 2010

Online at <https://mpra.ub.uni-muenchen.de/37630/>
MPRA Paper No. 37630, posted 28 Mar 2012 16:51 UTC

BELIEF HETEROGENEITY IN THE ARROW-BORCH-RAVIV INSURANCE MODEL

MARIO GHOSSOUB
UNIVERSITÉ DE MONTRÉAL

FIRST DRAFT: JUNE 10, 2010

THIS DRAFT: MARCH 22, 2012

LATEST DRAFT: <https://sites.google.com/site/marioghossoub>

ABSTRACT. In the classical Arrow-Borch-Raviv problem of demand for insurance contracts, it is well-known that the optimal insurance contract for an insurance buyer – or decision maker (DM) – is a *deductible* contract, when the insurer is a risk-neutral Expected-Utility (EU) maximizer, and when the DM is a risk-averse EU-maximizer. In the Arrow-Borch-Raviv framework, however, both parties share the same probabilistic beliefs about the realizations of the underlying insurable loss. This paper argues for heterogeneity of beliefs in the classical insurance model, and considers a setting where the DM and the insurer have preferences yielding different subjective beliefs. The DM seeks the insurance contract that will maximize her (subjective) expected utility of terminal wealth with respect to her subjective probability measure, whereas the insurer sets premiums on the basis of his subjective probability measure. I show that in this setting, and under a consistency requirement on the insurer’s subjective probability that I call *vigilance*, there exists an event to which the DM assigns full (subjective) probability and on which an optimal insurance contract for the DM takes the form of what I will call a *generalized deductible contract*. Moreover, the class of all optimal contracts for the DM that are nondecreasing in the loss is fully characterized in terms of their distribution under the DM’s probability measure. Finally, the assumption of *vigilance* is shown to be a weakening of the assumption of a *monotone likelihood ratio*, when the latter can be defined, and it is hence a useful tool in situations where the likelihood ratio cannot be defined.

Key words and phrases. Optimal insurance, deductible contract, subjective probability, heterogeneous beliefs, vigilance, Agreement Theorem, Harsányi Doctrine, Wilson Doctrine.

JEL Classification: C02, C65, D86, G22.

This paper is a revision of Chapter 3 of my thesis [26], and was previously entitled “*The Demand for Insurance under Heterogeneous Beliefs*”. I am very grateful to Max Amarante and Marc Henry for their advice and insightful comments, and for many helpful discussions. I also thank Carole Bernard, Andrew Heunis, Adam Kolkiewicz, Fabio Maccheroni, Massimo Marinacci, Marcus Pivato, and Yves Sprumont for their comments on earlier drafts. Any errors or omissions are entirely mine. Financial support from the Natural Sciences and Engineering Research Council of Canada, from the Social Sciences and Humanities Research Council of Canada, from the Waterloo Research Institute in Insurance, Securities, and Quantitative Finance, and from the Society of Actuaries is gratefully acknowledged. Part of this research was done while I was visiting the department of Decision Sciences of Università Bocconi, which I thank for its hospitality.

1. INTRODUCTION

The problem of demand for insurance contracts has become part of the folklore of economic theory, as it were, and unequivocally one of the cornerstones of actuarial thought. From the outset, the problem was studied within the framework of Expected-Utility Theory (EUT), as in the seminal work of Arrow [3], Borch [14], and Raviv [45], where it was shown that full insurance above a deductible is optimal for the DM when the premium principle depends on the actuarial value (expected value) of the indemnity, when the decision maker (DM) is a risk-averse Expected-Utility (EU) Maximizer, and when both the DM and the insurer share the same probabilistic beliefs about the realization of a given insurable loss. These foundational results were then extended in several directions, all the while maintaining the assumption of homogeneity of beliefs¹.

This homogeneity of beliefs is a consequence of the objectivity of the underlying uncertainty, in the sense that likelihoods are given independently of the preferences of the parties involved. Indeed, in the classical Arrow-Borch-Raviv approach, the insurable loss is taken to be a (nonnegative) random variable X on a given probability space (Ω, \mathcal{G}, P) , where the probability measure P is independent of the preferences of both the DM and the insurer. This classical formulation of the insurance demand problem is essentially due to Arrow [3], and is a heritage of the von Neumann and Morgenstern (vNM) [55] approach to the definition of uncertainty and decision under uncertainty, where an individual's preferences are over a collection of *objective* lotteries. "Objectivity" is to be understood here as the fact that likelihoods are given exogenously to the individuals' choice problem and independently of their preferences. This framework guarantees *de facto* a perfect homogeneity of beliefs, since probabilities are totally objective and hence common to all parties involved.

Regardless of the way uncertainty is formulated, the case can also be made for an assumption of homogeneity of beliefs in an insurance model relying only on Aumann's [4] celebrated *Agreement Theorem*, within the differential information approach (or information-structure approach) to the formation of beliefs initially introduced by Aumann. The Agreement Theorem roughly states that when two parties share a common prior belief, and when their posterior beliefs² are common knowledge³, then these posterior beliefs must coincide. In this sense, two individuals having the same prior cannot agree to disagree, when the information is completely shared, and hence disagreement about probabilities must vanish.

To see how the Agreement Theorem can be incorporated into the Arrow-Borch-Raviv insurance model, suppose that before engaging in the problem of selecting an insurance contract, that is at a stage prior to the problem under consideration, both the insurer and the DM have the opportunity to formulate prior beliefs about the realizations of the insurable loss. Once they receive some relevant information, both parties proceed to updating their beliefs based on this information, hence yielding their posterior beliefs. They then engage in the insurance activity based on these posterior beliefs. If an epistemic foundation for the theory of insurance demand were to be rooted in the Agreement Theorem, then homogeneity of (posterior) beliefs can be obtained automatically as a consequence of both (i) common prior beliefs about the realization of the insurable loss, and, (ii) complete sharing of relevant information so that posterior beliefs become common knowledge. However, as soon as either (i), or (ii), or both fail to hold, there would be no rationale for the assumption of homogeneous beliefs in the classical insurance model, at least as far as the information-structure approach is concerned.

¹See [23, p. 59] for a brief survey.

²Obtained by Bayesian updating of the priors.

³In Aumann's framework, an event is said to be common knowledge if both individuals know it, if each individual knows that the other individual knows it, if each individual knows that other individuals knows that the former knows it, and so on.

After reviewing the classical Arrow-Borch-Raviv model, I will argue that heterogeneity of beliefs in insurance markets is, on a practical level, a very natural assumption, and, on a theoretical level, an assumption that is as justified as that of belief homogeneity. The argument is based on a criticism of the *Common Priors Assumption* (CPA); a criticism of the the assumption that posteriors are common knowledge, or that information is totally shared; the spirit of the *Wilson Doctrine*; and, a criticism of the objective approach to uncertainty.

I then consider an insurance model in the spirit of the Arrow-Borch-Raviv model, with the exception that the DM and the insurer are allowed to entertain different subjective beliefs about the realization of the insurable loss. I adopt a decision-theoretic approach to belief formation and rely on the heterogeneity of preferences as a proxy for the heterogeneity of (subjective) beliefs. Both agents are Subjective Expected-Utility (SEU) maximizers, with different subjective probability measures. I do not assume that the DM's actions influence the realization of the random loss under consideration. The model considered does not allow for moral hazard or for information asymmetry. Rather, it allows only for *belief asymmetry*, and the rest is identical to the Arrow-Borch-Raviv setup.

I introduce a consistency requirement on the subjective probability measure of the insurer with respect to that of the DM that I call *vigilance*, and I show that if this condition holds, then the DM's demand problem admits a solution \mathcal{Y}^* which is a nondecreasing function of the underlying loss X , and which has the same distribution (for her subjective probability measure) as a function of the form

$$(1.1) \quad Z := \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(\lambda^* h) \right] \right) \right]$$

for some $\lambda^* \geq 0$ and a nonnegative measurable function h which is entirely characterized from the subjective probabilities of both parties (Theorem 6.2 on p. 17). The function u is the DM's utility function, W_0 is the DM's initial level of wealth, and Π is the premium paid by the DM, as in the Arrow-Borch-Raviv model. Moreover, any other indemnity schedule which is a nondecreasing function of the underlying loss and which has the same distribution as \mathcal{Y}^* for the DM is almost surely equal to \mathcal{Y}^* (for the DM's probability measure). The monotonicity of an insurance indemnity schedule is usually desired so as to eliminate *ex post* moral hazard issues that might arise from the DM's possible misreporting of the actual amount of the loss suffered. Furthermore, and almost surely for the DM, an optimal indemnity scheme takes the form of a *generalized deductible contract*, defined hereafter (Corollary 6.3 on p. 18). In particular, a generalized deductible contract includes a deductible provision, whereby the insurer will not compensate losses of an amount lesser than a given threshold (the deductible). Also, when the beliefs of both parties coincide, the function h appearing in eq. (1.1) is the constant function equal to 1, and hence the function Z is simply a deductible contract, which is a nondecreasing function of the loss X . In this case, the main result of this paper (Theorem 6.2 on p. 17) boils down to the classical Arrow-Borch-Raviv result (Theorem 2.2 on p. 6).

The *vigilance* condition is a probabilistic consistency requirement on the insurer's subjective belief with respect to the DM's subjective belief. In particular, vigilance is trivially satisfied in the case of perfect homogeneity of beliefs, since it will be apparent from the definition of vigilance that any probability measure is vigilant with respect to itself. Vigilance can be interpreted as a kind of credibility that the insurer gives to the DM's assessment of the riskiness of a given insurance contract, and of its variability as a function of the underlying loss. Indeed, heuristically, one can think of an insurance scheme $Y = I(X)$ as having two sources of "randomness": a *baseline randomness* associated with the state space itself, that is, with the variability of Y with the state of nature; and, an *idiosyncratic randomness* associated with the variability of Y with the loss X itself, that is, the variability of the

function I with respect to the identity function⁴. Baseline randomness is belief-specific, in the sense that it depends on the distribution of X , whereas idiosyncratic randomness is belief-free. Vigilance of the insurer’s belief with respect to the DM’s belief can then be intuitively understood as requiring that, for a given baseline randomness fixed according to the DM’s belief, a comparison of the overall riskiness of two contracts from the point of view of the insurer can be restricted to a comparison of their idiosyncratic randomness only. Moreover, it will be understood that for two functions I and J , if function I ’s variability is similar to that of the identity function Id (that is, I is comonotonic with Id), then less idiosyncratic risk is attributed to I than to J . This is tantamount to a kind of credibility that the insurer gives to the DM’s probabilistic assessment of the riskiness of a contract.

Vigilance is simply a consistency requirement on two different probability measures. When the beliefs of the DM and the insurer are such that one can define probability density functions (pdf-s) for the loss X , then another well-known probabilistic consistency requirement that could be imposed on the agents’ beliefs is the *monotone likelihood ratio condition* (MLR). The MLR simply assumes that the ratio of the pdf-s is a monotone function (either nonincreasing or nondecreasing, depending on how the ratio is taken). I will show that the assumption of vigilance is (strictly) weaker than the MLR, which is commonly used in economics. Vigilance can then serve as substitute for, and as a weakening of the MLR in situations where a likelihood ratio can be defined, and it can be seen as a useful extension of the MLR to situations where densities do not exist and hence a likelihood ratio cannot be defined.

Related Literature. Notwithstanding the Harsányi Doctrine, that is the idea that disagreements about probabilities result only from difference in information, one should be cautious not to confound a problem of *belief heterogeneity* in an insurance model with a problem of *information asymmetry*. Problems of information asymmetry in insurance markets were usually studied in the context of adverse selection or moral hazard. The classical setup of adverse selection in insurance markets (e.g. [46, 53]) considers a risk-neutral EU-maximizing insurer and two types of risk-averse EU-maximizing insureds: a high-risk type (h) and a low risk-type (l). There are only two states of the world: an *accident* state and a *no-accident* state. An insurable loss X then takes the value 0 with a probability $1 - p_i$, for $i = h, l$, and a fixed value $L > 0$ with probability p_i . However, for each risk type i , both the insurer and the i -type insured have perfectly homogeneous beliefs, in that they both assign the distribution $(L, p_i; 0, 1 - p_i)$ to the loss. In this framework, “information asymmetry” refers to the fact that the type of the insured is private information, not the insured’s perception of the loss distribution. In insurance models with moral hazard (e.g. [2, 52, 54]), there is risk-neutral EU-maximizing insurer, a risk-averse EU-maximizing insured, and a two-state world (accident and no-accident) where the loss X can take a value $L > 0$ (in case of accident), with a known probability p_e that depends on the insured’s effort (e) in preventing the loss. In the no-accident state, the loss takes the value 0 with probability $1 - p_e$. However, both the insurer and the insured have perfectly homogeneous beliefs at each effort level. That is, for a given effort level e , both the insurer and the insured assign the distribution $(L, p_e; 0, 1 - p_e)$ to the loss. In this framework, “information asymmetry” refers to the fact that the effort level of the insured is private information, not the insured’s perception of the loss distribution for a given effort level.

Although the effect of heterogeneity of beliefs in an insurance market on the shape of an optimal contract is a very natural problem to examine, in light of the previous discussion, this problem was by and far left open. The only two exceptions that I am aware of are Marshall [41] and Jeleva and Villeneuve [39]. The latter extend the two-state adverse selection model of Stiglitz [53] to account for

⁴Indeed, if Id denotes the identity function on the range of X , then the insurance scheme $Y_0 = X = Id \circ X$ can be interpreted as the reference point for idiosyncratic randomness, i.e. randomness with respect to the loss X itself.

belief heterogeneity. In their framework, for each type of insured, the insurer and the insured have different beliefs about the probability of the loss taking the positive value L . Although the authors give an interesting analysis, one might argue that the two-state framework is of limited interest for at least two reasons: (i) typically, financial and insurance risks are not binary risks as in the two-state model; and (ii) assuming the classical constraint that an indemnity $I(X)$ be nonnegative and not larger than the loss itself, a two-state model where the loss X has a distribution $(L, p; 0, 1 - p)$ cannot determine the *shape* of the optimal indemnity schedule. For instance, the indemnity can be a *deductible* contract of the form $I(X) = \max(X - d, 0) := (X - d)^+$, for some $d \geq 0$. Indeed, in the no-accident state, the loss is 0, and so the indemnity is 0. In the accident state, the loss is $L > 0$, and so the indemnity is N , for some $N \in (0, L]$. Letting $d = L - N \geq 0$, one can then write the indemnity as $I(X) = (X - d)^+$. However, the indemnity can also be of the *coinsurance* type, i.e. of the form $I(X) = \alpha X$, for some $\alpha \in (0, 1]$. Indeed, in the no-accident state, the loss is 0, and so the indemnity is 0. In the accident state, the loss is $L > 0$, and so the indemnity is N , for some $N \in (0, L]$. Letting $\alpha = N/L$, one can then write the indemnity as $I(X) = \alpha X$.

Marshall [41] (hereafter, Marshall) considers a more general setup than that of Jeleva and Villeneuve [39], but remains restrictive nonetheless. The way in which the heterogeneity of beliefs is introduced in Marshall's insurance model is very specific: the DM (Marshall refers to this agent as "the client") assigns a probability density function (pdf) $f(t)$ to the insurable loss, whereas the insurer attributes the pdf $g(t)$ to the loss. Moreover, conditional on the event that the loss is nonzero, $f(t)$ and $g(t)$ agree. However, the probability of a zero loss is higher for the DM than for the insurer. This is a very restrictive approach to belief heterogeneity, since this heterogeneity is reduced only to the likelihood that each party attaches to the event of zero loss. In Section 5.3, I will examine Marshall's setup in more detail and show how it can be reduced to a particular case of the model proposed in this paper. In particular, I will show that the assumptions made in the setting of Marshall imply a form of vigilance.

The rest of this paper is organized as follows. In Section 2, I review the Arrow-Borch-Raviv model. In Section 3, I argue for the introduction of heterogeneity of beliefs into the Arrow-Borch-Raviv model. In Section 4, I introduce the formal model of insurance demand in the presence of belief heterogeneity. Section 5 is dedicated to the notion of vigilance. In Section 6, the DM's problem of demand for insurance coverage is examined, and the main results of this paper (Theorem 6.2 and Corollary 6.3) are stated. Section 7 concludes. All proofs, and some related analysis are collected in the [Appendices](#). The supplement to this paper contains some extensions of this paper's results and of the techniques used in the proofs of this paper's results. The supplement also illustrates how the idea of vigilance can be used to obtain a general monotone comparative statics result in a setting of heterogeneous uncertainty.

2. THE CLASSICAL ARROW-BORCH-RAVIV MODEL

The Arrow-Borch-Raviv model starts with a given random loss, modeled as an essentially bounded nonnegative random variable X on some exogenously given probability space (Ω, \mathcal{G}, P) . A DM seeks an insurance coverage against this random loss she is facing, and the market gives the DM the opportunity to purchase a coverage I from an insurer, for a premium Π set by the latter. Both the DM and the insurer are assumed to know what the distribution of X is and to agree on this distribution, in the sense that both assign to X the law $P \circ X^{-1}$ on the range $\mathcal{D} \subseteq \mathbb{R}^+$ of X . The insurance coverage is modeled as a Borel-measurable mapping $I : \mathcal{D} \rightarrow \mathbb{R}^+$, and \mathcal{D} is typically assumed to take the form $[0, M]$, for some $M < +\infty$. Moreover, the insurer is assumed to be a risk-neutral EU-maximizer, and the DM is assumed to be a strictly risk-averse EU-maximizer, with a strictly increasing, concave, and

twice continuously differentiable utility function u . The DM has an initial wealth of amount W_0 , and her wealth in state $\omega \in \Omega$ is given by

$$W(\omega) := W_0 - \Pi - X(\omega) + I(X(\omega))$$

The DM's demand problem is to find the indemnity I^* that will maximize her expected utility of wealth, subject to a feasibility constraint and a premium constraint. Formally, the problem is the following:

Problem 2.1. *For a given loading factor $\rho > 0$,*

$$\sup_I \left\{ \int u(W_0 - \Pi - X + I \circ X) dP \right\} : \begin{cases} 0 \leq I \circ X \leq X \\ \Pi \geq (1 + \rho) \int I \circ X dP \end{cases}$$

The first constraint is standard (see Arrow [3] and Raviv [45]), and says that an indemnity is nonnegative and cannot exceed the loss itself. The latter requirement simply rules out situations where the DM has an incentive to create damage (see Huberman, Mayers and Smith [37]), which would yield to *ex-post* moral hazard. The second constraint is simply the insurer's participation constraint, restated as a premium constraint. The classical literature has studied Problem 2.1 extensively, and it was shown that the solution is given by a *deductible contract* (e.g. [3, p. 212] or [45, Corollary 1]). Figure 1 below illustrates the shape of a deductible contract, in comparison with a full insurance contract.

Theorem 2.2 (Arrow-Borch-Raviv). *There exists some $d > 0$ such that $(I_d \circ X)$ is optimal for Problem 2.1, where I_d is a deductible contract defined by:*

$$I_d(t) = \begin{cases} 0 & \text{if } t < d \\ t - d & \text{if } t \geq d \end{cases}$$

That is, an optimal solution for Problem 2.1 takes the form $Y^ = \min[X, \max(0, X - d)]$, for some $d > 0$.*

The objectivity of the probability space (Ω, \mathcal{F}, P) means that the probability measure P is not determined from the preferences of the parties involved, and it hence cannot be a reflection of the subjective beliefs of the insurer and the DM. In this classical approach, both parties completely agree on the likelihoods attached to the different realizations of the underlying insurable loss, as measured by the objective probability measure P . In this sense, the uncertainty inherent in the insurable loss is assumed to be totally objective, *a priori*. While Aumann's [4] Agreement Theorem might give a justification for such an assumption, as was argued earlier, instances of divergence of beliefs are the most interesting ones to examine in practice, and the more realistic. The next section also gives some theoretical justifications for the introduction of belief heterogeneity into the classical Arrow-Borch-Raviv framework.

3. THE CASE FOR HETEROGENEITY OF BELIEFS IN THE ARROW-BORCH-RAVIV MODEL

As stated earlier, one might give two justifications for the assumption of homogeneity of beliefs in the Arrow-Borch-Raviv model. The first relies on the specific formulation of the idea of uncertainty

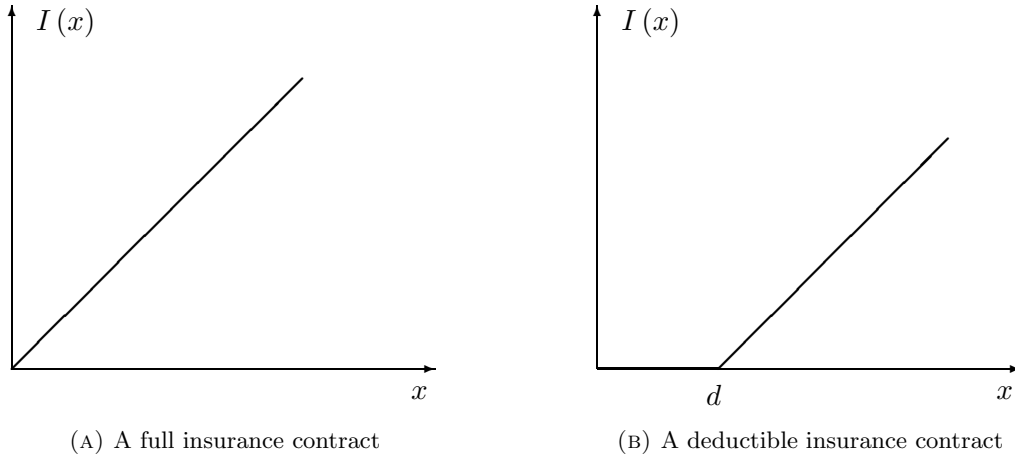


FIGURE 1. Full insurance above a deductible.

introduced by von Neumann and Morgenstern [55], whereby uncertainty is an “objective” concept, given prior to, and independently of an individual’s preferences over the objects of choice. The second justification relies on Aumann’s [4] information model and the Agreement Theorem: if the DM and the insurer have common prior beliefs, and if relevant information is shared, then the parties’ posterior beliefs must coincide. This then suggests a three-fold argument for the heterogeneity of beliefs in insurance markets:

- (a) first, a criticism of the CPA;
- (b) second, a criticism of the the assumption that posteriors are common knowledge, or that information is totally shared; and,
- (c) third, a criticism of the objectivity of the underlying uncertainty, a view that is inherited from the vNM approach and heavily challenged by subjectivist Bayesianism.

The CPA and the Harsányi Doctrine. In a series of groundbreaking papers, Harsányi [34, 35, 36] introduced the idea – which came to be known as the *Harsányi Doctrine* – that any disagreement about probabilities must be solely a result of difference in information. As Aumann [5] puts it:

“[P]eople with different information may legitimately entertain different probabilities, but there is no rational basis for people who have always been fed precisely the same information to do so.”

Aumann’s [4] Agreement Theorem can be viewed as a formalization of the Harsányi Doctrine. In light of the Agreement Theorem, the Harsányi Doctrine is equivalent to an assumption of common priors, but this is only one of the numerous reasons why the CPA has become so widespread; and, indeed, many fundamental results in economic theory are dependent on it⁵. Morris [43] gives an

⁵For instance, no-trade results à la Milgrom and Stokey [42] and Sebenius and Geanakoplos [51]; or results à la Aumann [5] and Aumann and Brandenburger [7] that aim at giving a decision-theoretic foundation of equilibrium reasoning in games.

excellent and still timely review and criticism of the CPA. The now famous Gul-Aumann [31, 6] exchange is also a debate over the validity of the CPA. It is not the object of this paper to partake in such a debate, but merely to draw attention to it, for by its very being it ought to serve as an indication that there might be reasonable grounds for abandoning the CPA altogether. After all, even Aumann [6] himself writes that the CPA “*embodies a reasonable and useful approach to interactive decision problems, though by no means the only such approach.*” Once the CPA is abandoned, heterogeneity of (prior and hence posterior) beliefs is obtained naturally.

Common Knowledge of Posteriors and the Wilson Doctrine. The second tenet of the information-structure justification of the assumption of belief homogeneity in the Arrow-Borch-Raviv framework, based on Aumann’s [4] Agreement Theorem, is the assumption of common knowledge of posteriors. When exactly can such common knowledge exist? One obvious case, in light of the Agreement Theorem, is that when the two agents have common priors and different posteriors, then they cannot have common knowledge of these posteriors. However, when this happens, heterogeneity of (posterior) beliefs is given by assumption, and this is precisely the case that is the object of this paper. Whether or not posterior beliefs are common knowledge becomes irrelevant.

In less trivial situations, Halpern and Moses [32] show that when communication is not guaranteed, common knowledge cannot be attained, and, indeed, common knowledge will not be achieved in many cases of interest. They also show that, even if communication is guaranteed, common knowledge cannot be attained if there is no time limit for messages to be delivered. Recently, Lehrer and Samet [40] examined instances where it is impossible *in principle* (not incidentally) for two agents to have common knowledge of their (posterior) beliefs, regardless of their priors. At any rate, dropping the assumption of common knowledge of posteriors seems to be equally justified as maintaining it.

On another level, one might argue for relaxing the assumption of common knowledge in an effort to espouse what has come to be known as the *Wilson Doctrine*, which is expressed in this quote from Wilson [56]:

“Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one agents probability assessment about anothers preferences or information. [...] I foresee the progress of game theory as depending on successive reduction in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality.”

There has been a movement in economic theory, in the past 10 years or so, towards incorporating the Wilson Doctrine, or some reinterpretation thereof, into the formulation of many theoretical problems, especially in problems of mechanism design⁶. In the Arrow-Borch-Raviv model, the Wilson Doctrine then stipulates that one should endeavour to relax the assumption of common knowledge among both parties. This will then lead to dropping the assumption of homogeneous (posterior) beliefs.

Subjective Uncertainty and Personal Probability. The approach to uncertainty in economic theory at large, and especially in decision theory, has been in constant modification since the work of vNM [55], which defines uncertainty as *totally objective*. A state-of-affairs where different individuals may entertain different beliefs regarding the realizations of an underlying uncertainty is *de facto* left out of consideration. However, the idea that uncertainty can be totally objective has been severely

⁶See Bergemann and Morris [8, 9, 10, 11], or Chung and Ely [18], for instance.

criticized in decision theory⁷, starting from the advocates of a *personal view of probability*, such as De Finetti [20], Ramsey [44], and especially Savage [49]. Contrary to the vNM approach, Savage formulates uncertainty in the objects of choice *without* objectively given probabilities. Beliefs and likelihoods are then derived from preferences, and probabilities are hence *subjective*, in the sense that the individual's preferences over the *acts* of choice induce a (unique) subjective probability measure representing the individual's beliefs. I am not concerned here with the philosophical questioning of this purely subjective approach to probability. My intent is merely to draw attention to the fact that the personalistic view of probability is by now a well-established foundational view of belief formation, and that some of its proponents feel very strongly that this ought to be the only approach to the idea of probability. Suffice it to quote Savage [49] who writes:

“[T]he concept of personal probability [...] is [...] the only probability concept essential to science and other activities that call upon probability.”

More passionately, De Finetti [21] writes:

“The abandonment of superstitious belief about [...] Fairies and Witches [...] was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs. [...] Probabilistic reasoning – always to be understood as subjective – merely stems from our being uncertain about something. It makes no difference whether the uncertainty relates to an unforeseeable future, or to an unnoticed past, or to a past doubtfully reported or forgotten [...] The only relevant thing is uncertainty – the extent of our knowledge and ignorance. The actual fact of whether or not the events considered are in some sense determined, or known by other people, and so on, is of no consequence.”

This subjectivist view of probability seems to be inconsistent with an assumption of common priors, given that the preferences of different agents typically differ. Consequently, as soon as one is willing to accept a purely personalistic view of uncertainty and belief formation, the heterogeneity of (subjective) beliefs arises then naturally as a consequence of the heterogeneity of preferences. Incidentally, the latter seems to be a rather uncontroversial assumption in economic theory.

Introducing Belief Heterogeneity into the Arrow-Borch-Raviv Model. The arguments presented above show that there are reasonable grounds for rejecting either (i) an assumption of common priors, or (ii) an assumption of common knowledge of posteriors, or both. When either (i), or (ii), or both are rejected, the Agreement Theorem cannot guarantee homogeneity of posterior beliefs. Consequently, belief heterogeneity can very well exist between the insurer and the DM.

Furthermore, on a practical level, the assumption of a common prior is a strong assumption in the insurance framework, and it is hardly justifiable in practice. Often, an insurer has far more experience with a particular kind of insurable loss than a DM does, for having encountered the same kind of loss in previous dealings with different insurees, for instance. The DM, on the other hand, may be a novice when facing this particular loss against which an insurance coverage is sought. One cannot then expect both parties to formulate identical prior beliefs about the uncertain loss. Moreover, even if both parties are assumed to have common priors, an assumption of complete information sharing is very strong and rather unrealistic. In practice, sharing information is costly and time-consuming, and

⁷The debate over the objectivity or subjectivity of probabilistic beliefs is also vivid in epistemology and the philosophy of science, especially in what might be called the philosophical foundations of probability theory.

hence typically uncared for. Therefore, practical considerations of the very nature of the insurance market leads one to deem conditions (i) and (ii) stated above as inapplicable, and hence no grounds for the assumption of homogeneous beliefs can be reasonably given. Additionally, as discussed above, heterogeneity of beliefs can be taken as a natural inherent characteristic of the insurance model, in keeping with the spirit of the Wilson Doctrine.

Moreover, it was argued that the conception of uncertainty inherent in the Arrow-Borch-Raviv insurance model is a heritage of the *objectivist* approach to uncertainty, championed by von Neumann and Morgenstern (vNM) [55]. I see no reason why uncertainty should be perceived only in this fashion, especially that the subjectivist tradition of Ramsey-De Finetti-Savage has proven to be very fruitful in economic theory. As soon as a subjectivist approach to uncertainty is adopted, whereby beliefs are the personal probabilities determined from preferences, belief heterogeneity in the Arrow-Borch-Raviv model would be obtained naturally as a consequence of the heterogeneity of preferences. One can then consider that heterogeneity of beliefs is an inherent trait of insurance markets, and is a very natural state-of-affairs.

4. AN INSURANCE MODEL WITH HETEROGENEOUS BELIEFS

4.1. Setup. Let S denote the set of states of the world. The DM faces a loss X , taken to be a mapping of S onto a closed interval $[0, M]$, for some $M < +\infty$. In particular, there are states of the world in which the loss takes a zero value, that is, $\{s \in S : X(s) = 0\} \neq \emptyset$. The DM seeks an insurance coverage against this loss. Henceforth, I will denote by Σ the σ -algebra $\sigma\{X\}$ of subsets of S generated by the random loss X .

Denote by $B(\Sigma)$ the supnorm-normed Banach space of all bounded, \mathbb{R} -valued and Σ -measurable functions on (S, Σ) , and denote by $B^+(\Sigma)$ the collection of all \mathbb{R}^+ -valued elements of $B(\Sigma)$. For any $Y \in B(\Sigma)$, the supnorm of Y is given by $\|Y\|_s := \sup\{|Y(s)| : s \in S\} < +\infty$. Then by Doob's measurability theorem [1, Theorem 4.41], for any $Y \in B(\Sigma)$ there exists a Borel-measurable map $\zeta : \mathbb{R} \rightarrow \mathbb{R}$ such that $Y = \zeta \circ X$. For $C \subseteq S$, denote by $\mathbf{1}_C$ the indicator function of C . For any $A \subseteq S$ and for any $B \subseteq A$, denote by $A \setminus B$ the complement of B in A .

Definition 4.1. Two functions $Y_1, Y_2 \in B(\Sigma)$ are said to be comonotonic if

$$\left[Y_1(s) - Y_1(s') \right] \left[Y_2(s) - Y_2(s') \right] \geq 0, \text{ for all } s, s' \in S$$

For instance any $Y \in B(\Sigma)$ is comonotonic with any $c \in \mathbb{R}$. Moreover, if $Y_1, Y_2 \in B(\Sigma)$, and if Y_2 is of the form $Y_2 = I \circ Y_1$, for some Borel-measurable function I , then Y_2 is comonotonic with Y_1 if and only if the function I is nondecreasing.

The insurance market gives the DM the possibility of entering into an insurance contract with the insurer. Such a contract is represented by a pair (Π, I) , where $\Pi > 0$ is the premium paid by the DM in return of the indemnity I . The indemnity is a Borel-measurable map $I : [0, M] \rightarrow [0, M]$, such that $0 \leq I(X(s)) \leq X(s)$ for all $s \in S$. Then $Y := I \circ X \in B^+(\Sigma)$, $Y \leq X$, and $Y(s) = 0$ for all $s \in \{s \in S : X(s) = 0\}$, that is, $I(0) = 0$.

Most of the assumptions that will be made here are for the sake of comparison with the classical Arrow-Borch-Raviv framework. Both the DM and the insurer have preferences over the elements of $B^+(\Sigma)$ that have a Subjective Expected-Utility (SEU) representation. The DM's preferences induce a utility function $u : \mathbb{R} \rightarrow \mathbb{R}$, unique up to a positive linear transformation, and the insurer's preferences induce a utility function $v : \mathbb{R} \rightarrow \mathbb{R}$, also unique up to a positive linear transformation. Both the DM's

and the insurer's preferences are also assumed to satisfy the Arrow-Villegas *Monotone Continuity* axiom [16] hence yielding a unique countably additive subjective probability measure on the measurable space (S, Σ) , for each. The subjectivity of the beliefs of each of the DM and the insurer is reflected in the different subjective probability measure that each has over the measurable space (S, Σ) . Formally, the DM's beliefs are represented by the countably additive probability measure μ on (S, Σ) , and the insurer's beliefs are represented by the countably additive probability measure ν on (S, Σ) .

Additionally, as in the Arrow-Borch-Raviv framework, I suppose that the DM is risk averse, having a utility index u such that following holds:

Assumption 4.2. *The DM's utility is bounded and satisfies Inada's [38] conditions. Specifically,*

- (1) u is bounded;
- (2) $u(0) = 0$;
- (3) u is strictly increasing and strictly concave;
- (4) u is continuously differentiable; and,
- (5) $u'(0) = +\infty$ and $\lim_{x \rightarrow +\infty} u'(x) = 0$.

Remark 4.3. Assumption 4.2 (1) above on the DM's utility function is, strictly speaking, redundant. Indeed, the utility function given from the DM's preferences in Savage's SEU representation is a bounded function ([24, Theorem 14.1] or [27, Theorem 10.2]). Moreover, boundedness of utility functions has been widely accepted as a necessary feature for utilities to be able to properly discriminate between alternative choices [3, 17].

Furthermore, Assumption 4.2 (3) is as in the Arrow-Borch-Raviv model. Finally, assuming that u is strictly concave and continuously differentiable implies that u' is both continuous and strictly decreasing. This then implies that $(u')^{-1}$ is continuous and strictly decreasing, by the Inverse Function Theorem [47, pp. 221-223].

The DM has initial wealth $W_0 > \Pi$, and her total state-contingent wealth is the Σ -measurable, \mathbb{R} -valued and bounded function on S defined by

$$W(s) := W_0 - \Pi - X(s) + Y(s), \quad \forall s \in S$$

I will also make the assumption that the random loss X (with closed range $[0, M]$) has a nonatomic law induced by the probability measure μ , that the subjective probability measures μ and ν are not mutually singular⁸, and that the DM is almost certain that the random loss she will incur is not larger than her remaining wealth after the premium has been paid. Specifically:

Assumption 4.4. *Assume that:*

- (1) $\mu \circ X^{-1}$ is nonatomic;
- (2) $X \leq W_0 - \Pi$, μ -a.s. In other words, $\mu(\{s \in S : X(s) > W_0 - \Pi\}) = 0$;
- (3) μ and ν are not mutually singular.

⁸Two finite nonnegative measures m_1 and m_2 on the measurable space (S, Σ) are said to be mutually singular, denoted by $m_1 \perp m_2$, if there is some $A \in \Sigma$ such that $m_1(S \setminus A) = m_2(A) = 0$. In other words, $m_1 \perp m_2$ if there is a Σ -partition $\{A, B\}$ of the set S of states of nature such that μ_1 is concentrated on A and μ_2 is concentrated on B .

Assumption 4.4 (1) is a technical requirement that is needed for defining the *equimeasurable monotone rearrangement*, as in Appendix B. It means that the loss is diffused enough, and is a very common assumption in many situations (e.g. when it is assumed that a probability density function for X exists). Assumption 4.4 (2) simply states that the DM is well-diversified so that the particular loss exposure X against which she is seeking an insurance coverage is sufficiently small. Assumption 4.4 (3) means that the insurer and the DM do not have beliefs that are totally incompatible. However, this does not prevent the agents from assigning different probabilities to events, and they typically do not assign same likelihoods to the realizations of the uncertainty X . For instance, they might disagree on zero-probability events.

Finally, I assume that the insurer is risk-neutral. This assumption is common in contracting problems, principal-agent problems, and especially in the insurance framework, as in Arrow-Borch-Raviv model. Since the insurer's utility function v is unique up to a positive linear transformation, one can then assume, without loss of generality, that v is simply the identity function. The total state-contingent wealth of the insurer is the Σ -measurable, \mathbb{R} -valued and bounded function on S defined by

$$W^{ins}(s) := W_0^{ins} + \Pi - Y(s) - \rho Y(s), \quad \forall s \in S$$

where W_0^{ins} is the insurer's initial wealth and $\rho > 0$ is such that ρY is a (proportional) cost associated with handling the insurance contract Y , as in the model of Arrow [3] (or section III of Raviv [45]).

4.2. The DM's problem. The DM seeks the indemnity that will maximize her expected utility of wealth, under her subjective probability measure, subject to a constraint on the premium and to some constraints on the indemnity function. Specifically, the DM's problem is the following:

Problem 4.5. For a given loading factor $\rho > 0$,

$$\begin{aligned} & \sup_{Y \in B^+(\Sigma)} \left\{ \int u(W_0 - \Pi - X + Y) d\mu \right\} : \\ & \begin{cases} 0 \leq Y \leq X \\ \int v(W_0^{ins} + \Pi - (1 + \rho)Y) d\nu \geq v(W_0^{ins}) \end{cases} \end{aligned}$$

The first constraint is as in the classical Arrow-Raviv model. The second constraint is simply the insurer's participation constraint, or individual rationality constraint, where $R = v(W_0^{ins})$ is the insurer's reservation utility. Since v was assumed to be the identity function (by risk-neutrality of the insurer), the insurer's individual rationality constraint can be re-written as the premium constraint $W_0^{ins} + \Pi - (1 + \rho) \int Y d\nu \geq W_0^{ins}$, that is,

$$\Pi \geq (1 + \rho) \int Y d\nu$$

Hence, one can write the DM's problem as follows:

Problem 4.6. For a given loading factor $\rho > 0$,

$$\begin{aligned} & \sup_{Y \in B^+(\Sigma)} \left\{ \int u(W_0 - \Pi - X + Y) d\mu \right\} : \\ & \begin{cases} 0 \leq Y \leq X \\ \Pi \geq (1 + \rho) \int Y d\nu \end{cases} \end{aligned}$$

The second constraint in Problem 4.6 is commonplace in insurance problems. Raviv [45], for instance, assumes that the premium is at least equal to $(1 + \alpha) \int (I \circ X) dP$, where P in Raviv's [45] context is the probability measure common to both the DM and the insurer, and $\alpha \geq 0$ is a loading factor. Also typical in the insurance problem is to impose an additional monotonicity constraint on the desired optimal indemnity by requiring that it be nondecreasing in the loss. This constraint, first introduced by Huberman, Mayers and Smith [37], is meant to prevent moral hazard issues that might result from a downward misrepresentation of the loss by the DM. Here, I do not impose this monotonicity as a constraint, but rather achieve it as a property of an optimal indemnity.

4.3. An Information-Structure Interpretation. The model of heterogeneous beliefs introduced here can be interpreted in terms of the usual information structure approach of the differential information literature. Specifically, consider a set S of states of the world, equipped with a σ -algebra \mathcal{G} of events. The random loss X is a fixed \mathcal{G} -measurable mapping $X : S \rightarrow [0, M]$, for some $M < +\infty$. An information structure, or information model, for the problem of insurance demand is an object $I = \{(S, \mathcal{G}), (\mathcal{F}_1, \mu_0), (\mathcal{F}_2, \nu_0)\}$, where \mathcal{F}_1 and \mathcal{F}_2 are sub- σ -algebras of \mathcal{G} , μ_0 is the DM's prior probability distribution over the space (S, \mathcal{F}_1) , and $\nu_0 \neq \mu_0$ is the insurer's prior probability distribution over the space (S, \mathcal{F}_2) . Both parties are assumed to perfectly observe X , so that there is no difference between them in the received information regarding the realizations of the underlying loss. Consequently, one can then assume that $\mathcal{F}_1 = \mathcal{F}_2 = \Sigma$, where Σ is the σ -algebra $\sigma\{X\}$ of subsets of S generated by the loss X . Both parties are assumed to be Bayesian, and, upon receiving some information (the same information), they will update their prior beliefs μ_0 and ν_0 into interim beliefs μ and ν . Since μ_0 and ν_0 differ, so will μ and ν . Once these interim beliefs have been determined, the parties will engage in the economic activity.

5. VIGILANCE

The novelty of the insurance model presented in Section 4 is precisely the fact that it allows for heterogeneity of beliefs, which was lacking in the Arrow-Borch-Raviv framework and in the vast majority of the subsequent insurance literature. Here, the analysis will be restricted to a class of beliefs ν that are consistent with the belief μ in a specific sense that will be made precise below. This restriction is nevertheless general enough to encompass, for instance, cases where these heterogeneous beliefs induce a likelihood ratio that is monotone, as discussed later on.

5.1. Vigilant Beliefs.

Definition 5.1. The probability measure ν is said to be (μ, X) -vigilant, or the insurer is said to be vigilant, if for any two possible insurance contracts $Y_1, Y_2 \in B^+(\Sigma)$ such that

- (i) Y_1 and Y_2 have the same distribution under μ , i.e. $\mu \circ Y_1^{-1} = \mu \circ Y_2^{-1}$, and,
- (ii) Y_2 and X are comonotonic, i.e. $\left[Y_2(s) - Y_2(s') \right] \left[X(s) - X(s') \right] \geq 0$, for all $s, s' \in S$,

the following holds:

$$\int Y_2 d\nu \leq \int Y_1 d\nu$$

Clearly, μ is (μ, X) -vigilant. One possible interpretation of (μ, X) -vigilance, when one thinks of $\int Y d\nu$ as the minimum premium level that the insurer requires in exchange for paying the state-contingent indemnity Y to the DM, is as follows. Suppose that ν is (μ, X) -vigilant and let $Y_1 \in B^+(\Sigma)$

be a given possible insurance contract. Suppose that Y_2 is another possible insurance contract such that Y_2 is nondecreasing in X and such that the DM believes Y_2 to be identically distributed as Y_1 under her subjective probability measure μ (and hence the DM is indifferent between these two functions). Vigilance of the insurer's subjective belief means that the insurer will not perceive Y_2 to be riskier than Y_1 , in that the insurer is willing to accept a lower minimum premium level when promising to pay the indemnity Y_2 than when promising a payment of Y_1 . In this sense, *the insurer is vigilant in his assessment of the riskiness of Y_2 as a function of X* . In a sense, the insurer assigns some *credibility* to the DM's subjective assessment of Y_2 in reference to both Y_1 and X . This implies a certain *probabilistic consistency* between the DM's and the insurer's subjective beliefs for this particular class of risks. As will be seen later on, this consistency requirement is crucial to rule out moral hazard problems that might result from a downward misrepresentation of the loss by the DM.

Another interpretation of (μ, X) -vigilance – that reinforces the above idea of “credibility” – comes from the nature of the class of insurance contracts over which the parties' preferences are assumed to be given. These contracts have been defined as functions $Y : S \rightarrow \mathbb{R}^+$ of the form $Y = I(X)$. Contracts are then (measurable) functions of the loss X , which can itself be written as $Id(X)$ where Id denotes the identity function (on the range of X). Intuition then suggests that a given insurance scheme $Y = I(X)$ inherently possesses two sources of “variability”, or “randomness”:

- (i) a randomness that stems directly from the uncertainty inherent in the state space and in the loss X , and that will be called the *baseline randomness*; and,
- (ii) an *idiosyncratic randomness* that is associated with the function I itself, and with its variability with respect to the identity function Id , which is seen as a reference point for idiosyncratic randomness. If a given function I 's variability is similar to that of the identity function (that is, I is comonotonic with Id), then less idiosyncratic risk is attributed to I than to another function J .

Baseline randomness is hence subjective, in the sense that its evaluation is related to the distribution of X on the state space, which is itself related to the subjective beliefs of each party. Idiosyncratic randomness, on the other hand, is not belief-specific. It is an objective evaluation of the variability of a given function $I : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with respect to the identity function. The overall evaluation of a given insurance contract is then intuitively an aggregation of two levels of randomness: its (subjective) baseline randomness and its (objective) idiosyncratic randomness.

Now, consider two agents, agent 1 and agent 2, with respective beliefs P_1 and P_2 over (S, Σ) . If, for a given insurance scheme $I(X)$, both of the aforementioned measures of randomness are given to individual 2, then he is able to assess the overall “riskiness” of $I(X)$ by the process of aggregation that was heuristically described above. However, if only the idiosyncratic randomness is observable or trusted by agent 2, then vigilance of agent 2's belief with respect to agent 1's belief means that fixing the baseline randomness according to agent 1's belief, a comparison of the overall riskiness of contracts from the point of view of agent 2 can be restricted to a comparison of their idiosyncratic randomness only. In this sense, vigilance of agent 2's belief with respect to agent 1's belief is tantamount to a form of credibility that agent 2 gives to agent 1's belief.

5.2. Vigilance and the Monotone Likelihood Ratio Condition. Suppose that, in the setting of Section 4, the laws $\mu \circ X^{-1}$ and $\nu \circ X^{-1}$ are both absolutely continuous with respect to Lebesgue measure, with Radon-Nikodým derivatives f and g , respectively, where $f(t)$ is interpreted as the pdf that the DM assigns to the loss X and $g(t)$ is interpreted as the pdf that the insurer assigns to the loss X . Then $f(t)$ and $g(t)$ are both continuous functions with support $[0, M]$. The *likelihood ratio* is

the function $LR : [0, M] \rightarrow \mathbb{R}^+$ defined by

$$(5.1) \quad LR(t) := g(t)/f(t)$$

for all $t \in [0, M]$ such that $f(t) \neq 0$.

Consider the following two conditions that one might impose.

Condition 5.2 (Monotone Likelihood Ratio). *LR is a nonincreasing function on its domain.*

Condition 5.3 (Vigilance). *ν is (μ, X) -vigilant.*

The following proposition shows that the *vigilance* condition is strictly weaker than the *monotone likelihood ratio* condition in this particular setting. Its proof is given in Appendix C.

Proposition 5.4. *If Condition 5.2 (Monotone Likelihood Ratio) holds then condition 5.3 (Vigilance) holds. However, the converse is not true.*

The Likelihood Ratio is only defined in situations where densities exist. In other words, when the DM and the insurer assign distributions to the underlying loss random variable, with probability density functions, then the likelihood ratio can be defined. What Proposition 5.4 asserts is that not only is vigilance a weaker assumption than the MLR when a likelihood ratio can be defined, but also, in other situations where densities do not necessarily exist and hence likelihood ratios cannot be defined, the notion of vigilance might serve as a substitute.

5.3. Vigilance in Marshall's [41] Setting. In Marshall's setup the heterogeneity of beliefs is restricted to the following situation. The loss has a pdf $f(t)$ with a mass point of p_{DM}^0 at 0 for the DM, and a pdf $g(t)$ with a mass point of p_{In}^0 at 0 for the insurer, such that $p_{DM}^0 > p_{In}^0$.

Suppose that, in the setting of Section 4 of this paper, the laws $\mu \circ X^{-1}$ and $\nu \circ X^{-1}$ are both absolutely continuous with respect to Lebesgue measure, with Radon-Nikodým derivatives f and g , respectively, where $f(t)$ is interpreted as the pdf that the DM assigns to the loss X and $g(t)$ is interpreted as the pdf that the insurer assigns to the loss X . Suppose also that the pdf's $f(t)$ and $g(t)$ satisfy the setup of Marshall. Then for any $Y_1, Y_2 \in B^+(\Sigma)$ such that:

- (1) Y_1 and Y_2 have the same distribution under μ ,
- (2) $Y_1(s) = Y_2(s) = 0$ for each $s \in \{s \in S : X(s) = 0\}$ (e.g. $Y_1 \leq X$ and $Y_2 \leq X$),

the following holds:

$$\begin{aligned} \int Y_2 d\nu &= \mathbb{E}_\nu[Y_2] = \mathbb{E}_\nu[Y_2|X=0] \nu\{X=0\} + \mathbb{E}_\nu[Y_2|X>0] \nu\{X>0\} \\ &= \mathbb{E}_\nu[Y_2|X=0] p_{In}^0 + \mathbb{E}_\nu[Y_2|X>0] (1 - p_{In}^0) = \mathbb{E}_\nu[Y_2|X>0] (1 - p_{In}^0) \\ &= \mathbb{E}_\mu[Y_2|X>0] (1 - p_{In}^0) \quad (\text{since } f \text{ and } g \text{ are identical conditional on } X > 0) \\ &= \mathbb{E}_\mu[Y_1|X>0] (1 - p_{In}^0) \quad (\text{since } Y_1 \text{ and } Y_2 \text{ have the same distribution under } \mu) \\ &= \mathbb{E}_\nu[Y_1|X>0] (1 - p_{In}^0) \quad (\text{since } f \text{ and } g \text{ are identical conditional on } X > 0) \\ &= \mathbb{E}_\nu[Y_1] = \int Y_1 d\nu \end{aligned}$$

Therefore, in particular, ν is (μ, X) -vigilant, if the definition of vigilance is restricted to the class of functions $Y \in B^+(\Sigma)$ that take the value zero in each state of the world where the loss takes the value zero. For instance, if $0 \leq Y \leq X$, then this automatically holds. This is not a limitation, since these are precisely the only elements of $B^+(\Sigma)$ that are of interest to the DM, by the very nature of the DM's problem (Problem 4.6). One should also note that the MLR condition in this setting is artificial since the likelihood ratio only takes two distinct values and since any function taking only two distinct values is monotone.

6. THE DM'S DEMAND FOR INSURANCE CONTRACTS

Recall from Section 4 that the DM's problem is the following:

Problem 6.1. *For a given loading factor $\rho > 0$,*

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int u(W_0 - \Pi - X + Y) d\mu \right\} : \\ \left\{ \begin{array}{l} 0 \leq Y \leq X \\ \Pi \geq (1 + \rho) \int Y d\nu \end{array} \right.$$

The difference between Problem 6.1 and the classical Arrow-Borch-Raviv problem (Problem 2.1) is the fact that the probability measures ν and μ differ. When $\mu = \nu$, denoting this common probability measure by P takes one back to the classical framework. The main complication in the setting where $\nu \neq \mu$ is dealing with this belief heterogeneity. One insight into this comes from Lebesgue's decomposition theorem [19, Theorem 4.3.1].

By Lebesgue's decomposition theorem, there exists a unique pair (ν_{ac}, ν_s) of (nonnegative) finite measures on (S, Σ) such that $\nu = \nu_{ac} + \nu_s$, $\nu_{ac} \ll \mu$, and $\nu_s \perp \mu$. That is, for all $B \in \Sigma$ with $\mu(B) = 0$, one has $\nu_{ac}(B) = 0$, and there is some $A \in \Sigma$ such that $\mu(S \setminus A) = \nu_s(A) = 0$. It then also follows that $\nu_{ac}(S \setminus A) = 0$ and $\mu(A) = 1$. Note also that for all $Z \in B^+(\Sigma)$, $\int Z d\nu = \int_A Z d\nu_{ac} + \int_{S \setminus A} Z d\nu_s$. Furthermore, by the Radon-Nikodým theorem [19, Theorem 4.2.2] there exists a μ -a.s. unique Σ -measurable and μ -integrable function $h : S \rightarrow [0, +\infty)$ such that $\nu_{ac}(C) = \int_C h d\mu$, for all $C \in \Sigma$.

6.1. The Main Results. The Lebesgue decomposition of ν with respect to μ suggests a re-writing of the premium constraint appearing in Problem 6.1 as

$$\Pi / (1 + \rho) \geq \int Y d\nu = \int_A Y h d\mu + \int_{S \setminus A} Y d\nu,$$

and one can then re-write Problem 6.1 as follows: For a given loading factor $\rho > 0$,

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_A u(W_0 - \Pi - X + Y) d\mu + \int_{S \setminus A} u(W_0 - \Pi - X + Y) d\mu \right\} : \\ \left\{ \begin{array}{l} 0 \leq Y \mathbf{1}_A + Y \mathbf{1}_{S \setminus A} \leq X \mathbf{1}_A + X \mathbf{1}_{S \setminus A} \\ \Pi / (1 + \rho) \geq \int_A Y h d\mu + \int_{S \setminus A} Y d\nu \end{array} \right.$$

This then suggests a splitting of Problem 6.1 into two problems. Each one of these problems is then solved separately, and the individual solutions hence obtained are then combined appropriately

so as to obtain a solution for Problem 6.1. All details are provided in Appendix D, but heuristically, consider the problems:

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_A u(W_0 - \Pi - X + Y) d\mu : Y \mathbf{1}_A \leq X \mathbf{1}_A, \int_A Y h d\mu = \beta \right\}$$

for an appropriately chosen β , and

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_{S \setminus A} u(W_0 - \Pi - X + Y) d\mu : Y \mathbf{1}_{S \setminus A} \leq X \mathbf{1}_{S \setminus A}, \int_{S \setminus A} Y d\nu \leq \alpha \right\}$$

for an appropriately chosen α . Since $\mu(S \setminus A) = 0$, any feasible Y for the second problem is also optimal for that problem. Since $\mu(A) = 1$, the first problem can be written as

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int u(W_0 - \Pi - X + Y) d\mu : Y \mathbf{1}_A \leq X \mathbf{1}_A, \int Y h d\mu = \beta \right\}$$

The main part of solving Problem 6.1 becomes that of solving the latter problem above. This is a considerably simpler problem than Problem 6.1, since dealing with the heterogeneity of beliefs appearing in Problem 6.1 has been reduced to dealing simply with the function h . The following theorem characterizes an optimal solution of Problem 6.1. Its proof is given in Appendix D.

Theorem 6.2. *Suppose that the previous assumptions hold, and for each $\lambda \geq 0$ define the function $Y_\lambda^* \in B^+(\Sigma)$ by:*

$$(6.1) \quad Y_\lambda^* := \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(\lambda h) \right] \right) \right]$$

If the insurer's subjective probability measure ν is (μ, X) -vigilant, then there exists a $\lambda^ \geq 0$ and an optimal solution \mathcal{Y}^* to Problem 6.1 which is nondecreasing in the loss X , and such that \mathcal{Y}^* has the same distribution as $Y_{\lambda^*}^*$ under μ .*

Moreover, any other \mathcal{Z}^ which is nondecreasing in X and which has the same distribution as $Y_{\lambda^*}^*$ under μ is such that $\mathcal{Z}^* = \mathcal{Y}^*$, μ -a.s.*

Theorem 6.2 characterizes a class of solutions to Problem 6.1 in terms of their distribution for the DM, that is for the probability measure μ . Of course, when $\mu = \nu$, so that there is perfect homogeneity of beliefs as in the classical model, then $h = 1$, and so

$$Y_\lambda^* = \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(\lambda) \right] \right) \right] = \min \left[X, (X - d_\lambda)^+ \right],$$

where $d_\lambda := W_0 - \Pi - (u')^{-1}(\lambda)$. Since Y_λ^* is then a nondecreasing function of X , Theorem 6.2 simply says that there is some λ^* such that an optimal contract \mathcal{Y}^* for the DM is such that $\mathcal{Y}^* = \min \left[X, (X - d_{\lambda^*})^+ \right]$, μ -a.s., which is a result similar to the classical Arrow-Borch-Raviv theorem (Theorem 2.2 on p. 6).

Corollary 6.3 below goes a step further than Theorem 6.2 and characterizes the shape of an optimal insurance contract for the DM. It turns out that such an optimal contract takes the form of what I will refer to as a *generalized deductible contract*. Formally, an indemnity schedule $I : [0, M] \rightarrow [0, M]$ will be called a *generalized deductible insurance contract* when there is some $d \in [0, M]$ such that

$$I(t) = \begin{cases} 0 & \text{if } t \in [0, d) \\ f(t) & \text{if } t \in [d, M] \end{cases}$$

for some nondecreasing Borel-measurable function $f : [0, M] \rightarrow [0, M]$ such that $0 \leq f(t) \leq t$ for $t \in [0, M]$. Figure 2 below illustrates the shape of an example of a generalized deductible insurance contract.

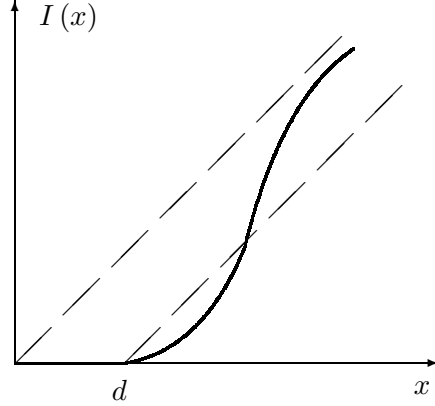


FIGURE 2. An example of a generalized deductible insurance contract.

Corollary 6.3. *Under the previous assumptions, and provided the insurer's subjective probability measure ν is (μ, X) -vigilant, there exists an optimal solution \mathcal{Y}^* to Problem 6.1 which is nondecreasing in the loss X , and such that for μ -a.a. $s \in S$,*

$$(6.2) \quad \mathcal{Y}^*(s) = \begin{cases} 0 & \text{if } X(s) \in [0, a^*) \\ f(X(s)) & \text{if } X(s) \in [a^*, M] \end{cases}$$

for an $a^* \geq 0$ and a nondecreasing, left-continuous, and Borel-measurable function $f : [0, M] \rightarrow [0, M]$ such that $0 \leq f(t) \leq t$ for each $t \in [a^*, M]$.

Sufficient conditions for a^* appearing in eq. (6.2) to be strictly positive are somewhat technical, and they are relegated to Appendix F.

The proof of Corollary 6.3 is given in Appendix E. Corollary 6.3 essentially says that when *vigilance* holds, there is a measurable set D to which the DM assigns full (subjective) probability, and such that an optimal indemnity schedule I^* will pay the DM, in the state of the world $s \in D$, the amount $I_{op}(s)$, where I_{op} is a *generalized deductible* contract on D . Two immediate implications of Corollary 6.3 are:

- (1) The existence of the deductible a^* . This is interesting mainly because of the resemblance with the classical result of Arrow [3], Borch [14], and Raviv [45] (see Theorem 2.2 on p. 6).
- (2) The nonlinearity of the insurance contract above the deductible a^* . I do not provide an explicit characterization of the function f that appears in Corollary 6.3, although it is possible to do so (see Remark E.14 on p. 39).

7. CONCLUSION AND OPEN QUESTIONS

The subjectivity of beliefs in problems of demand for insurance contracts was largely overlooked. The classical Arrow-Borch-Raviv approach to insurance demand has traditionally assumed that the

insurer and the insured – or decision maker (DM) – share the same probabilistic beliefs about the realization of a given insurable loss. While, intuitively, instances of divergence of beliefs in problems of insurance are the more realistic, Aumann’s [4] Agreement Theorem and the von Neumann-Morgenstern approach to uncertainty might give a justification for an assumption of belief homogeneity. I proposed an argument for the introduction of belief heterogeneity into the Arrow-Borch-Raviv framework based on a critique of the Common Priors Assumption (CPA) and the Harsányi Doctrine, a criticism of the assumption of common knowledge of posterior beliefs, the Wislon Doctrine, and the idea of subjective uncertainty and personal probability.

I then considered an insurance model in the spirit of the Arrow-Borch-Raviv model, with the exception that both parties have different beliefs about the realizations of the underlying insurable loss. I showed that under a specific probabilistic consistency assumption on the subjective probability measure of the insurer with respect to that of the DM that I called *vigilance*, an optimal insurance contract for the DM takes a *generalized deductible* form on an event to which the DM assigns full subjective probability. I also characterized the class of monotone optimal contracts for the DM in terms of their distribution for the DM.

The assumption of vigilance is always satisfied in the case of perfect homogeneity of beliefs. In cases of belief heterogeneity, vigilance can be intuitively interpreted in terms of the credibility that the insurer gives to the decision maker’s subjective assessment of the risk inherent in the insurance contract, and of the variability of the contract as a function of the underlying loss. Heuristically, two sources of randomness were identified that contribute to the overall assessment of the randomness in every insurance scheme: a *baseline randomness* that stems from the randomness inherent in the state space; and, an *idiosyncratic randomness* that results from the variability of the insurance contract as a function of the underlying loss, where it is understood that the more the variability of an insurance scheme is similar to that of the insurable loss, the less idiosyncratic risk is attributed to it. The former kind of randomness is belief-dependent, and depends on the specific probability measure that one has over the state space. The latter is belief-independent, and is hence an objective source of randomness associated with a given contract. Vigilance of the insurer’s subjective probability with respect to the DM’s subjective probability was then interpreted as requiring that, on the collection of all insurance schemes that are identically distributed for the DM – and hence between which the DM is indifferent, a comparison of the overall riskiness of two contracts from the point of view of the insurer can be restricted to a comparison of their idiosyncratic randomness only.

The assumption of vigilance was then shown to be strictly weaker than the monotone likelihood ratio assumption, which is commonly used in economics, whenever such a likelihood ratio can be defined – that is, whenever densities exist. Technically, this assumption of vigilance is essential to show existence of optimal indemnities which are comonotonic with the random loss X , and hence to avoid problems of *ex-post* moral hazard arising from a downward misrepresentation of losses by the DM.

Three problems of direct relevance to the insurance problem studied here are left for future research. First, in this paper I did not examine how the shape of an optimal contract changes with the “distance” between the subjective beliefs of the DM and the insurer, for an appropriately defined notion of “distance”. It is interesting to examine how, for instance, the deductible a^* given in eq. (6.2) varies with the “distance” between μ and ν , for a given fixed generalized deductible contract of the form given in eq. (6.2). Intuition suggests that a^* would tend to the Arrow-Borch-Raviv deductible d (given in Theorem 2.2) when ν “approaches” μ (letting P denote this common probability at the limit).

Another interesting problem to examine is a characterization of the notion of “*more* (μ, X) -*vigilant than*”, which is an issue that might arise when a DM is faced with the choice between two or more

potential insurers that she could contract with. Also, a related issue that arises in the presence of multiple potential insurers is a DM's "ranking" of these insurers according to some natural criterion. One such natural criterion would be defining a preference relation over a collection of vigilant insurers, and seeking a representation of this preference that might be helpful.

Finally, the tools developed by Ghossoub [25, 26] might suggest an extension of this paper's setting to situations of *ambiguity*, or *Knightian uncertainty*. These are situations of decision under uncertainty where the information available to a person is too coarse for that person to be able to formulate an additive probability measure over the list of contingencies. Rather, the person's beliefs are represented by either a collection of possible probability measures [12, 28, 30] or a non-additive probability measure [50]. I refer to Gilboa and Marinacci [29] for a review of models of decision under ambiguity.

APPENDIX A. TWO USEFUL RESULTS

Lemma A.1. *Let (Ω, \mathcal{F}) be a given measurable space, and suppose that η is a finite non-negative measure on (Ω, \mathcal{F}) . Let Z be any \mathbb{R}^+ -valued, bounded, and \mathcal{F} -measurable function on Ω . If $A \in \mathcal{F}$ is such that $\eta(A) > 0$, then the following are equivalent:*

- (1) $\int_A Z d\eta = 0$
- (2) $Z = 0$, η -a.s. on A .

Proof. [1, Theorem 11.16–(3)]. □

Lemma A.2. *Let (S, Σ, μ) be a finite nonnegative measure space. If $\{A_n\}_n \subset \Sigma$ is such that $\mu(A_n) = \mu(S)$, for each $n \geq 1$, then $\mu(\bigcap_{n=1}^{+\infty} A_n) = \mu(S)$.*

Proof. Since for each $n \geq 1$ one has $\mu(A_n) = \mu(S)$, it follows that $\mu(S \setminus A_n) = 0$, for each $n \geq 1$. Therefore, since μ is nonnegative, and by countable subadditivity of countably additive measures [19, Proposition 1.2.2], it follows that $0 \leq \mu(\bigcup_{n=1}^{+\infty} S \setminus A_n) \leq \sum_{n=1}^{+\infty} \mu(S \setminus A_n) = 0$. Therefore, $\mu(\bigcap_{n=1}^{+\infty} A_n) = \mu(S) - \mu(\bigcup_{n=1}^{+\infty} S \setminus A_n) = \mu(S)$. □

APPENDIX B. EQUIMEASURABLE REARRANGEMENTS AND SUPERMODULARITY

The classical theory of *monotone equimeasurable rearrangements* of Borel-measurable functions on \mathbb{R} dates back to the work of Hardy, Littlewood, and Pólya [33]. The theory was extended in several directions, and integral inequalities involving functions and their *rearrangements* were first given in [33] and then generalized. Here, the idea of an equimeasurable rearrangement of any element Y of $B^+(\Sigma)$ with respect to the fixed underlying uncertainty X is discussed. All of the results in this Appendix are taken from Ghossoub [25, 26] to which we refer the reader for proofs, additional results, and additional references on this topic. Ghossoub [25] extends these ideas to the case of non-additive probability measures, also known as capacities.

B.1. The Nondecreasing Rearrangement. Let (S, \mathcal{G}, P) be a probability space, and let $X \in B^+(\mathcal{G})$ be a continuous random variable (i.e. $P \circ X^{-1}$ is nonatomic) with range $[0, M] := X(S)$, where $M := \sup\{X(s) : s \in S\} < +\infty$, i.e. X is a mapping of S onto the closed interval $[0, M]$. Denote by Σ the σ -algebra $\sigma\{X\}$, and denote by ϕ the law of X defined by

$$\phi(B) := P(\{s \in S : X(s) \in B\}) = P \circ X^{-1}(B)$$

for any Borel subset B of \mathbb{R} .

Proposition B.1. *For any Borel-measurable map $I : [0, M] \rightarrow [0, M]$ there exists a ϕ -a.s. unique Borel-measurable map $\tilde{I} : [0, M] \rightarrow [0, M]$ such that:*

- (1) \tilde{I} is left-continuous and nondecreasing;
- (2) \tilde{I} is ϕ -equimeasurable with I , in the sense that for any Borel set B ,

$$\phi(\{t \in [0, M] : I(t) \in B\}) = \phi(\{t \in [0, M] : \tilde{I}(t) \in B\})$$

- (3) If $I_1, I_2 : [0, M] \rightarrow [0, M]$ are such that $I_1 \leq I_2$, ϕ -a.s., then $\tilde{I}_1 \leq \tilde{I}_2$; and,
- (4) If $Id : [0, M] \rightarrow [0, M]$ denotes the identity function, then $\tilde{Id} \leq Id$.

\tilde{I} will be called the nondecreasing ϕ -rearrangement of I . Now, define $Y := I \circ X$ and $\tilde{Y} := \tilde{I} \circ X$. Since both I and \tilde{I} are Borel-measurable mappings of $[0, M]$ into itself, it follows that $Y, \tilde{Y} \in B^+(\Sigma)$. Note also that \tilde{Y} is nondecreasing in X , in the sense that if $s_1, s_2 \in S$ are such that $X(s_1) \leq X(s_2)$ then $\tilde{Y}(s_1) \leq \tilde{Y}(s_2)$, and that Y and \tilde{Y} are P -equimeasurable, that is, for any $\alpha \in [0, M]$, $P(\{s \in S : Y(s) \leq \alpha\}) = P(\{s \in S : \tilde{Y}(s) \leq \alpha\})$. The function \tilde{Y} will be called a **nondecreasing P -rearrangement of Y with respect to X** , and it will be denoted by \tilde{Y}_P to avoid confusion in case a different measure on (S, \mathcal{G}) is also considered. Note that \tilde{Y}_P is P -a.s. unique. Note also that if Y_1 and Y_2 are P -equimeasurable and if $Y_1 \in L_1(S, \mathcal{G}, P)$, then $Y_2 \in L_1(S, \mathcal{G}, P)$ and $\int \psi(Y_1) dP = \int \psi(Y_2) dP$, for any measurable function ψ such that the integrals exist.

Similarly to the previous construction, for a given a Borel-measurable $B \subseteq [0, M]$ with $\phi(B) > 0$, there exists a ϕ -a.s. unique (on B) nondecreasing, Borel-measurable mapping $\tilde{I}_B : B \rightarrow [0, M]$ which is ϕ -equimeasurable with I on B , in the sense that for any $\alpha \in [0, M]$,

$$\phi(\{t \in B : I(t) \leq \alpha\}) = \phi(\{t \in B : \tilde{I}_B(t) \leq \alpha\})$$

\tilde{I}_B is called the nondecreasing ϕ -rearrangement of I on B . Since X is \mathcal{G} -measurable, there exists $A \in \mathcal{G}$ such that $A = X^{-1}(B)$, and hence $P(A) > 0$. Now, define $\tilde{Y}_A := \tilde{I}_B \circ X$. Since both I and \tilde{I}_B are bounded Borel-measurable mappings, it follows that $Y, \tilde{Y}_A \in B^+(\Sigma)$. Note also that \tilde{Y}_A is nondecreasing in X on A , in the sense that if $s_1, s_2 \in A$ are such that $X(s_1) \leq X(s_2)$ then $\tilde{Y}_A(s_1) \leq \tilde{Y}_A(s_2)$, and that Y and \tilde{Y}_A are P -equimeasurable on A , that is, for any $\alpha \in [0, M]$, $P(\{s \in S : Y(s) \leq \alpha\} \cap A) = P(\{s \in S : \tilde{Y}_A(s) \leq \alpha\} \cap A)$. The function \tilde{Y}_A will be called a **nondecreasing P -rearrangement of Y with respect to X on A** , and it will be denoted by $\tilde{Y}_{A,P}$ to avoid confusion in case a different measure on (S, \mathcal{G}) is also considered. Note that $\tilde{Y}_{A,P}$ is P -a.s. unique. Note also that if $Y_{1,A}$ and $Y_{2,A}$ are P -equimeasurable on A and if $\int_A Y_{1,A} dP < +\infty$, then

$\int_A Y_{2,A} dP < +\infty$ and $\int_A \psi(Y_{1,A}) dP = \int_A \psi(Y_{2,A}) dP$, for any measurable function ψ such that the integrals exist.

Lemma B.2. *Let $Y \in B^+(\Sigma)$ and let $A \in \mathcal{G}$ be such that $P(A) = 1$ and $X(A)$ is a Borel set⁹. Let \tilde{Y}_P be the nondecreasing P -rearrangement of Y with respect to X , and let $\tilde{Y}_{A,P}$ be the nondecreasing P -rearrangement of Y with respect to X on A . Then $\tilde{Y}_P = \tilde{Y}_{A,P}$, P -a.s.*

B.2. Supermodularity and Hardy-Littlewood-Pólya Inequalities. A partially ordered set (poset) is a pair (T, \succcurlyeq) where \succcurlyeq is a reflexive, transitive and antisymmetric binary relation on T . For any $x, y \in S$ denote by $x \vee y$ (resp. $x \wedge y$) the least upper bound (resp. greatest lower bound) of the set $\{x, y\}$. A poset (T, \succcurlyeq) is called a *lattice* when $x \vee y, x \wedge y \in T$, for each $x, y \in T$.

For instance, the Euclidian space \mathbb{R}^n is a lattice for the partial order \succcurlyeq defined as follows: for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, write $x \succcurlyeq y$ when $x_i \geq y_i$, for each $i = 1, \dots, n$. It is then easy to see that $x \vee y = (\max(x_1, y_1), \dots, \max(x_n, y_n))$ and $x \wedge y = (\min(x_1, y_1), \dots, \min(x_n, y_n))$.

Definition B.3. Let (T, \succcurlyeq) be a lattice. A function $L : T \rightarrow \mathbb{R}$ is said to be *supermodular* if for each $x, y \in T$,

$$(B.1) \quad L(x \vee y) + L(x \wedge y) \geq L(x) + L(y)$$

In particular, a function $L : \mathbb{R}^2 \rightarrow \mathbb{R}$ is *supermodular* if for any $x_1, x_2, y_1, y_2 \in \mathbb{R}$ with $x_1 \leq x_2$ and $y_1 \leq y_2$, one has

$$(B.2) \quad L(x_2, y_2) + L(x_1, y_1) \geq L(x_1, y_2) + L(x_2, y_1)$$

Equation (B.2) then implies that a function $L : \mathbb{R}^2 \rightarrow \mathbb{R}$ is supermodular if and only if the function $\eta(y) := L(x+h, y) - L(x, y)$ is nondecreasing on \mathbb{R} , for any $x \in \mathbb{R}$ and $h \geq 0$.

Example B.4. The following are supermodular functions:

- (1) If $g : \mathbb{R} \rightarrow \mathbb{R}$ is concave, and $a \in \mathbb{R}$, then the function $L_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $L_1(x, y) = g(a - x + y)$ is supermodular.
- (2) If $\psi, \phi : \mathbb{R} \rightarrow \mathbb{R}$ are both nonincreasing or both nondecreasing functions, then the function $L_4 : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $L_4(x, y) = \phi(x) \psi(y)$ is supermodular.

Lemma B.5 (Hardy-Littlewood-Pólya Inequalities). *Let $Y \in B^+(\Sigma)$ and let $A \in \mathcal{G}$ be such that $P(A) > 0$ and $X(A)$ is a Borel set. Let \tilde{Y}_P be the nondecreasing P -rearrangement of Y with respect to X , and let $\tilde{Y}_{A,P}$ be the nondecreasing P -rearrangement of Y with respect to X on A . If L is supermodular then:*

$$(1) \quad \int L(X, Y) dP \leq \int L(X, \tilde{Y}_P) dP; \text{ and,}$$

$$(2) \quad \int_A L(X, Y) dP \leq \int_A L(X, \tilde{Y}_{A,P}) dP,$$

⁹Note that if $A \in \Sigma = \sigma\{X\}$ then $X(A)$ is automatically a Borel set, by definition of $\sigma\{X\}$. Indeed, for any $A \in \sigma\{X\}$, there is some Borel set B such that $A = X^{-1}(B)$. Then $X(A) = B \cap X(S)$ [22, p. 7]. Thus $X(A) = B \cap [0, M]$ is a Borel subset of $[0, M]$.

provided the integrals exist.

Lemma B.6. *Let $Y \in B^+(\Sigma)$ and let $A \in \mathcal{G}$ be such that $P(A) > 0$ and $X(A)$ is a Borel set. Let \tilde{Y}_P be the nondecreasing P -rearrangement of Y with respect to X , and let $\tilde{Y}_{A,P}$ be the nondecreasing P -rearrangement of Y with respect to X on A . Then the following hold:*

- (1) *If $0 \leq Y \leq X$, P -a.s., then $0 \leq \tilde{Y}_P \leq X$; and,*
- (2) *If $0 \leq Y \leq X$, P -a.s. on A , then $0 \leq \tilde{Y}_{A,P} \leq X$, P -a.s. on A .*

B.3. Approximation of the Rearrangement.

Lemma B.7. *If f and f_n are $[0, +\infty)$ -valued, Σ -measurable functions on S such that the sequence $\{f_n\}_n$ converges pointwise P -a.s. to f monotonically downwards, then the sequence $\{\tilde{f}_{n,P}\}_n$ converges pointwise P -a.s. to \tilde{f}_P monotonically downwards, where \tilde{f}_P is the nondecreasing P -rearrangement of f with respect to X , and $\tilde{f}_{n,P}$ is the nondecreasing P -rearrangement of f_n with respect to X , for each $n \in \mathbb{N}$.*

Lemma B.8. *Let f and f_n be $[0, +\infty)$ -valued, Σ -measurable functions on S . If $f_n \in B^+(\Sigma)$, for each $n \geq 1$, and if the sequence $\{f_n\}_n$ converges uniformly to $f \in B^+(\Sigma)$, then*

- (1) *The functions \tilde{f}_P and $\tilde{f}_{n,P}$ are in L_∞ , for each $n \geq 1$, where \tilde{f}_P is the nondecreasing P -rearrangement of f with respect to X , and $\tilde{f}_{n,P}$ is the nondecreasing P -rearrangement of f_n with respect to X , for each $n \in \mathbb{N}$; and,*
- (2) *The sequence $\{\tilde{f}_{n,P}\}_n$ converges to \tilde{f}_P in the L_∞ norm.*

APPENDIX C. PROOF OF PROPOSITION 5.4

First, note that Condition 5.2 implies that the map $L : [0, M] \times [0, M] \rightarrow \mathbb{R}$ defined by

$$L(x, y) := -yLR(x)$$

is supermodular (see Example B.4 (2)).

Suppose that Condition 5.2 holds. To show that Condition 5.3 is implied, choose any $Y_1, Y_2 \in B^+(\Sigma)$ such that Y_1 and Y_2 have the same distribution under μ , and Y_2 is comonotonic with X . Then by the μ -a.s. uniqueness of the nondecreasing μ -rearrangement, Y_2 is μ -a.s. equal to $\tilde{Y}_{1,\mu}$, where \tilde{Y}_1 is the nondecreasing μ -rearrangement of Y_1 with respect to X . Furthermore, the function $L(x, y) = -yLR(x)$ is supermodular, as observed above. Consequently, by Lemma B.5, it follows that

$$\int L(X, \tilde{Y}_{1,\mu}) d\mu \geq \int L(X, Y_1) d\mu$$

that is,

$$\int \tilde{Y}_{1,\mu} Z d\mu \leq \int Y_1 Z d\mu$$

where Z is as defined above. Since $Y_2 = \tilde{Y}_{1,\mu}$, μ -a.s., we then have

$$\int Y_2 Z \, d\mu \leq \int Y_1 Z \, d\mu$$

which yields¹⁰ the following:

$$\int Y_2 \, d\nu \leq \int Y_1 \, d\nu$$

as required. Condition 5.3 hence follows from Condition 5.2.

The following example gives a situation where both the DM and the insurer assign a different pdf to the insurable loss and where the MLR condition fails but the Vigilance condition holds, hence showing that Condition 5.3 cannot imply Condition 5.2, i.e. Vigilance is strictly weaker than the MLR condition.

Example C.1. Suppose that $f(t) = 1/M$, that is, f is the pdf of a continuous uniformly distributed random variable on the interval $[0, M]$. Let $Y_1 = X \in B^+(\Sigma)$, and let $Y_2 = I \circ X \in B^+(\Sigma)$ be such that Y_1 and Y_2 are identically distributed under μ , and Y_2 is comonotonic with X . Then by the μ -a.s. uniqueness of the nondecreasing rearrangement, $Y_2 = \tilde{Y}_{1,\mu}$, μ -a.s., that is, $I = \tilde{I}d$, $\mu \circ X^{-1}$ -a.s., where Id denotes the identity function.

Fix $x^* \in (0, M)$ such that¹¹ $I(x^*) = \tilde{I}d(x^*)$, and let $g(t) = \delta(t - x^*)$, the Dirac delta function¹² centered at x^* . The likelihood ratio $LR = g/f$ is defined on the whole of $[0, M]$ and is clearly not nonincreasing. Therefore, the MLR condition fails. However, to see that the Vigilance condition still holds, note that

$$\begin{aligned} \int Y_2 \, d\nu &= \int I \circ X \, d\nu = \int_{X(S)} I \, d\nu \circ X^{-1} = \int_0^M I(t) g(t) \, dt \\ &= \int_0^M I(t) \delta(t - x^*) \, dt = I(x^*) = \tilde{I}d(x^*) \end{aligned}$$

Similarly, $\int Y_1 \, d\nu = Id(x^*) = x^*$. Hence, by Proposition B.1 (4),

$$\int Y_2 \, d\nu = \tilde{I}d(x^*) \leq Id(x^*) = \int Y_1 \, d\nu$$

and hence ν is (μ, X) -vigilant. This completes the proof of Proposition 5.4. \square

APPENDIX D. PROOF OF THEOREM 6.2

As in Section 6.1, there exists a unique pair (ν_{ac}, ν_s) of (nonnegative) finite measures on (S, Σ) such that $\nu = \nu_{ac} + \nu_s$, $\nu_{ac} \ll \mu$, and $\nu_s \perp \mu$. That is, for all $B \in \Sigma$ with $\mu(B) = 0$, one has $\nu_{ac}(B) = 0$, and there is some $A \in \Sigma$ such that $\mu(S \setminus A) = \nu_s(A) = 0$. It then also follows that $\nu_{ac}(S \setminus A) = 0$ and

¹⁰By two ‘‘changes of variable’’, as in [1, Theorem 13.46], and using the definition of f and g as Radon-Nikodým derivatives of $\mu \circ X^{-1}$ and $\nu \circ X^{-1}$, respectively, with respect to the Lebesgue measure.

¹¹This choice of x^* is possible for if it were not, than that would contradict the fact that Y_1 and Y_2 are identically distributed for μ .

¹²For any $a \in \mathbb{R}$, the Dirac delta function centered at a is zero everywhere except at a where it is infinite. Moreover, $\int_{a-\varepsilon}^{a+\varepsilon} \phi(t) \delta(t - a) \, dt = \phi(a)$, for any $\varepsilon > 0$ and for any function ϕ . See, e.g. [15, Chap. 5].

$\mu(A) = 1$. In the following, the Σ -measurable set A on which μ is concentrated and $\nu_s(A) = 0$ is assumed to be fixed all throughout.

D.1. “Splitting” the Initial Problem.

Lemma D.1. *Let Y^* be an optimal solution for Problem 6.1, and suppose that ν is (μ, X) -vigilant. Let \tilde{Y}_μ^* be the nondecreasing μ -rearrangement of Y^* with respect to X . Then:*

- (1) \tilde{Y}_μ^* is optimal for Problem 6.1; and,
- (2) $\tilde{Y}_\mu^* = \tilde{Y}_{\mu,A}^*$, μ -a.s., where $\tilde{Y}_{\mu,A}^*$ is the nondecreasing μ -rearrangement of Y^* with respect to X on A . In particular, Y^* and $\tilde{Y}_{\mu,A}^*$ have the same distribution under μ .

Proof. Since the function $\mathcal{U} : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $\mathcal{U}(x, y) := u(W_0 - \Pi - x + y)$ is supermodular (see Example B.4 (1)), it follows from Lemma B.5 that

$$\int u(W_0 - \Pi - X + \tilde{Y}_\mu^*) d\mu \geq \int u(W_0 - \Pi - X + Y^*) d\mu$$

Moreover, since $0 \leq Y^* \leq X$, it follows from Lemma B.6 that $0 \leq \tilde{Y}_\mu^* \leq X$. Finally, since ν is (μ, X) -vigilant, it follows that

$$\Pi/(1 + \rho) \geq \int Y^* d\nu \geq \int Y_\mu^* d\nu$$

and so \tilde{Y}_μ^* is optimal for Problem 6.1.

Now, let $\tilde{Y}_{\mu,A}^*$ be the nondecreasing μ -rearrangement of Y^* with respect to X on A . Since $\mu(A) = 1$, then by Lemma B.2 one has that $\tilde{Y}_\mu^* = \tilde{Y}_{\mu,A}^*$, μ -a.s. Therefore, \tilde{Y}_μ^* and $\tilde{Y}_{\mu,A}^*$ have the same distribution under μ . Hence, from the equimeasurability of Y^* and \tilde{Y}_μ^* , it follows that Y^* and $\tilde{Y}_{\mu,A}^*$ have the same distribution under μ . \square

Lemma D.2. *Let an optimal solution for Problem 6.1 be given by:*

$$(D.1) \quad Y^* = Y_1^* \mathbf{1}_A + Y_2^* \mathbf{1}_{S \setminus A}$$

for some $Y_1^*, Y_2^* \in B^+(\Sigma)$. Let \tilde{Y}_μ^* be the nondecreasing μ -rearrangement of Y^* with respect to X , and let $Y_{1,\mu}^*$ be the nondecreasing μ -rearrangement of Y_1^* with respect to X . Then $\tilde{Y}_\mu^* = \tilde{Y}_{1,\mu}^*$, μ -a.s., and hence Y^* and $\tilde{Y}_{1,\mu}^*$ have the same distribution under μ .

Proof. Let $\tilde{Y}_{\mu,A}^*$ be the nondecreasing μ -rearrangement of Y^* with respect to X on A . Since $\mu(A) = 1$, then by Lemma B.2 one has $\tilde{Y}_\mu^* = \tilde{Y}_{\mu,A}^*$, μ -a.s. Similarly, let $\tilde{Y}_{1,\mu,A}^*$ be the nondecreasing μ -rearrangement of Y_1^* with respect to X on A . Then $\tilde{Y}_{1,\mu}^* = \tilde{Y}_{1,\mu,A}^*$, μ -a.s. Therefore, it suffices to show that $\tilde{Y}_{\mu,A}^* = \tilde{Y}_{1,\mu,A}^*$, μ -a.s. Since both $\tilde{Y}_{\mu,A}^*$ and $\tilde{Y}_{1,\mu,A}^*$ are nondecreasing functions of X on A , then by the μ -a.s. uniqueness of the nondecreasing rearrangement, it remains to show that they are μ -equimeasurable with Y^* on A . Now, for each $t \in [0, M]$,

$$\begin{aligned} \mu(\{s \in A : \tilde{Y}_{\mu,A}^*(s) \leq t\}) &= \mu(\{s \in A : Y^*(s) \leq t\}) = \mu(\{s \in A : Y_1^*(s) \leq t\}) \\ &= \mu(\{s \in A : \tilde{Y}_{1,\mu,A}^*(s) \leq t\}) \end{aligned}$$

where the first equality follows from the definition of $\tilde{Y}_{\mu,A}^*$ (equimeasurability), the second equality follows from equation (D.1), and the third equality follows from the definition of $\tilde{Y}_{1,\mu,A}^*$ (equimeasurability). Therefore, $\tilde{Y}_\mu^* = \tilde{Y}_{1,\mu}^*$, μ -a.s., and hence \tilde{Y}_μ^* and $\tilde{Y}_{1,\mu}^*$ have the same distribution under μ . Consequently, by equimeasurability of Y^* and \tilde{Y}_μ^* , it follows that Y^* and $\tilde{Y}_{1,\mu}^*$ have the same distribution under μ . \square

What Lemma D.1 asserts is that when the insurer's subjective probability measure is vigilant with respect to the DM's subjective probability measure in regards to the risk X , and if there exists an indemnity schedule which is perceived by the DM as optimal for her initial problem, then there exists another indemnity schedule which is perceived by the DM as optimal for her initial problem, and which rules out any possibility of moral hazard resulting from a voluntary downward misrepresentation of losses by the DM. Indeed, as long as an indemnity schedule is nondecreasing in the insurable loss, there is no incentive for the DM to misrepresent the loss downwards. Consider now the following three problems:

Problem D.3. For a given $\beta \in [0, \min(\Pi/(1+\rho), \int_A X d\nu)]$,

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_A u(W_0 - \Pi - X + Y) d\mu \right\} : \begin{cases} 0 \leq Y \mathbf{1}_A \leq X \mathbf{1}_A \\ \int_A Y d\nu = \beta \end{cases}$$

Problem D.4.

$$\sup_{Y \in B^+(\Sigma)} \left\{ \int_{S \setminus A} u(W_0 - \Pi - X + Y) d\mu \right\} : \begin{cases} 0 \leq Y \mathbf{1}_{S \setminus A} \leq X \mathbf{1}_{S \setminus A} \\ \int_{S \setminus A} Y d\nu \leq \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S \setminus A} X d\nu\right), \text{ for the same } \beta \text{ as in Problem D.3} \end{cases}$$

Problem D.5.

$$\sup_{\beta} \left[F_A^*(\beta) + F_A^*\left(\frac{\Pi}{1+\rho} - \beta\right) : 0 \leq \beta \leq \min\left(\Pi/(1+\rho), \int_A X d\nu\right) \right] : \begin{cases} F_A^*(\beta) \text{ is the supremum value of Problem D.3, for a fixed } \beta \\ F_A^*\left(\frac{\Pi}{1+\rho} - \beta\right) \text{ is the supremum value of Problem D.4, for the same fixed } \beta \end{cases}$$

Remark D.6. The feasibility sets of Problems D.3 and D.4 are nonempty. To see why this is true, first note that:

- (1) Since μ and ν are not mutually singular, by Assumption 4.4, and since $\mu(S \setminus A) = 0$, it follows that $\nu(A) > 0$;
- (2) Since $\nu(A) > 0$, $h \geq 0$, and $\nu(A) = \nu_{ac}(A) + \nu_s(A) = \nu_{ac}(A) = \int_A h d\mu$, it follows from Lemma A.1 that there exists some $B \in \Sigma$ such that $B \subseteq A$, $\mu(B) > 0$, and $h > 0$ on B .

If $\int_A X d\nu = \int_A Xh d\mu = 0$, then by Lemma A.1 one has $Xh = 0$, μ -a.s. on A . However, $h > 0$ on B . Thus, $X = 0$, μ -a.s. on B . Consequently, there is some $C \in \Sigma$, with $C \subseteq B$ and $\mu(C) > 0$, such that $X = 0$ on C and $\mu(B \setminus C) = 0$. Therefore, $\mu(B) = \mu(C)$. Now, since $X(s) = 0$, for each $s \in C$, it follows that $C \subseteq \{s \in S : X(s) = 0\}$. Thus, by monotonicity of μ , $\mu(C) \leq \mu(\{s \in S : X(s) = 0\}) = \mu \circ X^{-1}(\{0\})$. But $\mu \circ X^{-1}(\{0\}) = 0$, by nonatomicity of $\mu \circ X^{-1}$ (Assumption 4.4). Therefore, $\mu(C) = 0$, a contradiction. Hence $\int_A X d\nu > 0$.

Now, for a given $\beta \in [0, \min(\Pi/(1+\rho), \int_A X d\nu)]$, the function $Y_1 := \frac{\beta X}{\int_A X d\nu}$ is feasible for Problem D.3 with parameter β .

If $\int_{S \setminus A} X d\nu = 0$, then $Y_2 := 0$ is feasible for Problem D.4. If $\int_{S \setminus A} X d\nu > 0$, then $Y_3 := \frac{\alpha X}{\int_{S \setminus A} X d\nu}$, with $\alpha := \min\left(\frac{\Pi}{1+\rho} - \beta, \int_{S \setminus A} X d\nu\right) / 2$, is feasible for Problem D.4 with parameter β , for any given $\beta \in [0, \min(\Pi/(1+\rho), \int_A X d\nu)]$.

Note also that by boundedness of the utility function u (Assumption 4.2), the supremum value of each of the above three problems is finite.

Lemma D.7. *If β^* is optimal for Problem D.5, Y_3^* is optimal for Problem D.3 with parameter β^* , and Y_4^* is optimal for Problem D.4 with parameter β^* , then $Y_2^* := Y_3^* \mathbf{1}_A + Y_4^* \mathbf{1}_{S \setminus A}$ is optimal for Problem 6.1.*

Proof. Feasibility of Y_2^* for Problem 6.1 is immediate. To show optimality of Y_2^* for Problem 6.1, let \tilde{Y} be any other feasible solution for Problem 6.1, and define $\alpha := \int_A \tilde{Y} d\nu$. Then $\alpha = \int_A \tilde{Y}h d\mu$ and $\int_A X d\nu = \int_A Xh d\mu$, since $\nu_s(A) = 0$. Moreover, $\alpha \in [0, \min(\Pi/(1+\rho), \int_A X d\nu)]$ since \tilde{Y} is feasible for Problem 6.1. Consequently, α is feasible for Problem D.5. Furthermore, $\tilde{Y} \mathbf{1}_A$ (resp. $\tilde{Y} \mathbf{1}_{S \setminus A}$) is feasible for Problem D.3 (resp. Problem D.4) with parameter α , and hence

$$\begin{cases} F_A^*(\alpha) \geq \int_A u(W_0 - \Pi - X + \tilde{Y}) d\mu \\ F_A^*\left(\frac{\Pi}{1+\rho} - \alpha\right) \geq \int_{S \setminus A} u(W_0 - \Pi - X + \tilde{Y}) d\mu \end{cases}$$

Now, since β^* is optimal for Problem D.5, it follows that

$$F_A^*(\beta^*) + F_A^*\left(\frac{\Pi}{1+\rho} - \beta^*\right) \geq F_A^*(\alpha) + F_A^*\left(\frac{\Pi}{1+\rho} - \alpha\right)$$

However,

$$\begin{cases} F_A^*(\beta^*) = \int_A u(W_0 - \Pi - X + Y_3^*) d\mu \\ F_A^*\left(\frac{\Pi}{1+\rho} - \beta^*\right) = \int_{S \setminus A} u(W_0 - \Pi - X + Y_4^*) d\mu \end{cases}$$

Therefore,

$$\int u(W_0 - \Pi - X + Y_2^*) d\mu \geq \int u(W_0 - \Pi - X + \tilde{Y}) d\mu$$

Hence, Y_2^* is optimal for Problem 6.1. \square

Remark D.8. By Lemmata D.1, D.2, and D.7, if ν is (μ, X) -vigilant, β^* is optimal for Problem D.5, Y_1^* is optimal for Problem D.3 with parameter β^* , and Y_2^* is optimal for Problem D.4 with parameter β^* , then \tilde{Y}_μ^* is optimal for Problem 6.1, and $\tilde{Y}_\mu^* = \tilde{Y}_{1,\mu}^*$, μ -a.s., where \tilde{Y}_μ^* (resp. $\tilde{Y}_{1,\mu}^*$) is the μ -a.s.

unique nondecreasing μ -rearrangement of $Y^* := Y_1^* \mathbf{1}_A + Y_2^* \mathbf{1}_{S \setminus A}$ (resp. of Y_1^*) with respect to X . In particular, Y^* and $\tilde{Y}_{1,\mu}^*$ have the same distribution under μ .

Lemma D.9. *If β^* is optimal for Problem D.5, then $\beta^* > 0$.*

Proof. First note the following:

- (i) Since $\mu(S \setminus A) = 0$, it follows that $\int_{S \setminus A} Z \, d\mu = 0$, for each $Z \in B(\Sigma)$, and so $F_A^* \left(\frac{\Pi}{1+\rho} - \beta \right) = 0$, for each $\beta \in [0, \min(\Pi/(1+\rho), \int_A X \, d\nu)]$. Consequently, $F_A^*(\beta) + F_A^* \left(\frac{\Pi}{1+\rho} - \beta \right) = F_A^*(\beta)$, for each $\beta \in [0, \min(\Pi/(1+\rho), \int_A X \, d\nu)]$. Therefore, in particular, $F_A^*(\beta^*) + F_A^* \left(\frac{\Pi}{1+\rho} - \beta^* \right) = F_A^*(\beta^*)$.
- (ii) Since μ and ν are not mutually singular, by Assumption 4.4, it follows that $\nu(A) > 0$.
- (iii) Since $\nu(A) > 0$, $h \geq 0$, and $\nu(A) = \nu_{ac}(A) + \nu_s(A) = \nu_{ac}(A) = \int_A h \, d\mu$, it follows from Lemma A.1 that there exists some $B \in \Sigma$ such that $B \subseteq A$, $\mu(B) > 0$, and $h > 0$ on B .

Now, suppose, *per contra*, that $\beta^* = 0$ is optimal for Problem D.5, and let Y_0 be optimal for Problem D.3 with parameter 0, so that $F_A^*(0) = \int_A u(W_0 - \Pi - X + Y_0) \, d\mu$. Since $\beta^* = 0$ is optimal for Problem D.5, one has $F_A^*(0) \geq F_A^*(\beta)$, for each $\beta \in [0, \min(\Pi/(1+\rho), \int_A X \, d\nu)]$. Since Y_0 is feasible for Problem D.3 with parameter $\beta^* = 0$, one has $\int_A Y_0 \, d\nu = \int_A Y_0 h \, d\mu = \beta^* = 0$. Now, since $\mu(A) > 0$ and $Y_0 h \geq 0$, it follows from Lemma A.1 that $Y_0 h = 0$, μ -a.s. on A . Moreover, since $h > 0$ on B and $\mu(B) > 0$, it follows that $Y_0 = 0$, μ -a.s. on B . Define the function Z by $Z := Y_0 \mathbf{1}_{A \setminus B} + \min(X, \Pi/(1+\rho)) \mathbf{1}_B$, and let $K_Z := \int_A Z \, d\nu$. Then the following clearly hold:

- (i) $Z \in B^+(\Sigma)$;
- (ii) $0 \leq Z \mathbf{1}_A \leq X \mathbf{1}_A$;
- (iii) $0 \leq K_Z \leq \min\left(\int_A X \, d\nu, \Pi/(1+\rho)\right)$.

Therefore, in particular, K_Z is feasible for Problem D.5 and Z is feasible for Problem D.3 with parameter K_Z . Moreover,

$$\begin{aligned} 0 \leq K_Z &= \int_{A \setminus B} Y_0 \, d\nu + \int_B \min(X, \Pi/(1+\rho)) \, d\nu \\ &= \int_B \min(X, \Pi/(1+\rho)) \, d\nu = \int_B \min(X, \Pi/(1+\rho)) h \, d\mu \end{aligned}$$

If $K_Z = 0$, then $\int_B \min(X, \Pi/(1+\rho)) h \, d\mu = 0$ and $\min(X, \Pi/(1+\rho)) h \geq 0$. Hence, by Lemma A.1, $\min(X, \Pi/(1+\rho)) h = 0$, μ -a.s. on B . However, $h > 0$ on B . Thus, $\min(X, \Pi/(1+\rho)) = 0$, μ -a.s. on B . Since $\Pi > 0$, this yields $X = 0$, μ -a.s. on B . Consequently, there is some $C \in \Sigma$, with $C \subseteq B$ and $\mu(C) > 0$, such that $X = 0$ on C and $\mu(B \setminus C) = 0$. Therefore, $\mu(B) = \mu(C)$. Now, since $X(s) = 0$, for each $s \in C$, it follows that $C \subseteq \{s \in S : X(s) = 0\}$. Thus, by monotonicity of μ , $\mu(C) \leq \mu(\{s \in S : X(s) = 0\}) = \mu \circ X^{-1}(\{0\})$. But $\mu \circ X^{-1}(\{0\}) = 0$, by nonatomicity of $\mu \circ X^{-1}$

(Assumption 4.4). Therefore, $\mu(C) = 0$, a contradiction. Hence $K_Z > 0$. Finally,

$$\begin{aligned} F_A^*(K_Z) &\geq \int_A u(W_0 - \Pi - X + Z) \, d\mu \\ &= \int_{A \setminus B} u(W_0 - \Pi - X + Y_0) \, d\mu + \int_B u(W_0 - \Pi - X + \min(X, \Pi/(1 + \rho))) \, d\mu \\ &\geq \int_{A \setminus B} u(W_0 - \Pi - X + Y_0) \, d\mu + \int_B u(W_0 - \Pi - X) \, d\mu \\ &= \int_A u(W_0 - \Pi - X + Y_0) \, d\mu := F_A^*(0) = F_A^*(\beta^*) \end{aligned}$$

This contradicts the optimality of $\beta^* = 0$ for Problem D.5. Consequently, if β^* is optimal for Problem D.5 then $\beta^* > 0$. \square

D.2. **Solving Problem D.4.** Since $\mu(S \setminus A) = 0$, it follows that, for all $Y \in B^+(\Sigma)$, one has

$$\int_{S \setminus A} u(W_0 - \Pi - X + Y) \, d\mu = 0$$

Consequently, any Y which is feasible for Problem D.4 with parameter β is also optimal for Problem D.4 with parameter β . For instance, define $Y_4^* := \min \left[X, \max \{0, X - \bar{d}_\beta\} \right]$, where \bar{d}_β is chosen such that $\int_{S \setminus A} Y_4^* \, d\nu \leq \min \left(\frac{\Pi}{1 + \rho} - \beta, \int_{S \setminus A} X \, d\nu \right)$. Then $Y_4^* \mathbf{1}_{S \setminus A}$ is optimal for Problem D.4.

Remark D.10. The choice of \bar{d}_β so that $\int_{S \setminus A} Y_4^* \, d\nu \leq \min \left(\frac{\Pi}{1 + \rho} - \beta, \int_{S \setminus A} X \, d\nu \right)$ is justified by the following argument: define the function $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ by

$$\phi(\alpha) = \int_{S \setminus A} Y_{4,\alpha} \, d\nu$$

where $Y_{4,\alpha} := \min[X, \max\{0, X - \alpha\}]$, for each $\alpha \geq 0$. Then ϕ is a nonincreasing function of α . Moreover, by the continuity of the functions $\max(0, \cdot)$ and $\min(x, \cdot)$, and by Lebesgue's Dominated Convergence Theorem, ϕ is a continuous function of the parameter α . Now, by the continuity of the functions \max and \min , $\lim_{\alpha \rightarrow 0} Y_{4,\alpha} = X$ and $\lim_{\alpha \rightarrow +\infty} Y_{4,\alpha} = 0$. Therefore, by continuity of the function ϕ in α ,

$$\lim_{\alpha \rightarrow 0} \phi(\alpha) = \int_{S \setminus A} X \, d\nu \quad \text{and} \quad \lim_{\alpha \rightarrow +\infty} \phi(\alpha) = 0$$

Consequently, ϕ is a continuous nonincreasing function of α such that $\lim_{\alpha \rightarrow +\infty} \phi(\alpha) = 0$ and $\lim_{\alpha \rightarrow 0} \phi(\alpha) = \int_{S \setminus A} X \, d\nu$. Thus, by the Intermediate Value Theorem, one can always choose α such that $\phi(\alpha) \leq \min \left(\frac{\Pi}{1 + \rho} - \beta, \int_{S \setminus A} X \, d\nu \right)$, for any $\beta \in [0, \min(\Pi/(1 + \rho), \int_A X \, d\nu)]$.

D.3. Solving Problem D.3. For a fixed parameter $\beta \in [0, \min(\Pi/(1+\rho), \int_A X d\nu)]$, Problem D.3 will be solved “statewise”, as described below. Moreover, by Lemma D.9, one can restrict the analysis to the case where $\beta \in (0, \min(\Pi/(1+\rho), \int_A X d\nu)]$.

Lemma D.11. *If $Y^* \in B^+(\Sigma)$ satisfies the following:*

- (1) $0 \leq Y^*(s) \leq X(s)$, for all $s \in A$;
- (2) $\int_A Y^* h d\mu = \beta$, for some $\beta \in (0, \min(\Pi/(1+\rho), \int_A X d\nu)]$; and,
- (3) *There exists some $\lambda \geq 0$ such that for all $s \in A \setminus \{s \in S : h(s) = 0\}$,*

$$Y^*(s) = \arg \max_{0 \leq y \leq X(s)} \left[u(W_0 - \Pi - X(s) + y) - \lambda y h(s) \right]$$

Then the function $Z^ := Y^* \mathbf{1}_{A \setminus \{s \in S : h(s) = 0\}} + X \mathbf{1}_{A \cap \{s \in S : h(s) = 0\}}$ solves Problem D.3 with parameter β .*

Proof. Suppose that $Y^* \in B^+(\Sigma)$ satisfies (1), (2), and (3) above. Then Z^* is clearly feasible for Problem D.3 with parameter β . To show optimality of Z^* for Problem D.3 note that for any other $Y \in B^+(\Sigma)$ which is feasible for Problem D.3 with parameter β , one has, for all $s \in A \setminus \{s \in S : h(s) = 0\}$,

$$\begin{aligned} & u(W_0 - \Pi - X(s) + Z^*(s)) - u(W_0 - \Pi - X(s) + Y(s)) \\ &= u(W_0 - \Pi - X(s) + Y^*(s)) - u(W_0 - \Pi - X(s) + Y(s)) \\ &\geq \lambda [h(s) Y^*(s) - h(s) Y(s)] = \lambda [h(s) Z^*(s) - h(s) Y(s)] \end{aligned}$$

Furthermore, since u is increasing, since $0 \leq Y \leq X$ on A , and since $Z^*(s) = X(s)$ for all $s \in \{s \in S : h(s) = 0\} \cap A$, it follows that for all $s \in \{s \in S : h(s) = 0\} \cap A$,

$$u(W_0 - \Pi - X(s) + Z^*(s)) = u(W_0 - \Pi) \geq u(W_0 - \Pi - X(s) + Y(s))$$

Thus,

$$\int_{A \cap \{s \in S : h(s) = 0\}} u(W_0 - \Pi - X + Z^*) d\mu - \int_{A \cap \{s \in S : h(s) = 0\}} u(W_0 - \Pi - X + Y) d\mu \geq 0$$

Consequently,

$$\begin{aligned} & \int_A u(W_0 - \Pi - X + Z^*) d\mu - \int_A u(W_0 - \Pi - X + Y) d\mu \\ &\geq \int_{A \setminus \{s \in S : h(s) = 0\}} u(W_0 - \Pi - X + Z^*) d\mu - \int_{A \setminus \{s \in S : h(s) = 0\}} u(W_0 - \Pi - X + Y) d\mu \\ &\geq \lambda [\beta - \beta] = 0 \end{aligned}$$

which completes the proof. □

Lemma D.12. *For any $\lambda \geq 0$, the function given by*

$$(D.2) \quad Y_\lambda^* := \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(\lambda h) \right] \right) \right]$$

satisfies conditions (1) and (3) of Lemma D.11.

Proof. Fix $\lambda \geq 0$, fix $s \in A \setminus \{s \in S : h(s) = 0\}$, and consider the problem

$$(D.3) \quad \max_{0 \leq y \leq X(s)} f(y) := \left[u\left(W_0 - \Pi - X(s) + y\right) - \lambda y h(s) \right]$$

Since u is strictly concave (by Assumption 4.2), so is f , as a function of y . In particular, $f'(y)$ is a (strictly) decreasing function. Hence the first-order condition on f yields a global maximum for f at $y^* := X(s) - \left[W_0 - \Pi - (u')^{-1}(\lambda h(s)) \right]$. If $y^* < 0$, then since f' is decreasing, it is negative on the interval $[0, X(s)]$. Therefore, f is decreasing on the interval $[0, X(s)]$, and hence attains a local maximum of $f(0)$ at $y = 0$. If $y^* > X(s)$, then since f' is decreasing, it is positive on the interval $[0, X(s)]$. Therefore, f is increasing on the interval $[0, X(s)]$, and hence attains a local maximum of $f(X(s))$ at $y = X(s)$. If $0 < y^* < X(s)$, then the local maximum of f on the interval $[0, X(s)]$ is its global maximum $f(y^*)$. Consequently, the function $y^{**} := \min\left[X(s), \max(0, y^*)\right]$ solves the problem appearing in eq. (D.3). Since s and λ were chosen arbitrarily, this completes the proof of Lemma D.12. \square

Lemma D.13. For Y_λ^* defined in equation (D.2), the following holds:

$$Y_\lambda^* \mathbf{1}_{A \setminus \{s \in S : h(s) = 0\}} + X \mathbf{1}_{A \cap \{s \in S : h(s) = 0\}} = Y_\lambda^* \mathbf{1}_A$$

Therefore,

$$\int_A \left[Y_\lambda^* \mathbf{1}_{A \setminus \{s \in S : h(s) = 0\}} + X \mathbf{1}_{A \cap \{s \in S : h(s) = 0\}} \right] d\nu = \int_A Y_\lambda^* d\nu = \int_A Y_\lambda^* h d\mu$$

Proof. Indeed, if $s \in \{s \in S : h(s) = 0\}$, then $(u')^{-1}(\lambda h(s)) = (u')^{-1}(0) = +\infty$, by Assumption 4.2. Thus, for each $s \in \{s \in S : h(s) = 0\}$ one has

$$Y_\lambda^*(s) = \min \left[X(s), \max \left(0, X(s) - \left[W_0 - \Pi - (u')^{-1}(0) \right] \right) \right] = X(s)$$

The rest then follows trivially. \square

Lemma D.14. Define the function $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ as follows: for each $\lambda \in \mathbb{R}^+$,

$$\begin{aligned} \phi(\lambda) &:= \int_A \left[Y_\lambda^* \mathbf{1}_{A \setminus \{s \in S : h(s) = 0\}} + X \mathbf{1}_{A \cap \{s \in S : h(s) = 0\}} \right] d\nu \\ &= \int_A Y_\lambda^* d\nu = \int_A Y_\lambda^* h d\mu \end{aligned}$$

Then ϕ is a continuous nonincreasing function of the parameter λ .

Proof. First, recall that

$$Y_\lambda^* := \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(\lambda h) \right] \right) \right]$$

Continuity of ϕ is a direct consequence of Lebesgue's Dominated Convergence Theorem and of the continuity of each of the functions $(u')^{-1}$, $\max(0, \cdot)$, and $\min(x, \cdot)$ (the function $(u')^{-1}$ is continuous by Remark 4.3). The fact that ϕ is nonincreasing in λ results from the concavity of u , i.e. from the fact that u' is a nonincreasing function. \square

Lemma D.15. *Consider the function ϕ defined above. Then:*

- (1) $\lim_{\lambda \rightarrow 0} \phi(\lambda) = \int_A X \, d\nu$; and,
- (2) $\lim_{\lambda \rightarrow +\infty} \phi(\lambda) = 0$.

Proof. By continuity of the functions $(u')^{-1}$, $\max(0, \cdot)$, and $\min(x, \cdot)$, it follows that for each $s \in S$,
 $\lim_{\lambda \rightarrow 0} Y_\lambda^*(s) = \min \left[X(s), \max \left(0, X(s) - \left[W_0 - \Pi - (u')^{-1}(0) \right] \right) \right]$. Moreover, as was shown above,

$$\min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(0) \right] \right) \right] = X$$

Therefore, $\lim_{\lambda \rightarrow 0} Y_\lambda^*(s) = X(s)$, for each $s \in S$. Hence, by continuity of the function ϕ in λ , it follows that $\lim_{\lambda \rightarrow 0} \phi(\lambda) = \int_A X \, d\nu$. Similarly, by continuity of the functions $(u')^{-1}$, $\max(0, \cdot)$, and $\min(x, \cdot)$, one has that for each $s \in S$,

$$\lim_{\lambda \rightarrow +\infty} Y_\lambda^*(s) = \min \left[X(s), \max \left(0, X(s) - \left[W_0 - \Pi - (u')^{-1}(+\infty) \right] \right) \right]$$

However, by continuity of the function ϕ in λ , one has

$$\lim_{\lambda \rightarrow +\infty} \phi(\lambda) = \int_A \lim_{\lambda \rightarrow +\infty} Y_\lambda^* \, d\nu$$

But by Assumption 4.2, $(u')^{-1}(+\infty) = 0$, and by Assumption 4.4, $X \leq W_0 - \Pi$, μ -a.s. Moreover, $\mu(A) = 1$. Therefore,

$$\int_A \lim_{\lambda \rightarrow +\infty} Y_\lambda^* \, d\nu = \int_A \lim_{\lambda \rightarrow +\infty} Y_\lambda^* h \, d\mu = 0$$

□

Remark D.16. Hence, summing up, the function ϕ defined above is a nonincreasing continuous function of the parameter λ such that $\lim_{\lambda \rightarrow 0} \phi(\lambda) = \int_A X \, d\nu$ and $\lim_{\lambda \rightarrow +\infty} \phi(\lambda) = 0$. Therefore, $\phi(\lambda) \in [0, \int_A X \, d\nu]$, and so by the Intermediate Value Theorem, for each $\beta \in (0, \min(\Pi/(1+\rho), \int_A X \, d\nu)]$ one can chose $\bar{\lambda} = \bar{\lambda}_\beta \in [0, +\infty)$ such that

$$\beta = \phi(\bar{\lambda}) = \int \left[Y_{\bar{\lambda}}^* \mathbf{1}_{A \setminus \{s \in S: h(s)=0\}} + X \mathbf{1}_{A \cap \{s \in S: h(s)=0\}} \right] h \, d\mu$$

Therefore, by Lemmata D.11 and D.12, the function $Y_{\bar{\lambda}}^*$ defined above solves Problem D.3, with parameter β . Finally, let β^* be optimal for Problem D.5, let λ^* be chosen for β^* just as $\bar{\lambda}$ was chosen for β in Remark D.16, and let $Y_{\lambda^*}^*$ be a corresponding optimal solution for Problem D.4 with parameter β^* . The rest then follows from Remark D.8. The μ -a.s. uniqueness part of Theorem 6.2 follows from the uniqueness property of the nondecreasing rearrangement. This concludes the proof of Theorem 6.2. □

APPENDIX E. PROOF OF COROLLARY 6.3

Fix $\beta \in (0, \min(\Pi/(1+\rho), \int_A X d\nu)]$, and let $\bar{\lambda}$ be the corresponding λ , chosen as in to Remark D.16. Since h is nonnegative, Σ -measurable and μ -integrable, there is a sequence $\{h_n\}_n$ of nonnegative, μ -simple and μ -integrable functions on (S, Σ) that converges monotonically upwards and pointwise to h [19, Proposition 2.1.7]. Therefore, since u' is bicontinuous (so that, in particular, $(u')^{-1}$ is continuous), it follows that the sequence $\{Y_{\bar{\lambda},n}\}_n$, defined by

$$Y_{\bar{\lambda},n} := X - W_0 + \Pi + (u')^{-1}(\bar{\lambda}h_n),$$

converges pointwise to $Y_{\bar{\lambda}}$, defined by

$$Y_{\bar{\lambda}} := X - W_0 + \Pi + (u')^{-1}(\bar{\lambda}h)$$

Since the sequence $\{h_n\}_n$ converges monotonically upwards and pointwise to h , and since $(u')^{-1}$ is continuous and decreasing, it follows that the sequence $\{Y_{\bar{\lambda},n}\}_n$ converges monotonically downwards and pointwise to $Y_{\bar{\lambda}}$. Now, for each $n \in \mathbb{N}$, there is some $m_n \in \mathbb{N}$, a Σ -partition $\{B_{i,n}\}_{i=1}^{m_n}$ of S , and some nonnegative real numbers $\alpha_{i,n} \geq 0$, for $i = 1, \dots, m_n$, such that $h_n = \sum_{i=1}^{m_n} \alpha_{i,n} \mathbf{1}_{B_{i,n}}$. Since $X - W_0 + \Pi$ can be written as $\sum_{i=1}^{m_n} (X - W_0 + \Pi) \mathbf{1}_{B_{i,n}}$, it is then easy to see that

$$Y_{\bar{\lambda},n} = \sum_{i=1}^{m_n} \left((u')^{-1}(\bar{\lambda} \alpha_{i,n}) + X - W_0 + \Pi \right) \mathbf{1}_{B_{i,n}}$$

Define $Y_{\bar{\lambda},n}^*$ by

$$Y_{\bar{\lambda},n}^* := \min \left[X, \max \left(0, Y_{\bar{\lambda},n} \right) \right]$$

By continuity of the functions $\max(0, \cdot)$ and $\min(x, \cdot)$, and since $\max(0, t)$ and $\min(X(s), t)$ are nondecreasing functions of t for each $s \in S$, it follows that the sequence $\{Y_{\bar{\lambda},n}^*\}_n$ converges monotonically downwards and pointwise to $Y_{\bar{\lambda}}^*$ (given by equation (D.2)).

Remark E.1. For each $n \geq 1$, let $\tilde{Y}_{\bar{\lambda},n,\mu}^*$ denote the μ -a.s. unique nondecreasing μ -rearrangement of $Y_{\bar{\lambda},n}^*$ with respect to X . Then by Lemma B.7, the sequence $\{\tilde{Y}_{\bar{\lambda},n,\mu}^*\}_n$ converges monotonically downwards and pointwise μ -a.s. to $\tilde{Y}_{\bar{\lambda},\mu}^*$.

Note that, for each $n \in \mathbb{N}$, one can rewrite $Y_{\bar{\lambda},n}^*$ as

$$Y_{\bar{\lambda},n}^* = \sum_{i=1}^{m_n} I_{\bar{\lambda},n,i}^* \mathbf{1}_{B_{i,n}},$$

where, for $i = 1, \dots, m_n$,

$$I_{\bar{\lambda},n,i}^* := \min \left[X, \max \left(0, X - d_{\bar{\lambda},n,i} \right) \right]$$

and

$$d_{\bar{\lambda},n,i} := W_0 - \Pi - (u')^{-1}(\bar{\lambda}\alpha_{i,n})$$

Lemma E.2. For each $n \in \mathbb{N}$, and for each $i_0 \in \{1, 2, \dots, m_n\}$, $I_{\bar{\lambda},n,i_0}^*$ is either a full insurance contract or a deductible contract (with a strictly positive deductible) on the set $B_{i_0,n}$.

Proof. Fix $n \in \mathbb{N}$, and fix $i_0 \in \{1, 2, \dots, m_n\}$. If $\alpha_{i_0, n} > 0$ and $\bar{\lambda} \leq u'(W_0 - \Pi)/\alpha_{i_0, n}$, then since u' is decreasing (u is concave) it follows that $(u')^{-1}(\bar{\lambda}\alpha_{i_0, n}) \geq W_0 - \Pi$. Therefore, $(u')^{-1}(\bar{\lambda}\alpha_{i_0, n}) - W_0 + \Pi + X \geq X \geq 0$, and so $I_{\bar{\lambda}, n, i_0}^* = X$, a full insurance contract (on $B_{i_0, n}$).

If $\alpha_{i_0, n} = 0$, then $I_{\bar{\lambda}, n, i_0}^* = \min \left[X, \max \left(0, (u')^{-1}(0) + X - W_0 + \Pi \right) \right]$. But $(u')^{-1}(0) = +\infty$, by Assumption 4.2. Therefore, $(u')^{-1}(0) - W_0 + \Pi + X \geq X \geq 0$, and so $I_{\bar{\lambda}, n, i_0}^* = X$, a full insurance contract (on $B_{i_0, n}$).

If $\alpha_{i_0, n} > 0$ and $\bar{\lambda} > u'(W_0 - \Pi)/\alpha_{i_0, n}$, then since u' is strictly decreasing (u is strictly concave) it follows that $(u')^{-1}(\bar{\lambda}\alpha_{i_0, n}) < W_0 - \Pi$. Therefore, $0 < W_0 - \Pi - (u')^{-1}(\bar{\lambda}\alpha_{i_0, n}) = d_{\bar{\lambda}, n, i_0}^-$, and so $I_{\bar{\lambda}, n, i_0}^* = \left(X - d_{\bar{\lambda}, n, i_0}^- \right)^+$, a deductible insurance contract (on $B_{i_0, n}$) with a strictly positive deductible, where for any $a, b \in \mathbb{R}$, $(a - b)^+ := \max(0, a - b)$. \square

Remark E.3. Hence, a sequence $\{Y_{\bar{\lambda}, n}^*\}_n$ has been constructed that converges pointwise (on S and hence on A) to $Y_{\bar{\lambda}}^*$. Consequently, by Egoroff's theorem [48, Theorem 9.6], for each $\varepsilon > 0$, there exists some $B_\varepsilon \in \Sigma$, $B_\varepsilon \subseteq A$, with $\mu(A \setminus B_\varepsilon) < \varepsilon$, such that $\{Y_{\bar{\lambda}, n}^*\}_n$ converges to $Y_{\bar{\lambda}}^*$ uniformly on B_ε . In other words, for each $\varepsilon > 0$, there is some $B_\varepsilon \in \Sigma$, $B_\varepsilon \subseteq A$, with $\mu(A \setminus B_\varepsilon) < \varepsilon$, and there is some $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$, $|Y_{\bar{\lambda}, n}^*(s) - Y_{\bar{\lambda}}^*(s)| < \varepsilon/2^n$, for all $s \in B_\varepsilon$.

The following lemma is a direct consequence of Lemmata B.2 and B.7, and it is hence stated without a proof.

Lemma E.4. If $\tilde{Y}_{\bar{\lambda}, n, \mu}^*$ (resp. $\tilde{Y}_{\bar{\lambda}, \mu}^*$) denotes the nondecreasing μ -rearrangement of $Y_{\bar{\lambda}, n}^*$ (resp. $Y_{\bar{\lambda}}^*$) with respect to X , then $\{\tilde{Y}_{\bar{\lambda}, n, \mu}^*\}_n$ converges monotonically downwards and pointwise μ -a.s. to $\tilde{Y}_{\bar{\lambda}, \mu}^*$. Moreover, $\tilde{Y}_{\bar{\lambda}, n, \mu}^* = \tilde{Y}_{\bar{\lambda}, n, A, \mu}^*$, μ -a.s., where $\tilde{Y}_{\bar{\lambda}, n, A, \mu}^*$ denotes the nondecreasing μ -rearrangement of $Y_{\bar{\lambda}, n, \mu}^*$ with respect to X on A .

Let $C_{2, n} := \left\{ s \in S : Y_{\bar{\lambda}, n}^*(s) = X(s) \right\}$. Then $C_{2, n}$ is of the form¹³ $C_{2, n} = B_{k_1, n} \cup \dots \cup B_{k_N, n}$, for some $\{k_1, k_2, \dots, k_N\} \subseteq \{1, 2, \dots, m_n\}$. Therefore,

$$Y_{\bar{\lambda}, n}^* = \sum_{j \in J} \left(X - d_{\bar{\lambda}, n, j}^- \right)^+ \mathbf{1}_{B_{j, n}} + X \mathbf{1}_{C_{2, n}}$$

for $J = \{1, 2, \dots, m_n\} \setminus \{k_1, k_2, \dots, k_N\}$.

Lemma E.5. Fix $n \in \mathbb{N}$. If there exists some $i_0 \in \{1, 2, \dots, m_n\}$ such that $\alpha_{i_0, n} = 0$ and $B_{i_0, n} \setminus \{s \in S : X(s) = 0\} \neq \emptyset$, then $C_{2, n} \setminus \{s \in S : X(s) = 0\} \neq \emptyset$.

Proof. Immediate, in light of the second paragraph in the proof of Lemma E.2. \square

¹³Note that since the random loss X is a mapping of S onto the closed interval $[0, M]$, it follows that $\{s \in S : X(s) = 0\} \neq \emptyset$, as mentioned previously. Now, since $0 \leq Y_{\bar{\lambda}, n}^* \leq X$, it follows that $\emptyset \neq \{s \in S : X(s) = 0\} \subseteq C_{2, n}$. Therefore, $C_{2, n} \neq \emptyset$.

Lemma E.6. *If μ is not absolutely continuous with respect to ν , then for each $n \in \mathbb{N}$ there is some $i_0 \in \{1, 2, \dots, m_n\}$ such that $\alpha_{i_0, n} = 0$.*

Proof. Suppose, *per contra*, that μ is not absolutely continuous with respect to ν but that there is some $n \in \mathbb{N}$ such that $\alpha_{i_0, n} > 0$, for each $i_0 \in \{1, 2, \dots, m_n\}$. Then $h_n = \sum_{i=1}^{m_n} \alpha_{i, n} \mathbf{1}_{B_{i, n}} > 0$. But the sequence $\{h_n\}_n$ converges monotonically upwards, and pointwise, to $h := d\nu_{ac}/d\mu$. Hence, since $h_n > 0$, it follows that $h(s) \geq h_k(s) > 0$, for each $s \in S$ and for each $k \geq n$. Consequently, $h > 0$. Therefore μ and ν_{ac} are mutually absolutely continuous (i.e. equivalent [13, p. 179]). Furthermore, the finite measures ν , ν_{ac} , and ν_s are nonnegative, and hence $\nu_{ac} \ll \nu$. Thus, $\mu \ll \nu$, a contradiction. \square

Lemma E.7. *If $\mu = \nu$ then $C_{2, n} \setminus \{s \in S : X(s) = 0\} = \emptyset$, for each $n \geq 1$.*

Proof. Suppose that $\mu = \nu$. Then, in this case, $\nu_s = 0$, $\nu_{ac} = \nu = \mu$, and so $h = 1$ and $A = S$. Thus, $h_n = 1$, for all $n \in \mathbb{N}$.

I claim that $\bar{\lambda} > u'(W_0 - \Pi)$. Suppose, *per contra*, that $\bar{\lambda} \leq u'(W_0 - \Pi)$. Then by concavity of u , u' is decreasing, and so $(u')^{-1}(\bar{\lambda}) \geq W_0 - \Pi$. Therefore,

$$Y_{\bar{\lambda}}^* = \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1}(\bar{\lambda}) \right] \right) \right] = X,$$

contradicting the classical result that a deductible insurance contract, with a positive deductible, is optimal in this case (as in Raviv [45] or Theorem 2.2). Therefore, $\bar{\lambda} > u'(W_0 - \Pi)$. But then from the proof of Lemma E.2 it follows that $C_{2, n} \setminus \{s \in S : X(s) = 0\} = \emptyset$, for each $n \geq 1$. \square

Now, let $C_{1, n} := \{s \in S : Y_{\bar{\lambda}, n}^*(s) = 0\}$. Then $C_{1, n}$ is non-empty¹⁴ and of the form $C_{1, n} = C_{1, n}^{(i)} \cup C_{1, n}^{(ii)}$, where $C_{1, n}^{(i)} \subseteq C_{2, n}$ and $C_{1, n}^{(ii)} \subseteq S \setminus C_{2, n}$. Indeed, since $\{s \in S : X(s) = 0\} \neq \emptyset$ and $0 \leq Y_{\bar{\lambda}, n}^* \leq X$, it follows that for all $s \in \{s \in S : X(s) = 0\}$ one has $Y_{\bar{\lambda}, n}^*(s) = X(s) = 0$. It is then easily verified that

$$C_{1, n}^{(i)} := \{s \in C_{2, n} : Y_{\bar{\lambda}, n}^*(s) = 0\} = \{s \in S : X(s) = 0\} \neq \emptyset$$

Therefore, $C_{1, n} = \{s \in S : X(s) = 0\} \cup C_{1, n}^{(ii)}$. Moreover, one can write $C_{1, n}^{(ii)} = \bigcup_{j=k_{N+1}}^{k_Q} B_{j, n}$, for some $\{k_{N+1}, \dots, k_Q\} \subseteq J$. Letting $J' := J \setminus \{k_{N+1}, \dots, k_Q\}$, it follows that $0 < (X - d_{\bar{\lambda}, n, j})^+ = X - d_{\bar{\lambda}, n, j} < X$, for each $j \in J'$. Therefore,

$$Y_{\bar{\lambda}, n}^* = 0 \mathbf{1}_{C_{1, n}} + \sum_{j \in J'} (X - d_{\bar{\lambda}, n, j}) \mathbf{1}_{B_{j, n}} + X \mathbf{1}_{C_{2, n} \setminus \{s \in S : X(s) = 0\}}$$

One can assume, without loss of generality, that $\alpha_{j, n} < \alpha_{k, n}$, for all $j, k \in J'$ such that $j < k$. Then it is easily verified that $d_{\bar{\lambda}, n, j} < d_{\bar{\lambda}, n, k}$, because of the concavity of u .

Now, if $\tilde{Y}_{\bar{\lambda}, n, \mu}^*$ denotes the nondecreasing μ -rearrangement of $Y_{\bar{\lambda}, n}^*$ with respect to X , one has the following result:

¹⁴Since $0 \leq Y_{\bar{\lambda}, n}^* \leq X$ and $\{s \in S : X(s) = 0\} \neq \emptyset$.

Lemma E.8. *Let $\tilde{Y}_{\lambda,n,\mu}^*$ denotes the nondecreasing μ -rearrangement of $Y_{\lambda,n}^*$ with respect to X . There exists $a_n \in [0, M]$ such that for μ -a.a. $s \in S$,*

$$(E.1) \quad \tilde{Y}_{\lambda,n,\mu}^*(s) = \begin{cases} 0 & \text{if } X(s) \in [0, a_n) \\ f_n(X(s)) & \text{if } X(s) \in [a_n, M] \end{cases}$$

where $f_n : [0, M] \rightarrow [0, M]$ is a nondecreasing and Borel-measurable function such that $0 \leq f_n(t) \leq t$ for each $t \in [0, M]$, and, for $\mu \circ X^{-1}$ -a.a. $t \in [0, M]$, one has $f(t) > 0$ if $t > a_n$.

Proof. First note that $0 \leq \tilde{Y}_{\lambda,n,\mu}^* \leq X$, by Lemma B.6, since $0 \leq Y_{\lambda,n}^* \leq X$, by definition of $Y_{\lambda,n}^*$. Moreover, one has $Y_{\lambda,n}^* = I_{\lambda,n}^- \circ X$, for some Borel-measurable function $I_{\lambda,n}^- : [0, M] \rightarrow [0, M]$. Therefore, $\tilde{Y}_{\lambda,n,\mu}^* = \tilde{I}_{\lambda,n}^* \circ X$, where $\tilde{I}_{\lambda,n}^*$ is the nondecreasing $\mu \circ X^{-1}$ -rearrangement of $I_{\lambda,n}^-$. Let $f_n := \tilde{I}_{\lambda,n}^*$. Then $0 \leq f_n(t) \leq t$, for each $t \in [0, M]$, and $f_n : [0, M] \rightarrow [0, M]$ is nondecreasing and Borel-measurable. Now, note that

$$\begin{aligned} \mu\left(\{s \in S : Y_{\lambda,n}^*(s) \leq 0\}\right) &= \mu\left(\{s \in S : Y_{\lambda,n}^*(s) = 0\}\right) = \mu(C_{1,n}) \\ &= \mu\left(\{s \in S : Y_{\lambda,n}^*(s) \leq 0, X(s) = 0\}\right) \\ &\quad + \mu\left(\{s \in S : Y_{\lambda,n}^*(s) \leq 0, X(s) > 0\}\right) \\ &= \mu\left(\{s \in S : X(s) = 0\}\right) + \mu(C_{1,n}^{ii}) \\ &= \mu(C_{1,n}^{ii}) \end{aligned}$$

where the last equality follows from the nonatomicity of $\mu \circ X^{-1}$ (Assumption 4.4). Moreover, by equimeasurability, one has that

$$\mu\left(\{s \in S : Y_{\lambda,n}^*(s) \leq 0\}\right) = \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^* \leq 0\}\right)$$

However,

$$\begin{aligned} \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) \leq 0\}\right) &= \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) = 0\}\right) \\ &= \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) \leq 0, X(s) = 0\}\right) \\ &\quad + \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) \leq 0, X(s) > 0\}\right) \\ &= \mu\left(\{s \in S : X(s) = 0\}\right) \\ &\quad + \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) = 0, X(s) > 0\}\right) \\ &= \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) = 0, X(s) > 0\}\right) \end{aligned}$$

where the last equality follows from the nonatomicity of $\mu \circ X^{-1}$ (Assumption 4.4). Consequently,

$$\mu(C_{1,n}) = \mu(C_{1,n}^{ii}) = \mu\left(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) = 0, X(s) > 0\}\right)$$

Thus, if $\mu(C_{1,n}^{ii}) \neq 0$, then there exists $a_n > 0$ such that for μ -a.a. $s \in S$, $\tilde{Y}_{\lambda,n,\mu}^*(s) = 0$ if $X(s)$ belongs to $[0, a_n]$ or $[0, a_n)$, and $\tilde{Y}_{\lambda,n,\mu}^*(s) > 0$ if $X(s) > a_n$. Therefore, $f_n(t) > 0$ if $t > a_n$, for $\mu \circ X^{-1}$ -a.a. $t \in [0, M]$.

If $\mu(C_{1,n}^{ii}) = 0$, then $\mu(\{s \in S : \tilde{Y}_{\lambda,n,\mu}^*(s) = 0, X(s) > 0\}) = 0$, and so for μ -a.a. $s \in S$, $\tilde{Y}_{\lambda,n,\mu}^*(s) = 0$ if $X(s) = 0$, and $\tilde{Y}_{\lambda,n,\mu}^*(s) > 0$ if $X(s) > 0$. Thus, with $a_n = 0$, $\tilde{Y}_{\lambda,n,\mu}^*$ is μ -a.s. of the form equation (E.1), with $f_n(t) > 0$ if $t > a_n = 0$, for $\mu \circ X^{-1}$ -a.a. $t \in [0, M]$. \square

Remark E.9. For each $n \geq 1$, let $E_n \in \Sigma$ be the event such that $\mu(E_n) = 1$ and $\tilde{Y}_{\lambda,n,\mu}^*$ is of the form equation (E.1) on E_n . Let $E := \bigcap_{n=1}^{+\infty} E_n$. Then $E \in \Sigma$ and, by Lemma A.2, $\mu(E) = 1$. Moreover, for each $s \in E$, and for each $n \geq 1$, $\tilde{Y}_{\lambda,n,\mu}^*(s)$ is given by equation (E.1).

By Lemma E.4, the sequence $\{\tilde{Y}_{\lambda,m,\mu}^*\}_m$ defined by equation (E.1) converges pointwise μ -a.s. to $\tilde{Y}_{\lambda,\mu}^*$, the nondecreasing μ -rearrangement of Y_{λ}^* with respect to X .

Now, let $Y_{4,\beta}^*$ be an optimal solution to Problem D.4 with parameter β , as defined previously, and for each $m \in \mathbb{N}$ let

$$\tilde{Y}_{m,\beta}^* := \tilde{Y}_{\lambda,m,\mu}^* \mathbf{1}_A + Y_{4,\beta}^* \mathbf{1}_{S \setminus A}$$

Finally, let β^* be optimal for Problem D.5, let λ^* be chosen for β^* just as $\bar{\lambda}$ was chosen for β , and let Y_{4,β^*}^* be a corresponding optimal solution for Problem D.4 with parameter β^* . For each $n \geq 1$, let

$$\tilde{Y}_{m,\beta^*}^* := \tilde{Y}_{\lambda^*,m,\mu}^* \mathbf{1}_A + Y_{4,\beta^*}^* \mathbf{1}_{S \setminus A}$$

Then, by Remark D.8, the sequence $\{\tilde{Y}_{m,\beta^*}^*\}_m$ converges pointwise μ -a.s. to an optimal solution of the initial problem (Problem 6.1), which is μ -a.s. nondecreasing in the loss X . Henceforth, \mathcal{Y}^* will denote that optimal solution. Then

$$(E.2) \quad \mathcal{Y}^* \mathbf{1}_A = \tilde{Y}_{\lambda^*,\mu}^* \mathbf{1}_A$$

To conclude the proof of Corollary 6.3, it will now be shown that the optimal solution \mathcal{Y}^* to Problem 6.1 obtained above has the form of a *generalized deductible* contract, μ -a.s. That is, it will be shown that $\tilde{Y}_{\lambda^*,\mu}^*$ has the form of a *generalized deductible* contract, μ -a.s.

Recall that $\tilde{Y}_{\lambda^*,\mu}^*$ is the μ -a.s. unique nondecreasing μ -rearrangement of $Y_{\lambda^*}^*$ with respect to X , where

$$Y_{\lambda^*}^* := \min \left[X, \max \left(0, Y_{\lambda^*} \right) \right]$$

and

$$Y_{\lambda^*} := X - W_0 + \Pi + (u')^{-1}(\lambda^* h)$$

Moreover, the sequence $\{Y_{\lambda^*,m}\}_m$, defined by

$$Y_{\lambda^*,m} := X - W_0 + \Pi + (u')^{-1}(\lambda^* h_m),$$

converges pointwise to Y_{λ^*} . Since the sequence $\{h_m\}_m$ converges monotonically upwards and pointwise to h , and since $(u')^{-1}$ is continuous and decreasing, it follows that the sequence $\{Y_{\lambda^*,m}\}_m$ converges monotonically downwards and pointwise to Y_{λ^*} . Consequently, one can easily check that the sequence $\{Y_{\lambda^*,m}^*\}_m$ converges monotonically downwards and pointwise to $Y_{\lambda^*}^*$, where for each $m \geq 1$,

$$Y_{\lambda^*,m}^* := \min \left[X, \max \left(0, Y_{\lambda^*,m} \right) \right]$$

Remark E.10. For each $m \geq 1$, let $\tilde{Y}_{\lambda^*, m, \mu}^*$ denote the μ -a.s. unique nondecreasing μ -rearrangement of $Y_{\lambda^*, m}^*$ with respect to X . Then by Lemma B.7, the sequence $\{\tilde{Y}_{\lambda^*, m, \mu}^*\}_m$ converges monotonically downwards and pointwise μ -a.s. to $\tilde{Y}_{\lambda^*, \mu}^*$. That is, there is some $A^* \in \Sigma$ with $A^* \subseteq A$ and $\mu(A^*) = 1$, such that for each $s \in A^*$ the sequence $\{\tilde{Y}_{\lambda^*, m, \mu}^*(s)\}_m$ converges monotonically downwards to $\tilde{Y}_{\lambda^*, \mu}^*(s)$.

Now, as in Lemma E.8, for μ -a.a. $s \in S$,

$$\tilde{Y}_{\lambda^*, n, \mu}^*(s) = \begin{cases} 0 & \text{if } X(s) \in [0, a_n) \\ f_n(X(s)) & \text{if } X(s) \in [a_n, M] \end{cases}$$

for a given $a_n \geq 0$, and $f_n : [0, M] \rightarrow [0, M]$, a nondecreasing and Borel-measurable function such that $0 \leq f_n(t) \leq t$ for each $t \in [0, M]$, and $f(t) > 0$ if $t > a_n$ for $\mu \circ X^{-1}$ -a.a. $t \in [0, M]$.

Lemma E.11. *The sequence $\{a_m\}_m$ is bounded and nondecreasing.*

Proof. Since $\{a_m\}_m \subset [0, M]$, boundedness of the sequence $\{a_m\}_m$ is clear. I now show that it is nondecreasing. Fix $m \in \mathbb{N}$. Since the sequence $\{\tilde{Y}_{\lambda^*, m, \mu}^*\}_m$ is nonincreasing pointwise on A^* (as in Remark E.10), one has $\tilde{Y}_{\lambda^*, m, \mu}^*(s) \geq \tilde{Y}_{\lambda^*, m+1, \mu}^*(s)$, for each $s \in A^*$.

To show that $a_m \leq a_{m+1}$, first note that if $a_m = 0$, then $a_{m+1} \geq 0 = a_m$. If $a_m > 0$, let $E \in \Sigma$ be as in Remark E.9, let $A^* \in \Sigma$ be as in Remark E.10, and choose $s \in E \cap A^*$ such that $X(s) \in [0, a_m)$. Then $0 = \tilde{Y}_{\lambda^*, m, \mu}^*(s) \geq \tilde{Y}_{\lambda^*, m+1, \mu}^*(s) \geq 0$, and so $\tilde{Y}_{\lambda^*, m+1, \mu}^*(s) = 0$. Consequently, $X(s) \in [0, a_{m+1}]$, and so $[0, a_m) \subseteq [0, a_{m+1}]$. Therefore, $0 < a_m \leq a_{m+1}$. \square

Hence, the sequence $\{a_m\}_m$ is bounded and monotone. Therefore, it has a limit. Let

$$(E.3) \quad a := \lim_{m \rightarrow +\infty} a_m$$

Moreover, if there is some $n \geq 1$ such that $a_n > 0$, then for each $m \geq n$, one has $a_m \geq a_n > 0$.

Lemma E.12. *With a as defined in eq. (E.3) above, one has $0 \leq a \leq M$, and $a > 0$ if there is some $n \geq 1$ with $a_n > 0$.*

Proof. Since $0 \leq a_m \leq M$, for each $m \geq 1$, it follows that $0 \leq a \leq M$. Moreover, if there is some $n \geq 1$ such that $a_n > 0$, then for each $m \geq n$ one has $a_m \geq a_n > 0$. Therefore, $a \geq a_m > 0$, for each $m \geq n$, and so $a > 0$. \square

Lemma E.13. *There exist $a^* \geq 0$ such that for μ -a.a. $s \in S$,*

$$\mathcal{Y}^*(s) = \begin{cases} 0 & \text{if } X(s) \in [0, a^*) \\ f(X(s)) & \text{if } X(s) \in [a^*, M] \end{cases}$$

for some nondecreasing, left-continuous, and Borel-measurable function $f : [0, M] \rightarrow [0, M]$ such that $0 \leq f(t) \leq t$ for each $t \in [a^*, M]$.

Proof. Let $a^* := a$, where $a = \lim_{m \rightarrow +\infty} a_m$, as above, let $E \in \Sigma$ be as in Remark E.9, let $A^* \in \Sigma$ be as in Remark E.10, and let $E^* := E \cap A^*$. Suppose that there exists some $s_1 \in E^*$ such that $X(s_1) \in [0, a^*)$, but $\mathcal{Y}^*(s_1) > 0$. Then for each $m \geq 1$ one has $\tilde{Y}_{\lambda^*, m, \mu}^*(s_1) > 0$, since the sequence $\{\tilde{Y}_{\lambda^*, m, \mu}^*\}_m$ converges monotonically downwards and pointwise on E^* to $\tilde{Y}_{\lambda^*, \mu}^*$ and $\mathcal{Y}^* \mathbf{1}_{E^*} = \tilde{Y}_{\lambda^*, \mu}^* \mathbf{1}_{E^*}$,

by definition of \mathcal{Y}^* . Consequently, $X(s_1) \geq a_m$, for each $m \geq 1$. Therefore, $X(s_1) \geq a^* = a = \lim_{m \rightarrow +\infty} a_m$, a contradiction. Hence, for each $s \in E^*$, $X(s) \in [0, a^*) \Rightarrow \mathcal{Y}^*(s) = 0$. Also, since $\mu(E) = \mu(A^*) = 1$, it follows from Lemma A.2 that $\mu(E^*) = 1$.

Moreover, $\tilde{Y}_{\lambda^*, \mu}^* = \tilde{I} \circ X$, for some bounded, nonnegative, nondecreasing, left-continuous, and Borel-measurable function \tilde{I} on the range $[0, M]$ of X (see Section B.1). Let $f := \tilde{I}$. One then has, for each $s \in E^*$, $\mathcal{Y}^*(s) = f(X(s))$ if $X(s) \in [a^*, M]$. Furthermore, since $0 \leq \tilde{Y}_{\lambda^*, \mu}^* \leq X$, it follows that $0 \leq f(t) \leq t$, for each $t \in [0, M]$. In particular, $f(0) = 0$. This completes the proof of Corollary 6.3. \square

Remark E.14. I have mentioned that the function f that appears in Corollary 6.3 can be characterized. Indeed, the optimal solution appearing in Corollary 6.3 has been constructed as the limit of a sequence of rearrangements of nonnegative functions (each bounded by $M := \|X\|_{sup}$). For ease of notation, I will refer to the initial sequence as $(k_m)_{m \geq 1}$, so that the sequence $(\tilde{k}_m)_{m \geq 1}$ of nondecreasing rearrangements converges to the optimal solution appearing in Corollary 6.3. Now each element k_m of the initial sequence can in turn be approximated by a nondecreasing sequence $\{l_{m,i}\}_{i \geq 1}$ of Σ -simple nonnegative functions. Moreover, Ghossoub [26] completely characterizes the nondecreasing rearrangement of simple functions, and hence it is possible to completely characterize the nondecreasing rearrangement $\tilde{l}_{m,i}$ of each one of these simple functions $l_{m,i}$. Each sequence $\{\tilde{l}_{m,i}\}_{i \geq 1}$ of nondecreasing simple functions hence obtained converges to the nondecreasing rearrangement \tilde{k}_m of each element k_m of the initial sequence $\{k_m\}_{m \geq 1}$ (by Lemma B.8).

APPENDIX F. SUFFICIENT CONDITIONS FOR $a^* > 0$

This section gives some sufficient conditions for the a^* appearing in Corollary 6.3 (eq. (6.2) on p. 18) – or Lemma E.13 – to be strictly positive. First, note that If there is some $n \geq 1$ such that $a_n > 0$, then $a > 0$ by Lemma E.12 – where a is defined in eq. (E.3) – and hence it follows from the definition of a^* that $a^* > 0$.

Lemma F.1. *There exists an event $E^* \in \Sigma$ such that $\mu(E^*) = 1$, and $a^* > 0$ when $\mu(\mathcal{D}_{E^*}) \neq 0$, where:*

- (i) $\mathcal{D}_{E^*} := \left\{ s_0 \in E^* : X(s_0) > 0, h(s_0) > 0, \int_{E^*} \mathcal{Y}^* h d\mu < \bar{L}(s_0) \right\}$; and,
- (ii) $\bar{L}(s_0) := \int_{E^*} \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} \left(\frac{u'(W_0 - \Pi - X(s_0))}{h(s_0)} h \right) \right] \right) \right] h d\mu$.

Finally, there exists $\kappa \in \mathbb{R}^+$ such that $a^* > 0$ when $\mu(\mathcal{E}_{E^*}) \neq 0$, where:

$$(F.1) \quad \mathcal{E}_{E^*} := \left\{ s_0 \in E^* : h(s_0) > 0, \kappa h(s_0) > u'(W_0 - \Pi), \right. \\ \left. 0 < X(s_0) < W_0 - \Pi - (u')^{-1}(\kappa h(s_0)) \right\}$$

Proof. Let $E \in \Sigma$ be as in Remark E.9, let $A^* \in \Sigma$ be as in Remark E.10, and let $E^* := E \cap A^*$, as above. Then $\mu(E^*) = 1$, by Lemma A.2. For each $s_0 \in E^*$, define $\bar{L}(s_0)$ by:

$$\bar{L}(s_0) := \int_{E^*} \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} \left(\frac{u'(W_0 - \Pi - X(s_0))}{h(s_0)} h \right) \right] \right) \right] h d\mu$$

Then

$$\bar{L}(s_0) = \int_A \min \left[X, \max \left(0, X - \left[W_0 - \Pi - (u')^{-1} \left(\frac{u'(W_0 - \Pi - X(s_0))}{h(s_0)} h \right) \right] \right) \right] h d\mu$$

Now, let

$$\mathcal{D}_{E^*} := \left\{ s_0 \in E^* : X(s_0) > 0, h(s_0) > 0, \int_{E^*} Y^* h d\mu < \bar{L}(s_0) \right\}$$

Then

$$\mathcal{D}_{E^*} = \left\{ s_0 \in E^* : X(s_0) > 0, h(s_0) > 0, \int_A Y^* h d\mu < \bar{L}(s_0) \right\}$$

Suppose that $\mu(\mathcal{D}_{E^*}) \neq 0$. Then, in particular, $\mathcal{D}_{E^*} \neq \emptyset$. Fix some $s_0 \in \mathcal{D}_{E^*}$. Then $X(s_0) > 0$, $h(s_0) > 0$, and $\int_A Y^* h d\mu < \bar{L}(s_0)$. In other words,

$$\beta^* = \phi(\lambda^*) = \int_A Y^* h d\mu < \phi \left(u'(W_0 - \Pi - X(s_0)) / h(s_0) \right) = \bar{L}(s_0).$$

Therefore, $\lambda^* \geq u'(W_0 - \Pi - X(s_0)) / h(s_0)$, since ϕ is a nonincreasing function. Consequently, $X(s_0) \leq W_0 - \Pi - (u')^{-1}(\lambda^* h(s_0))$, and so

$$Y_{\lambda^*}^*(s_0) = \min \left[X(s_0), \max \left(0, X(s_0) - \left[W_0 - \Pi - (u')^{-1}(\lambda^* h(s_0)) \right] \right) \right] = 0$$

Hence, for each $s_0 \in \mathcal{D}_{E^*}$, one has $X(s_0) > 0$ and $Y_{\lambda^*}^*(s_0) = 0$. Since $\mu(\mathcal{D}_{E^*}) \neq 0$ by hypothesis, it follows that

$$\mu \left(\left\{ s \in E^* : X(s) > 0, Y_{\lambda^*}^*(s) = 0 \right\} \right) \neq 0$$

Thus, the fact that in this case one has $a^* > 0$ follows from the properties of the equimeasurable rearrangement (recall equation (E.2) and the proof of Lemma E.8).

Now, let $\kappa = \lambda^*$, and define the set \mathcal{E}_{E^*} as follows:

$$\mathcal{E}_{E^*} := \left\{ s_0 \in E^* : h(s_0) > 0, \kappa h(s_0) > u'(W_0 - \Pi), \right. \\ \left. 0 < X(s_0) < W_0 - \Pi - (u')^{-1}(\kappa h(s_0)) \right\}$$

Suppose that $\mu(\mathcal{E}_{E^*}) \neq 0$. Then, in particular, $\mathcal{E}_{E^*} \neq \emptyset$. Fix some $s_0 \in \mathcal{E}_{E^*}$. Then $h(s_0) > 0$, $\lambda^* > u'(W_0 - \Pi) / h(s_0)$, $X(s_0) > 0$, and $X(s_0) < W_0 - \Pi - (u')^{-1}(\lambda^* h(s_0))$. Since the sequence $\{h_n\}_n$ of nonnegative, μ -simple functions on (S, Σ) previously defined converges pointwise to h , one can choose n large enough so that $h_n(s_0)$ is close enough to $h(s_0)$ and the following hold:

- (1) $h_n(s_0) > 0$;
- (2) $\lambda^* > u'(W_0 - \Pi) / h_n(s_0)$; and,

$$(3) \quad 0 < X(s_0) < W_0 - \Pi - (u')^{-1}(\lambda^* h_n(s_0)).$$

Therefore, from the proof of Lemma E.2 (third paragraph), one has $X(s_0) > 0$ and $Y_{\lambda^*,n}^*(s_0) = 0$. Since $\mu(\mathcal{E}_{E^*}) \neq 0$ by hypothesis, it follows that

$$\mu\left(\left\{s \in E^* : X(s) > 0, Y_{\lambda^*,n}^*(s) = 0, \text{ for some } n \geq 1\right\}\right) \neq 0$$

Thus, there exists $n^* \geq 1$ such that $\mu\left(\left\{s \in E^* : X(s) > 0, Y_{\lambda^*,n^*}^*(s) = 0\right\}\right) \neq 0$. For such n^* , one has $a_{n^*} > 0$ by properties of the equimeasurable rearrangement (as in the proof of Lemma E.8), and by definition of the function $\tilde{Y}_{\lambda^*,n^*,\mu}^*$ given in equation (E.1). This then yields $a > 0$ (by Lemma E.12) and so $a^* > 0$. \square

REFERENCES

- [1] C.D. ALIPRANTIS and K.C. BORDER. *Infinite Dimensional Analysis - 3rd edition*. Springer-Verlag, 2006.
- [2] R.J. ARNOTT and J.E. STIGLITZ. The Basic Analytics of Moral Hazard. *The Scandinavian Journal of Economics*, 90(3):383–413, 1988.
- [3] K.J. ARROW. *Essays in the Theory of Risk-Bearing*. Chicago: Markham Publishing Company, 1971.
- [4] R.J. AUMANN. Agreeing to Disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.
- [5] R.J. AUMANN. Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, 55(1):1–18, 1987.
- [6] R.J. AUMANN. Common Priors: A Reply to Gul. *Econometrica*, 66(4):929–938, 1998.
- [7] R.J. AUMANN and A. BRANDENBURGER. Epistemic Conditions for Nash Equilibrium. *Econometrica*, 63(5):1161–1180, 1995.
- [8] D. BERGEMANN and S. MORRIS. Robust Mechanism Design. *Econometrica*, 73(6):1771–1813, 2005.
- [9] D. BERGEMANN and S. MORRIS. Ex Post Implementation. *Games and Economic Behavior*, 63(2):527–566, 2008.
- [10] D. BERGEMANN and S. MORRIS. Robust Virtual Implementation. *Theoretical Economics*, 4(1):45–88, 2009.
- [11] D. BERGEMANN and S. MORRIS. Robust Implementation in General Mechanisms. *Games and Economic Behavior*, 71(2):261–281, 2011.
- [12] T.F. BEWLEY. Knightian Decision Theory. Part I. Cowles Foundation Discussion Papers, Paper no. 807 (1986).
- [13] V.I. BOGACHEV. *Measure Theory - Volume 1*. Springer-Verlag, 2007.
- [14] K.H. BORCH. *The Mathematical Theory of Insurance: An Annotated Selection of Papers on Insurance Published 1960-1972*. Lexington Books, 1974.
- [15] R. BRACEWELL. *The Fourier Transform and Its Applications, 3rd ed.* New York: McGraw-Hill, 2000.
- [16] A. CHATEAUNEUF, F. MACCHERONI, M. MARINACCI, and J.M. TALLON. Monotone Continuous Multiple Priors. *Economic Theory*, 26(4):973–982, 2005.
- [17] H. CHERNOFF and L.E. MOSES. *Elementary Decision Theory*. New York: Wiley, 1959.
- [18] K.S. CHUNG and J.C. ELY. Foundations of Dominant-Strategy Mechanisms. *The Review of Economic Studies*, 74(2):447–476, 2007.
- [19] D.L. COHN. *Measure Theory*. Birkhauser, 1980.
- [20] B. DE FINETTI. La Pr evision: Ses Lois Logiques, Ses Sources Subjectives. *Annales de l'Institut Henri Poincar e*, 7(1):1–68, 1937.
- [21] B. DE FINETTI. *Theory of Probability: A Critical Introductory Treatment (revised edition) - 1st ed. 1974*. New York: Wiley, 1990.
- [22] J. DIEUDONN E. *Foundations of Modern Analysis*. Academic press, 1969.
- [23] L. ECKHOUDT, C. GOLLIER, and H. SCHLESINGER. *Economic and Financial Decisions under Risk*. Princeton University Press, 2005.
- [24] P.C. FISHBURN. *Utility Theory for Decision Making*. John Wiley & Sons, 1970.
- [25] M. GHOSOUB. Monotone Equimeasurable Rarrangements with Non-Additive Probabilities. mimeo (2011).
- [26] M. GHOSOUB. *Contracting under Heterogeneous Beliefs*. PhD thesis, University of Waterloo – Dept. of Statistics & Actuarial Science, 2011.
- [27] I. GILBOA. *Theory of Decision under Uncertainty*. Cambridge University Press, 2009.

- [28] I. GILBOA, F. MACCHERONI, M. MARINACCI, and D. SCHMEIDLER. Objective and Subjective Rationality in a Multiple Prior Model. *Econometrica*, 78(2):755–770, 2010.
- [29] I. GILBOA and M. MARINACCI. Ambiguity and the Bayesian Paradigm. mimeo (2011).
- [30] I. GILBOA and D. SCHMEIDLER. Maxmin Expected Utility with a Non-Unique Prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.
- [31] F. GUL. A Comment on Aumann’s Bayesian View. *Econometrica*, 66(4):923–927, 1998.
- [32] J.Y. HALPERN and Y. MOSES. Knowledge and Common Knowledge in a Distributed Environment. *Journal of the ACM (JACM)*, 37(3):549–587, 1990.
- [33] G.H. HARDY, J.E. LITTLEWOOD, and G. PÓLYA. *Inequalities*. Cambridge University Press, Cambridge. Reprint of the 1952 edition, 1988.
- [34] J.C. HARSÁNYI. Games with Incomplete Information Played by ‘Bayesian’ Players, I-III. Part I. The Basic Model. *Management Science*, 14(3):159–182, 1967.
- [35] J.C. HARSÁNYI. Games with Incomplete Information Played by ‘Bayesian’ Players, I-III. Part II. Bayesian Equilibrium Points. *Management Science*, 14(5):320–334, 1968.
- [36] J.C. HARSÁNYI. Games with Incomplete Information Played by ‘Bayesian’ Players, I-III. Part III. The Basic Probability Distribution of the Game. *Management Science*, 14(7):486–502, 1968.
- [37] G. HUBERMAN, D. MAYERS, and C.W. SMITH Jr. Optimal Insurance Policy Indemnity Schedules. *The Bell Journal of Economics*, 14(2):415–426, 1983.
- [38] K. INADA. On a Two-Sector Model of Economic Growth: Comments and a Generalization. *The Review of Economic Studies*, 30(2):119–127, 1963.
- [39] M. JELEVA and B. VILLENEUVE. Insurance Contracts with Imprecise Probabilities and Adverse Selection. *Economic Theory*, 23(4):777–794, 2004.
- [40] E. LEHRER and D. SAMET. Agreeing to Agree. *Theoretical Economics*, 6(2):269–287, 2011.
- [41] J. MARSHALL. Optimum Insurance with Deviant Beliefs. in G. Dionne (ed.), *Contributions to Insurance Economics*. Boston: Kluwer Academic Publishers, pages 255–274, 1991.
- [42] P. MILGROM and N. STOKEY. Information, Trade and Common Knowledge. *Journal of Economic Theory*, 26(1):17–27, 1982.
- [43] S. MORRIS. The Common Prior Assumption in Economic Theory. *Economics and Philosophy*, 11(2):227–253, 1995.
- [44] F.P. RAMSEY. *The Foundations of Mathematics and other Logical Essays*. The Humanities Press: New York (Ed. R.B. Braithwaite), 1950.
- [45] A. RAVIV. The Design of an Optimal Insurance Policy. *The American Economic Review*, 69(1):84–96, 1979.
- [46] M. ROTHSCCHILD and J.E. STIGLITZ. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics*, 90(4):629–649, 1976.
- [47] W. RUDIN. *Principles of Mathematical Analysis – 3rd ed.* New York: McGraw-Hill Book Company, 1976.
- [48] S. SAKS. *Theory of the Integral*. 2nd ed. Monografie Matematyczne, Warsaw; English transl. L.C. Young, 1937.
- [49] L.J. SAVAGE. *The Foundations of Statistics (2nd revised edition) – 1st ed. 1954*. New York: Dover Publications, 1972.
- [50] D. SCHMEIDLER. Subjective Probability and Expected Utility without Additivity. *Econometrica*, 57(3):571–587, 1989.
- [51] J.K. SEBENIUS and J. GEANAKOPOLOS. Don’t Bet on It: Contingent Agreements with Asymmetric Information. *Journal of the American Statistical Association*, pages 424–426, 1983.
- [52] S. SHAVELL. On Moral Hazard and Insurance. *The Quarterly Journal of Economics*, 93(4):541–562, 1979.
- [53] J.E. STIGLITZ. Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies*, 44(3):407–430, 1977.
- [54] J.E. STIGLITZ. Risk, Incentives and Insurance: The Pure Theory of Moral Hazard. *The Geneva Papers on Risk and Insurance*, 8(1):4–33, 1983.
- [55] J. VON NEUMANN and O. MORGENSTERN. *Theory of Games and Economic Behavior*. Princeton University Press, 1947.
- [56] R. WILSON. Game-Theoretic Analyses of Trading Processes. In T. Bewley (ed.), *Advances in Economic Theory: Fifth World Congress*, pages 33–70. Cambridge: Cambridge University Press, 1987.

UNIVERSITÉ DE MONTRÉAL – DÉPARTEMENT DE SCIENCES ÉCONOMIQUES – C.P. 6128, SUCCURSALE CENTRE-VILLE – MONTRÉAL, QC, H3C 3J7 – CANADA

E-mail address: mario.ghossoub@umontreal.ca