Variance Insurance Contracts

Yichun Chi[*]   Xun Yu Zhou[†]   Sheng Chao Zhuang[‡]

Abstract

We study the design of an optimal insurance contract in which the insured maximizes her expected utility and the insurer limits the variance of his risk exposure while maintaining the principle of indemnity and charging the premium according to the expected value principle. We derive the optimal policy semi-analytically, which is coinsurance above a deductible when the variance bound is binding. This policy automatically satisfies the incentive-compatible condition, which is crucial to rule out ex post moral hazard. We also find that the deductible is absent if and only if the contract pricing is actuarially fair. Focusing on the actuarially fair case, we carry out comparative statics on the effects of the insured's initial wealth and the variance bound on insurance demand. Our results indicate that the expected coverage is always larger for a wealthier insured, implying that the underlying insurance is a *normal* good, which supports certain recent empirical findings. Moreover, as the variance constraint tightens, the prudent insured cedes less losses, while the insurer is exposed to less tail risk.

JEL codes: C02, G22

Keywords: Insurance design; expected value principle; variance; incentive compatibility; comparative statics.

---

[*]China Institute for Actuarial Science, Central University of Finance and Economics, Beijing 102206, China. Email: yichun@cufe.edu.cn.

[†]Department of Industrial Engineering and Operations Research and The Data Science Institute, Columbia University, New York, NY 10027, USA. Email: xz2574@columbia.edu.

[‡]Department of Finance, University of Nebraska-Lincoln, NE, USA. Email: szhuang3@unl.edu.

1   Introduction

Insurance is an important mechanism to facilitate risk reallocation between two parties. Borch (1960) was the first to study the insurance contract design problem and to prove that given a fixed premium, a stop-loss (or deductible) insurance policy (i.e., full coverage above a deductible) achieves the smallest variance of the insured's share of payment.[1] Arrow (1963) instead considers to maximize the expected utility of the final wealth of a risk-averse insured, and obtains a similar result in favor of stop-loss insurance when the insurance premium depends only on the expected indemnity. This is a foundational result that has earned the name of *Arrow's theorem of the deductible* in the literature. Mossin (1968), Arrow (1974) and Schlesinger (1981) further prove that the deductible is strictly positive if and only if the insurance price is actuarially unfair (i.e., the safety loading is strictly positive).

In the literature, the insurance premium is often assumed to be calculated by only using the expected indemnity. See, for example, Mossin (1968), Arrow (1974), Gollier (1996), Bernard et al. (2015) and Xu et al. (2019). This is based on the assumption that the insurer undertakes a sufficiently large number of independent and homogenous policies, such that his risk, by the law of large numbers, is sufficiently diversified to be nearly zero. Considering risk-averse insurers becomes relevant in situations when this diversification is not possible, e.g. when there are catastrophic risks or when designing tailor-made contracts for insuring one-off events (e.g., the shipment of a highly valuable painting). Proposition 2 in Arrow (1963) stipulates that, when the insurer is risk-averse and insurance cost is absent, an optimal contract must involve coinsurance. Raviv (1979) extends this result to include nonlinear insurance costs and shows that an optimal policy involves both a deductible and coinsurance. Much of the recent research in this area has focused on the insurer's tail risk exposure. Cummins and Mahul (2004) and Zhou et al. (2010) extend Arrow's model by introducing an *exogenous* upper bound on indemnity and thereby limiting the insurer's liability with respect to catastrophic losses. From a regulatory perspective, Zhou and Wu (2008) propose a model in which the insurer's expected loss above a prescribed level is controlled, and they conclude that an optimal policy is generally piecewise linear. Doherty et al. (2015) investigate the case in which losses are nonverifiable and deduce that a contract with a deductible and an *endogenous* upper limit is optimal.

As important as tail risks are for both parties in insurance contracts, in practice insurers are also concerned with other parts of the loss distribution. Kaye (2005), in a Casualty Actuarial Society report, writes

---

[1]Borch (1960) presents the problem in a reinsurance setting, in which the ceding insurer corresponds to the insured in an insurance setting.

"Different stakeholders have different levels of interest in different parts of the distribution - the perspective of the decision-maker is important. Regulators and rating agencies will be focused on the extreme downside where the very existence of the company is in doubt. On the other hand, management and investors will have a greater interest in more near-term scenarios towards the middle of the distribution and will focus on the likelihood of making a profit as well as a loss" (p. 4).

Assuming the insurer to be risk-averse with a concave utility function indeed takes the whole risk distribution into consideration, as studied in Arrow (1971) and Raviv (1979). However, one can rarely obtain, analytically, optimal policies with a general utility, hindering post-optimality analyses such as comparative statics. For instance, when the premium is a fixed constant, Raviv (1979) derives a differential equation satisfied by the optimal indemnity, which takes a rather complex form depending on the utility function used.

By contrast, variance, as a measure of risk originally put forth in Markowitz's pioneering work (Markowitz, 1952), is also related to the whole distribution, yet it is more intuitive and transparent. Moreover, if the number of insureds is not sufficiently large, then variance would be a significant risk for the insurer according to the law of large numbers. Therefore, Borch (1960) designs a contract that aims to minimize the variance of the insurer's retained risk. Kaluszka (2001) generalizes Borch (1960)'s work by incorporating a variance-related premium principle and shows that the optimal contract to minimize the variance of the insured's payment can be stop-loss, quota share (i.e., the insurer covers a *constant* proportion of the loss) or change-loss (i.e., a combination of stop-loss and quota share). For a fixed premium, Kaluszka (2005) extends the study of Arrow (1974) by considering variance-related premium principles and provides a characterization of the optimal contracts under some restrictive assumptions. Vajda (1962) studies the problem from the *insurer's* perspective, and shows that a quota share policy minimizes the variance of indemnity in an actuarially fair contract. However, his result depends critically on restricting the admissible contracts to be such that the *ratio* between indemnity and loss increases as the loss increases, a feature that enables the derivation of a solution through rather simple calculus.[2]

In this paper, we revisit the work of Arrow (1974) by assuming the expected value principle for premium but imposing a variance constraint on the *insurer*'s risk exposure.[3] Unlike Vajda (1962),

---

[2]Vajda (1962) claims that this feature "agrees with the spirit of (re)insurance, at least in most cases" (p. 259). However, for a larger loss, it is indeed in the spirit of insurance that the insurer should pay more, but it is not clear why he should be responsible for a higher *proportion*. Interestingly, our results will show that the optimal policies of our model possess this property *if* the insured is prudent; see Corollary 3.9.

[3]Problems with higher moment constraints are widely studied in the portfolio optimization literature; refer to Mandal and Thakur (2023) for a comprehensive review on this topic. In particular, a recent study of Boudt et al. (2022) derives an optimal payoff for an investor under a variance constraint. However, surprisingly, higher moment constraints have been seldom considered in the design of optimal insurance policies, despite their importance in controlling the insurance

we consider the general actuarially unfair case and remove the restriction that the proportion of the insurer's payment increases with the size of the loss. Because the expected value premium principle ensures the expected profit of the insurer, our model is essentially a *mean–variance* model *à la* Markowitz for the *insurer*. We stress, and indeed will show, that our formulation is *different* from the Pareto formulation with a variance premium principle, nor is it a special case of Raviv (1979)'s model with a quadratic utility function for the insurer.

The presence of the variance constraint causes substantial technical challenges in solving the problem. In the literature, there are generally two approaches used to study variants of Arrow's model: those involving sample-wise optimization and stochastic orders. However, the former fails to work for our problem due to the nonlinearity of the variance constraint, and the latter is not readily applicable either because the presence of the variance constraint invalidates the claim that any admissible contract is dominated by a stop-loss one. The first contribution of this paper is methodological: we develop a new approach by *combining* the techniques of stochastic orders, calculus of variations and the Lagrangian duality to derive optimal insurance policies. The solutions are semi-analytical in the sense that they can be computed by solving some *algebraic* equations.

Our second contribution is actuarial: we show that the optimal contract is coinsurance above a deductible when the variance constraint is binding. Moreover, the deductible disappears if and only if the insurance price is actuarially fair. These results are *qualitatively* similar to those of Raviv (1979) that uses a concave utility function for the insurer. A natural question is why one would bother to study the mean–variance version of a problem that would generate contracts with similar characteristics to its expected utility counterpart beyond the reason that, as pointed out earlier, the former is not a special case of the latter. This question can be answered in the same way as in the field of financial portfolio selection, where there is an enormously large body of study on the Markowitz mean–variance model along with its popularity in practice, *despite* the existence of the equally well-studied expected utility maximization models. In other words, expected utility and mean–variance are two *different* frameworks, and, as argued earlier, the latter underlines a more transparent and explicit return–risk tradeoff, which usually leads to explicit solutions.

Our optimal policies involve coinsurance, which is widely utilized in the insurance industry. As pointed out by Arrow (1963) and Raviv (1979), risk aversion on the part of the insurer could be a cause for coinsurance, but other attributes such as the nonlinearity of the insurance cost function could also lead to coinsurance. Another explanation for coinsurance is to mitigate the moral hazard risk due to asymmetric information; see Hölmstrom (1979) and Dionne and St-Michel (1991). From the insured's perspective, Doherty and Schlesinger (1990) argue that default risk of the insurer can motivate the insured to choose coinsurance. Picard (2000) also shows that coinsurance is optimal in order to reduce the risk premium paid to the auditor. In this paper, we prove that optimal policies

business risk (Kaye, 2005).

4

can turn from full insurance to coinsurance as the variance bound tightens, thereby providing a novel yet simple reason for the prevalent feature of coinsurance in insurance theory and practice: a variance bound on the insurer's risk exposure.

Intriguingly, our optimal insurance polices *automatically* satisfy the so-called *incentive-compatible* condition that *both* the insured and the insurer pay more for a larger loss (or, equivalently, the marginal indemnity is between 0 and 1).[4] In Arrow (1963)'s setting, the optimal contract turns out to be incentive-compatible; however, this is generally untrue. Gollier (1996) considers an insured facing an additional background risk that is not insurable. Under the expected value principle, he discovers that the optimal insurance, which relies heavily on the dependence between the background risk and the loss, may render the marginal indemnity strictly larger than 1. Bernard et al. (2015) generalize the insured's risk preference from expected utility to rank-dependent utility involving probability distortion (weighting), and also find that the optimal indemnity may decrease when the loss increases. In both of these papers, the derived optimal contracts would incentivize the insured to misreport the actual losses, leading to ex post moral hazard. Equally absurd would be the case in which the insurer pays less for a larger loss. To address this issue, Huberman et al. (1983) propose the incentive-compatible condition as a *hard* constraint on admissible insurance policies, in addition to the *principle of indemnity* (i.e., the insurer's reimbursement is non-negative and smaller than the loss). Xu et al. (2019) add this constraint to the model of Bernard et al. (2015), painstakingly developing a completely different approach in order to overcome the difficulty arising out of this additional constraint and deriving qualitatively very different contracts. On the other hand, Raviv (1979) discovers that his optimal solution is incentive-compatible, assuming that the loss has a strictly positive probability density function. Carlier and Dana (2005) use a Hardy–Littlewood rearrangement argument to prove that any optimal contract is dominated by an incentive-compatible contract, establishing the optimality of the latter. However, both these studies assume that the loss is non-atomic, which rules out the important and practical case in which the loss is atomic at 0. By contrast, in the presence of the variance constraint, we show that the optimal policy is naturally incentive-compatible even without the corresponding hard constraint or the assumption of an atom-less loss.

Our final contribution is a comparative statics analysis examining the respective effects of the insured's wealth and the variance constraint on insurance demand under actuarially fair pricing. Our results indicate that the presence of a variance bound fundamentally changes the insurance strategy

---

[4]In economics, the term "incentive-compatible condition" has different meanings depending on the context. For instance, in the asymmetric information literature, the term is on the constraint specifying that the agent takes the decision which is best for his/her self-interest (Hölmstrom, 1979). In our context, such a condition removes the incentive for the insured to misreport the true level of the loss after its occurrence (see e.g., Huberman et al., 1983; Cai et al., 2017; Asimit et al., 2020; Jiang et al., 2021; Birghila et al., 2023; Cheung et al., 2023). The incentive-compatible condition discussed in this paper is also termed the *no-sabotage* condition in Carlier and Dana (2003).

– it makes the insured's wealth relevant and it changes the way in which the two parties share the risk. In particular, *ceteris paribus,* the expected coverage is always *larger* for a wealthier insured who has strictly decreasing absolute prudence (DAP), rendering the insurance product a *normal good*. This finding provides some theoretical foundation for the empirical observations of Millo (2016) and Armantier et al. (2018). Moreover, we show that the insurer has less downside risk when contracting with a wealthier insured with strictly DAP.[5] On the other hand, we establish that the variance bound significantly changes a prudent insured's risk transfer decision – she would consistently transfer more losses as the variance bound loosens. A corollary of this result is, rather surprisingly, that the insurer can reduce the tail risk by simply tightening the variance constraint. This suggests that our variance contracts do, *after all*, address the issue of tail exposure.

The rest of the paper proceeds as follows. In Section 2 we formulate the problem, compare it with Pareto formulations and present some preliminaries about risk preferences. In Section 3 we develop the solution approach and present the optimal insurance contracts. In Section 4 we conduct a comparative analysis by examining the effects of the insured's initial wealth and the variance constraint on insurance demand. Section 5 concludes the paper. Some auxiliary results and all proofs are relegated to the appendices.

2   Problem Formulation and Preliminaries

2.1   Variance contracts formulation

An insured endowed with an initial wealth w_0 faces an insurable loss X, which is a non-negative, essentially bounded random variable defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with the cumulative distribution function (c.d.f.) $F_X(x) := \mathbb{P}\{X \leqslant x\}$ and the essential supremum $\mathcal{M} < \infty$. An insurance contract design problem is to partition X into two parts, $I(X)$ and $X - I(X)$, where $I(X)$ (the *indemnity*) is the portion of the loss that is ceded to the insurer ("he") and $R_I(X) := X - I(X)$ (the *retention*) is the portion borne by the insured ("she"). I and R_I are also called the insured's ceded and retained loss functions, respectively. It is natural to require a contract to satisfy the *principle of indemnity*, namely the indemnity is non-negative and less than the amount of loss. Thus, the feasible set of indemnity functions is

$$\mathfrak{C} := \{I : 0 \leqslant I(x) \leqslant x, \forall x \in [0, \mathcal{M}]\}.$$

As the insurer covers part of the loss for the insured, he is compensated by collecting the premium from her. We assume that the insurer calculates the premium using the *expected value principle* that is widely used in insurance economics (see e.g., Mossin, 1968; Arrow, 1974; Gollier, 1996; Bernard et al.,

---

[5]Menezes et al. (1980) introduce the notion of downside risk to compare two risks with the same mean and variance. The formal definition is given in Section 2.3.

2015; Xu et al., 2019); namely, the premium on making a non-negative random payment Y is charged as

$$\pi(Y) = (1 + \rho)\mathbb{E}[Y]$$

where $\rho \geqslant 0$ is the so-called *safety loading* coefficient. His risk exposure under a contract I for a loss X is hence

$$e_I(X) = I(X) - \pi(I(X)).$$

The insurer may evaluate this risk using different measures for different purposes, as Kaye (2005) notes. In this paper, we assume that the insurer has sufficient regulatory capital and therefore focuses on the volatility of the underwriting risk. Specifically, he uses the variance to measure the risk and requires

$$var[e_I(X)] \equiv var[I(X)] \leqslant \nu$$

for some prescribed $\nu > 0$.

On the other hand, denote by $W_I(X)$ the insured's final wealth under contract I upon its expiration, namely,

$$W_I(X) = w_0 - X + I(X) - \pi(I(X)).$$

The insured's risk preference is characterized by a von Neumann–Morgenstern utility function U satisfying $U' > 0$ and $U'' < 0$.

Our optimal contracting problem is, therefore,

$$
\begin{aligned}
\max_{I \in \mathfrak{C}} \quad & \mathbb{E}[U(W_I(X))] \\
\text{subject to} \quad & var\,[e_I(X)] \leqslant \nu.
\end{aligned}
\tag{2.1}
$$

Note that this model reduces to the classical Arrow (1963, 1974)'s model

$$\max_{I \in \mathfrak{C}} \quad \mathbb{E}[U(W_I(X))]$$

by setting the upper bound ν to be $\mathbb{E}[X^2]$. This is because $var[e_I(X)] = var\,[I(X)] \leqslant \mathbb{E}[X^2]$ for all $I \in \mathfrak{C}$.

In Problem (2.1), the insured's benefit–risk consideration is captured by the utility function U, whereas the insurer's return–risk tradeoff is reflected by the "mean" (the expected value principle) and the "variance" (the variance bound). One may interpret the problem as one faced by an insurer who likes to design a contract with the best interest of a representative insured in mind, so as to remain marketable and competitive, while maintaining the desired profitability and risk control in the mean–variance sense.[6] Problem (2.1) can also model a tailor-made contract design for insuring a

---

[6]Representative insureds in different wealth classes or different regions may have different "typical" levels of initial wealth. Moreover, when the economy grows, a representative insured's initial wealth may change substantially. As shown in Subsection 4.1, the change in the insured's initial wealth may affect her demand for insurance.

one-off event from an insured's perspective. The insured aims to maximize her expected utility while accommodating the insurer's participation constraint reflected by the mean and variance specifications.

2.2 Comparison to Pareto formulations

Our model follows the setting of Arrow (1974) with an additional variance constraint. In the insurance contracting literature it is also standard to use Pareto formulations. In this subsection we discuss the connection and difference between these formulations.

For a Pareto efficiency model, an insurance contract is a pair (P, I), where $P \in \mathbb{R}_+$ is a premium and $I \in \mathfrak{C}$ is an indemnity function. As before, the insured has a concave utility U and initial endowment w_0, and evaluates the insurance contract via

$$U(P, I) := \mathbb{E}[U(w_0 - X + I(X) - P)].$$

We assume that the insurance cost $c(I(X))$ is proportional to the indemnity, namely, $c(I(X)) = \rho I(X)$ for some $\rho \geq 0$. Then, the insurer's profit from the contract (P, I) can be determined by

$$Z = P - (1 + \rho)I(X).$$

Suppose the insurer has mean-variance preference of the form

$$V(Z) = \mathbb{E}[Z] - \frac{\gamma}{2}var[Z],$$

where γ is the insurer's risk-aversion coefficient. This gives rise to the following objective for the insurer

$$\begin{aligned}
\pi(P, I) :=& \mathbb{E}[P - (1 + \rho)I(X)] - \frac{\gamma}{2}var[P - (1 + \rho)I(X)] \\
=& P - (1 + \rho)\mathbb{E}[I(X)] - \frac{\gamma}{2}(1 + \rho)^2 var[I(X)].
\end{aligned}$$

A contract (P, I) is Pareto-efficient (PE) if there is no other contract (P', I') such that $U(P', I') \geqslant U(P, I)$ and $\pi(P', I') \geqslant \pi(P, I)$, with at least one strict inequality. It is well known that a contract (P, I) is PE if and only if there exists $k \in \mathbb{R}_+$ such that (P, I) is a solution of the constrained optimization problem (see e.g., Carlier and Dana, 2003, 2005)

$$\begin{aligned}
\max_{(P,I) \in \mathbb{R}_+ \times \mathfrak{C}} \quad & U(P, I) \\
\text{subject to} \quad & \pi(P, I) \geqslant k.
\end{aligned} \tag{2.2}$$

For any given I, since $\pi(P, I)$ is increasing and $U(P, I)$ is decreasing in P, the inequality constraint in (2.2) must be binding at optimality. Thus we only need to solve the following problem

$$\max_{I \in \mathfrak{C}} \quad \mathbb{E}\left[U\left(w_0 - k - (1 + \rho)\mathbb{E}[I(X)] - \frac{\gamma}{2}(1 + \rho)^2 var[I(X)] - X + I(X)\right)\right]. \tag{2.3}$$

If Problem (2.3) has a solution I_k^* for each $k \in \mathbb{R}_+$, then a PE contract is

$$\left(k + (1+\rho)\mathbb{E}[I_k^*(X)] + \frac{\gamma}{2}(1+\rho)^2 var[I_k^*(X)], I_k^*\right).$$

As the constant k varies in $\mathbb{R}_+$, we obtain the collection of all possible PE contracts.[7]

Problem (2.3) with $k = 0$ can be viewed as an insurance design problem in the Arrow (1963, 1974)'s setting in which the insured has expected utility preference and the insurer charges premium according to the *mean–variance premium principle*:

$$P = (1+\rho)\mathbb{E}[I(X)] + \frac{\gamma}{2}(1+\rho)^2 var[I(X)].$$

In contrast, our problem (2.1) can be rewritten as

$$\max_{I \in \mathfrak{C}} \quad \mathbb{E}\Big[U\big(w_0 - (1+\rho)\mathbb{E}[I(X)] - X + I(X)\big)\Big]$$
$$\text{subject to} \quad var[I(X)] \leq \nu.$$

When ν is sufficiently small (see Assumption 3.1), the results of this paper will indicate that the constraint above is always binding at optimality. So the problem is equivalent to one in which the variance constraint is an equality and, hence, further equivalent to the following by virtue of introducing a Lagrange multiplier β:

$$\max_{I \in \mathfrak{C}} \quad \mathbb{E}\Big[U\big(w_0 - (1+\rho)\mathbb{E}[I(X)] - X + I(X)\big)\Big] - \beta\left[var[I(X)] - \nu\right]. \tag{2.4}$$

Now we can see the clear difference between (2.3) with $k = 0$ and (2.4): the variance term $var[I(X)]$ is "inside" of U in the former while "outside" in the latter. The two problems are different, *unless* the insured is risk-neutral and U is a linear function.

There have been studies in literature solving some variants of Problem (2.3). For instance, when the insured's objective is measured by some convex function of the retained risk and the insurer takes the mean-variance premium principle, Kaluszka (2004, 2005) provide characterizing conditions of the optimal indemnity, without explicit forms of the solutions. To our best knowledge, solving (2.3) thoroughly, even if in some "semi-analytical" form, remains an open question.

Next, we show that our problem is *not* a special case of the problem studied in Raviv (1979) by simply taking the insurer's utility as a quadratic function. Specifically, Raviv (1979) formulates a PE problem with the insured and the insurer having the concave utility functions U and V respectively, which is equivalent to the following problem

$$\max_{(P,I) \in \mathbb{R}_+ \times \mathfrak{C}} \quad \mathbb{E}[U(w_0 - X + I(X) - P)]$$
$$\text{subject to} \quad \mathbb{E}\big[V(w_1 - I(X) - c(I(X)) + P)\big] \geqslant k. \tag{2.5}$$

---

[7]Technically, if we can solve Problem (2.3) with $k = 0$, then we can also solve its $k \neq 0$ counterpart, because k can be absorbed into w_0.

Here, w_1 is the insurer's initial wealth, $c(\cdot)$ is the insurance cost function with

$$c(0) = a \geqslant 0, \quad c' \geqslant 0, \quad c'' \geqslant 0,$$

and k is a constant such that $k \geqslant V(w_1)$. Now, take $c(x) = \rho x$ and $V(x) = x - \zeta x^2$, where $0 < \zeta < \frac{1}{2(w_1 + \mathcal{M})}$ to ensure that V is concave and increasing.[8] As before, the constraint in (2.5) is binding at optimality; thus it becomes a quadratic equation in P given that $V(x) = x - \zeta x^2$, which we can solve as $P = g(k, w_1, \mathbb{E}[I(X)], var[I(X)])$ for some nonlinear function g. Therefore, in this case Raviv's problem reduces to

$$\max_{I \in \mathfrak{C}} \quad \mathbb{E}\left[U\left(w_0 - g(k, w_1, \mathbb{E}[I(X)], var[I(X)]) - X + I(X) \right) \right]. \tag{2.6}$$

This problem is even more complex than Problem (2.3) and certainly does not include our problem (2.4) as a special case.

2.3 Absolute risk aversion and prudence

The Arrow–Pratt measure of absolute risk aversion (Pratt, 1964; Arrow, 1965), defined as

$$\mathcal{A}(x) := -\frac{U''(x)}{U'(x)},$$

captures the dependence of the level of risk aversion on the agent's wealth x. If $\mathcal{A}(x)$ is decreasing[9] in x, then the insured's risk preference is said to exhibit *decreasing absolute risk aversion* (DARA). The effect of an insured's initial wealth on the insurance demand under Arrow (1974)'s model has been widely studied in the literature. It is found that a wealthier DARA insured purchases a deductible insurance with a higher deductible. For a survey on how insureds' wealth impacts insurance, see e.g., Gollier (2001, 2013).

While risk aversion ($U'' < 0$) captures an insured's propensity for avoiding risk, *prudence* (i.e., $U''' > 0$) reflects her tendency to take precautions against future risk. Many commonly used utility functions, including those with hyperbolic absolute risk aversion (HARA) and mixed risk aversion, are prudent.[10] Based on an experiment with a large number of subjects, Noussair et al. (2014) observe that

---

[8]Note that $P \leqslant \mathcal{M}$; otherwise there is no incentive for the insured to buy the insurance. Thus $\zeta < \frac{1}{2(w_1 + \mathcal{M})}$ implies $V' \geqslant 0$.

[9]Throughout the paper, the terms "increasing" and "decreasing" mean "non-decreasing" and "non-increasing," respectively.

[10]A utility function is called HARA if the reciprocal of the Arrow–Pratt measure of absolute risk aversion is a linear function, i.e.,

$$\mathcal{A}_U(x) = -\frac{U''(x)}{U'(x)} = \frac{1}{px + q}$$

for some $p \geqslant 0$ and q. It includes exponential, logarithmic and power utility functions as special cases. For further discussion of HARA, see Gollier (2001). A utility function is said to be of mixed risk aversion if $(-1)^n U^{(n)}(x) \leqslant 0$ for all x and $n = 1, 2, 3, \cdots$, where $U^{(n)}$ denotes the nth derivative of U.

the majority of individuals' decisions are consistent with prudence. Eeckhoudt and Kimball (1992) and Gollier (1996) take into account the insured's prudence in designing optimal insurance policies. The degree of absolute prudence is defined as

$$\mathcal{P}(x) := -\frac{U'''(x)}{U''(x)} \tag{2.7}$$

for a three-time differentiable utility function U. If $\mathcal{P}(x)$ is strictly decreasing in x, then the insured is said to exhibit strictly *decreasing absolute prudence* (DAP). Kimball (1990) shows that DAP characterizes the notion that wealthier people are less sensitive to future risks. Moreover, DAP implies DARA, as noted in Proposition 21 of Gollier (2001).

A term related to prudence is *third-degree stochastic dominance* (TSD), which was introduced by Whitmore (1970). A non-negative random variable Z_1 is said to dominate another non-negative random variable Z_2 in TSD if

$$\mathbb{E}[Z_1] \geqslant \mathbb{E}[Z_2] \quad \text{and} \quad \int_0^x \int_0^y F_{Z_2}(z) - F_{Z_1}(z) \mathrm{d}z \mathrm{d}y \geqslant 0 \text{ for all } x \geqslant 0.$$

Equivalently, Z_1 dominates Z_2 in TSD if and only if $\mathbb{E}[u(Z_1)] \geqslant \mathbb{E}[u(Z_2)]$ for all functions u satisfying $u' > 0$, $u'' < 0$ and $u''' > 0$. TSD has been widely employed for decision making in finance and insurance. For instance, Gotoh and Konno (2000) use it to study mean-variance optimal portfolio problems. If Z_1 dominates Z_2 in TSD and they have the same mean and variance, then Z_1 is said to have less downside risk than Z_2. In fact, the latter is equivalent to $\mathbb{E}[u(Z_1)] \geqslant \mathbb{E}[u(Z_2)]$ for any function u with $u''' > 0$; see Menezes et al. (1980).

3 Optimal Contracts

In this section, we present our approach to solving Problem (2.1).

First, consider Problem (2.1) *without* the variance constraint:

$$\max_{I \in \mathfrak{C}} \quad \mathbb{E}[U(W_I(X))]. \tag{3.1}$$

This is the classical Arrow (1963, 1974)'s model, for which the optimal contract is a deductible one of the form $(x - d^*)_+$ for some non-negative deductible d^*, where $(x)_+ := \max\{x, 0\}$. This contract *automatically* satisfies the incentive-compatible condition. Moreover, Chi (2019) (see Theorem 4.2 therein) was the first to derive an *analytical* form of the optimal deductible level d^*. More precisely, define

$$VaR_{\frac{1}{1+\rho}}(X) := \inf \left\{ x \in [0, \mathcal{M}] : F_X(x) \geqslant \frac{\rho}{1+\rho} \right\}$$

and

$$\varphi(d) := \frac{\mathbb{E}[U'(W_{(x-d)_+}(X))]}{U'(w_0 - d - \pi((X-d)_+))}, \quad 0 \leqslant d < \mathcal{M},$$

11

where $\inf \emptyset := \mathcal{M}$ by convention. Then the optimal d^* is

$$d^* = \sup\left\{ VaR_{\frac{1}{1+\rho}}(X) \leqslant d < \mathcal{M} : \varphi(d) \geqslant \frac{1}{1+\rho} \right\} \vee VaR_{\frac{1}{1+\rho}}(X), \qquad (3.2)$$

where $\sup \emptyset := 0$ and $x \vee y := \max\{x, y\}$.[11] This leads immediately to the following proposition.

Proposition 3.1. *If $\nu \geqslant var[(X - d^*)_+]$, then $I(x) = (x - d^*)_+$ is the optimal solution to Problem* (2.1).

Intuitively, if the variance bound ν is set sufficiently high, then the variance constraint in Problem (2.1) is redundant and the problem reduces to the classical Arrow (1963, 1974)'s problem. Proposition 3.1 tells exactly and explicitly what the bound should be for the variance constraint to be active.

Therefore, it suffices to solve Problem (2.1) for the case in which $\nu < var[(X - d^*)_+]$, which we now set as an assumption.

Assumption 3.1. *The variance bound ν satisfies $\nu < var[(X - d^*)_+]$.*

The main thrust for finding the solution is to first restrict the analysis with a fixed level of expected indemnity and then find the optimal level of expected indemnity. To this end, we need to first identify the range in which the optimal expected indemnity possibly lies. Noting that $var[(X - d)_+]$ is strictly decreasing and continuous in d over $[\text{ess inf } X, \mathcal{M})$, we define

$$d_L := \inf\{d \geqslant d^* : var[(X - d)_+] \leqslant \nu\} \quad \text{and} \quad m_L := \mathbb{E}[(X - d_L)_+].$$

Intuitively, the insurer would demand a deductible higher than Arrow's level d^* due to the additional risk control reflected by the variance constraint, and d_L is the smallest deductible that makes this constraint binding.

Lemma 3.2. *Under Assumption 3.1, for Problem* (2.1)*, any admissible insurance policy I with $\mathbb{E}[I(X)] \leqslant m_L$ is no better than the deductible contract I_L where $I_L(x) = (x - d_L)_+$.*

Therefore, we can rule out any contract whose expected indemnity is strictly smaller than m_L; in other words, m_L is a *lower* bound of the optimal expected indemnity. In particular, no-insurance (i.e., $I^*(x) \equiv 0$) is *never* optimal under Assumption 3.1.

Next, we are to derive an *upper* bound of the optimal expected indemnity. Consider a loss-capped contract $X \wedge k$, where $x \wedge y := \min\{x, y\}$ and $k \geqslant 0$, which pays the actual loss up to the cap k.[12] Define

$$K_U := \inf\{k \geqslant 0 : var[X \wedge k] \geqslant \nu\} \quad \text{and} \quad m_U := \mathbb{E}[X \wedge K_U].$$

---

[11]The number d^* can be numerically computed easily, because $\varphi(d)$ is decreasing over $[VaR_{\frac{1}{1+\rho}}(X), \mathcal{M})$; see Chi (2019).

[12]A loss-capped contract is also called "full insurance up to a (policy) limit" or "full insurance with a cap."

In the above, K_U is well-defined because $X \wedge k - \mathbb{E}[X \wedge k]$ is increasing in k in the sense of convex order, according to Lemma A.2 in Chi (2012).[13] Clearly, both K_U and m_U depend on the variance bound ν. Since $var[X] \geqslant var[(X - d^*)_+] > \nu$, we have

$$K_U < \mathcal{M}, \quad m_U < \mathbb{E}[X] \quad \text{and} \quad var\left[X \wedge K_U\right] = \nu.$$

Lemma 3.3. *For any $I \in \mathfrak{C}$ with $var[I(X)] \leqslant \nu$, we must have $\mathbb{E}[I(X)] \leqslant m_U$. Moreover, if $I \in \mathfrak{C}$ satisfies $var[I(X)] \leqslant \nu$ and $\mathbb{E}[I(X)] = m_U$, then $I(X) = X \wedge K_U$ almost surely.*

This lemma stipulates that m_U is an upper bound of the optimal expected indemnity. Moreover, any admissible contract achieving this upper bound is equivalent to the loss-capped contract $X \wedge K_U$. An immediate corollary of the lemma is $m_L \leqslant m_U$, noting that $var[(X - d_L)_+] = \nu$.

The following result identifies the case $m_L = m_U$ as a trivial one.

Proposition 3.4. *If $m_L = m_U$, then the loss X must follow a Bernoulli distribution with values 0 and $d_L + K_U$. Moreover, under Assumption 3.1, the optimal contract of Problem (2.1) is*

$$I^*(0) = 0 \quad \text{and} \quad I^*(d_L + K_U) = K_U.$$

In what follows, we consider the general and interesting case in which $m_L < m_U$. For $m \in (m_L, m_U)$, define

$$\mathfrak{C}_m := \{I \in \mathfrak{C} : var[I(X)] \leqslant \nu, \, \mathbb{E}[I(X)] = m\}.$$

We now focus on the following optimization problem

$$\max_{I \in \mathfrak{C}_m} \quad \mathbb{E}[U(W_I(X))], \tag{3.3}$$

which is a "cross section" of the original problem (2.1) where the expected indemnity is fixed as m.

For $\lambda \in \mathbb{R}$ and $\beta \geqslant 0$, denote

$$I_{\lambda,\beta}(x) := \sup\left\{y \in [0, x] : U'\left(w_0 - x + y - (1 + \rho)m\right) - \lambda - 2\beta y \geqslant 0\right\}, \, x \in [0, \mathcal{M}]. \tag{3.4}$$

Actually, $I_{\lambda,\beta}$ is a contract that coinsures above a deductible or coinsures following full insurance, depending on the relative values between λ and $U'(w_0 - (1 + \rho)m)$. To see this, when $\lambda \geqslant U'(w_0 - (1 + \rho)m)$, we have

$$I_{\lambda,\beta}(x) = \begin{cases} 0, & 0 \leqslant x \leqslant w_0 - (1 + \rho)m - (U')^{-1}(\lambda), \\ f_{\lambda,\beta}(x), & w_0 - (1 + \rho)m - (U')^{-1}(\lambda) < x \leqslant \mathcal{M}, \end{cases} \tag{3.5}$$

---

[13] A random variable Y is said to be greater than a random variable Z in the sense of convex order, denoted as $Z \leqslant_{cx} Y$, if

$$\mathbb{E}[Y] = \mathbb{E}[Z] \text{ and } \mathbb{E}[(Z - d)_+] \leqslant \mathbb{E}[(Y - d)_+] \,\,\forall d \in \mathbb{R},$$

provided that the expectations exist. Obviously, $Z \leqslant_{cx} Y$ implies $var[Z] \leqslant var[Y]$.

and when $\lambda < U'(w_0 - (1 + \rho)m)$, we have

$$I_{\lambda,\beta}(x) = \begin{cases} x, & 0 \leqslant x \leqslant \frac{U'(w_0-(1+\rho)m)-\lambda}{2\beta}, \\ f_{\lambda,\beta}(x), & \frac{U'(w_0-(1+\rho)m)-\lambda}{2\beta} < x \leqslant \mathcal{M}, \end{cases} \tag{3.6}$$

where $f_{\lambda,\beta}(x)$ satisfies the following equation in y:[14]

$$U'(w_0 - x + y - (1 + \rho)m) - \lambda - 2\beta y = 0. \tag{3.7}$$

Moreover, it is easy to see that $0 \leqslant f_{\lambda,\beta}(x) \leqslant x$ either when $\lambda \geqslant U'(w_0-(1+\rho)m)$ and $w_0-(1+\rho)m-(U')^{-1}(\lambda) < x \leqslant \mathcal{M}$, or when $\lambda < U'(w_0 - (1 + \rho)m)$ and $\frac{U'(w_0-(1+\rho)m)-\lambda}{2\beta} < x \leqslant \mathcal{M}$. Furthermore,

$$f'_{\lambda,\beta}(x) = \frac{-U''(w_0 - x + f_{\lambda,\beta}(x) - (1 + \rho)m)}{2\beta - U''(w_0 - x + f_{\lambda,\beta}(x) - (1 + \rho)m)} \in [0, 1]. \tag{3.8}$$

The following result indicates that there exists an optimal solution to Problem (3.3) that is in the form of $I_{\lambda,\beta}(x)$ and binds both the mean and the variance constraints.

Proposition 3.5. *Suppose Assumption 3.1 holds and $m_L < m_U$. Then there exist $\lambda_m^* \in \mathbb{R}$ and $\beta_m^* > 0$ such that $I_{\lambda_m^*,\beta_m^*}$ satisfies*

$$\mathbb{E}[I_{\lambda_m^*,\beta_m^*}(X)] = m \qquad and \qquad var[I_{\lambda_m^*,\beta_m^*}(X)] = \nu, \tag{3.9}$$

and is an optimal solution to Problem (3.3).

Combining Lemma 3.2, Lemma 3.3 and Proposition 3.5 yields that we can always find an optimal contract in one of the following three types: a deductible one of the form $I_L(x) = (x - d_L)_+$, a loss-capped one of the form $I_U(x) = x \wedge K_U$ and a general one of the form $I_{\lambda_m^*,\beta_m^*}(x)$. In other words, the optimal solutions of the following maximization problem

$$\max_{I \in \left\{I_L,\ I_U,\ I_{\lambda_m^*,\beta_m^*} \text{ for } m \in (m_L, m_U)\right\}} \mathbb{E}[U(W_I(X))], \tag{3.10}$$

where $m_L < m_U$, also solve Problem (2.1).

Note that I_L, I_U and $I_{\lambda_m^*,\beta_m^*}$ all satisfy the incentive-compatible condition (see (3.8)); hence, so does at least one of the optimal contracts I^* of Problem (2.1). That is, $I^*(0) = 0$ and $0 \leqslant I^{*\prime}(x) \leqslant 1$ almost everywhere.[15] Therefore, it suffices to solve the following maximization problem

$$\begin{aligned} \max_{I \in \mathcal{IC}} \quad & \mathbb{E}[U(W_I(X))] \\ \text{subject to} \quad & var\left[e_I(X)\right] \leqslant \nu, \end{aligned} \tag{3.11}$$

---

[14]It can be shown easily that this equation has a unique solution.

[15]As will be evident in the sequel, the values of I' on a set with zero Lebesgue measure have no impact on I. Therefore, we will often omit the phrase "almost everywhere" in statements regarding the marginal indemnity function I' throughout this paper.

where

$$\mathcal{IC} := \left\{ I : I(0) = 0,\ 0 \leqslant I'(x) \leqslant 1,\ \forall x \in [0, \mathcal{M}] \right\} \subsetneq \mathfrak{C}, \tag{3.12}$$

to obtain an optimal contract for Problem (2.1).

Notice that $\mathcal{IC}$ is convex on which $\mathbb{E}[U(W_I(X))]$ is *strictly* concave. Using the convex property of variance and applying arguments similar to those in the proof of Proposition 3.1 in Chi and Wei (2020), we obtain the following proposition:

Proposition 3.6. (i) *There exist optimal solutions to Problem* (3.11).

(ii) *Assume either $\rho > 0$ or $\mathbb{P}\{X < \epsilon\} > 0$ for all $\epsilon > 0$. Then there exists a unique solution to Problem* (3.11) *in the sense that $I_1(X) = I_2(X)$ almost surely for any two solutions I_1 and I_2.*

Note that the assumptions in Proposition 3.6-(ii) are satisfied in most situations of practical interest because either an insurer naturally sets a positive safety loading, or a loss actually never occurs with a positive probability, or both happen. On the other hand, since any optimal solution to Problem (3.11) also solves Problem (2.1), Proposition 3.6-(i) establishes the existence of optimal solutions to the latter.[16] Moreover, the argument proving Proposition 3.1 in Chi and Wei (2020) can be used to show that Proposition 3.6-(ii) holds true for Problem (2.1) as well. Finally, now that we have the existence and uniqueness of the optimal solutions for both Problems (3.11) and (2.1), we conclude that these two problems are indeed *equivalent* under the assumptions of Proposition 3.6-(ii).

While the analysis of Problem (2.1) is simplified to Problem (3.10), it remains challenging to solve this problem because λ_m^* and β_m^* are implicit functions of m. Before solving this problem, we introduce a useful result that provides a general qualitative structure for the optimal indemnity function in Problem (3.11) or, equivalently, Problem (2.1).

Proposition 3.7. *Under Assumption 3.1, if I^* is a solution to Problem* (3.11), *then there exists $\beta^* > 0$ such that*

$$I^{*\prime}(x) = \begin{cases} 1, & \Phi_{I^*}(x) > 0, \\ c_{I^*}(x), & \Phi_{I^*}(x) = 0, \\ 0, & \Phi_{I^*}(x) < 0, \end{cases} \tag{3.13}$$

for some function c_{I^*} bounded on $[0, 1]$, where

$$\Phi_I(x) := \mathbb{E}\left[U'(W_I(X)) - 2\beta^* I(X) \,\middle|\, X > x\right] - \left((1 + \rho)\mathbb{E}\left[U'(W_I(X))\right] - 2\beta^* \mathbb{E}[I(X)]\right) \tag{3.14}$$

for $x \in [0, \mathcal{M})$ and $I \in \mathcal{IC}$.

---

[16]It is difficult to prove the existence of solutions to Problem (2.1) *directly* because its feasible set is not compact only under the principle of indemnity.

Note that (3.13) does not entail an *explicit* expression of $I^{*\prime}$ because its right hand side also depends on I^* as well as on an unknown parameter β^*. While deriving the optimal solution I^* *directly* from (3.13) seems challenging, the equation reveals the important property that $I^{*\prime}$ must take a value of either 0 or 1, except at point(s) x where $\Phi_{I^*}(x) = 0$.[17] This property will in turn help us to decide whether the optimal contract is of the form $(x - d_L)_+$, $x \wedge K_U$, or $I_{\lambda_m^*, \beta_m^*}$.

The following theorem presents a complete solution to Problem (2.1).

Theorem 3.8. *Under Assumption 3.1 and assuming that the c.d.f. F_X is strictly increasing on $(0, \mathcal{M})$, we have the following conclusions:*

(i) *If $\rho = 0$, then the optimal indemnity function is I^*, where $I^*(x)$ solves the following equation in y for all $x \in (0, \mathcal{M}]$:*

$$U'(w_0 - x + y - m^*) - 2\beta^* y - U'(w_0 - m^*) = 0, \ \ y \in (0, x), \tag{3.15}$$

with the parameters $m^* \in (m_L, m_U)$ and $\beta^* > 0$ determined by

$$\mathbb{E}[I^*(X)] = m^* \quad and \quad var[I^*(X)] = \nu. \tag{3.16}$$

(ii) *If $\rho > 0$, then the optimal indemnity function is*

$$I^*(x) = \begin{cases} 0, & 0 \leqslant x \leqslant \tilde{d}, \\ f^*(x), & \tilde{d} < x \leqslant \mathcal{M}, \end{cases} \tag{3.17}$$

where $f^*(x)$ satisfies $f^*(\tilde{d}) = 0$ and solves the following equation in y:

$$U'(w_0 - (1 + \rho)m^* - x + y) - U'(w_0 - (1 + \rho)m^* - \tilde{d}) \tag{3.18}$$
$$= \frac{y}{m^* \rho} \left(U'(w_0 - (1 + \rho)m^* - \tilde{d}) - (1 + \rho)\mathbb{E}[U'(w_0 - (1 + \rho)m^* - X \wedge \tilde{d})] \right), \ y \in (0, x),$$

and $\tilde{d} \in (VaR_{\frac{1}{1+\rho}}(X), \mathcal{M})$ and $m^* \in (m_L, m_U)$ are determined by (3.16).

Theorem 3.8 provides a thorough solution to Problem (2.1) under Assumption 3.1. It indicates that the optimal contract can *not* be a pure deductible of the form $(x - d_L)_+$, nor a pure loss-capped of the form $x \wedge K_U$. It can only be in the form $I_{\lambda_m^*, \beta_m^*}$ of (3.5) (rather than (3.6)). The optimal policies can be computed by solving a system of three *algebraic* equations; so the result is semi-analytic. The result also characterizes the mean–variance efficient frontier for the insurer:

$$\{((1 + \rho)\mathbb{E}[(X - d^*)_+], var[(X - d^*)_+])\} \cup \{((1 + \rho)\mathbb{E}[I^*(X)], \nu) : \nu < var[(X - d^*)_+]\}.$$

---

[17]From the control theory perspective, (3.13) corresponds to an optimal control problem in which I' is taken as the control variable. Moreover, the optimal control turns out to be of the so-called "bang-bang" type, whose values depend on the sign of the discriminant function Φ_I. This type of optimal control problem arises when the Hamiltonian depends linearly on control and the control is constrained between an upper bound and a lower bound. It is usually hard to solve for optimal control when the discriminant function is complex, which is the case here.

Actuarially, Theorem 3.8 reveals how the variance bound impacts the contract. When the bound ν is sufficiently low so that it is binding (hence the model does not degenerate into the classical Arrow's model), the optimal policy is always genuine coinsurance if there is no safety loading. Here, by "genuine" we mean the *strict* inequalities $0 < I^*(x) < x$ for all $x \in (0, \mathcal{M}]$, namely both the insurer and the insured pay positive portions of the loss incurred. When the safety loading coefficient is positive, the optimal contract demands genuine coinsurance above a positive deductible. So the variance bound translates into a change from the part of the full insurance in Arrow's contract to coinsurance. Our contracts are similar *qualitatively* to those of Raviv (1979), in which a utility function is in the place of a variance bound; however, ours are *quantitatively* different from Raviv (1979)'s.

On the other hand, the deductible $\tilde{d}$ is positive if and only if the safety loading coefficient is positive. So the existence of the deductible is completely determined by the loading coefficient in the insurance premium. This result is consistent with the findings in Mossin (1968), Arrow (1974) and Schlesinger (1981).

Corollary 3.9. *Under the assumptions of Theorem 3.8, if the insured is prudent, then the proportion between optimal indemnity and loss increases as the loss increases.*

So, with a prudent insured, the insurer pays more not only absolutely but also relatively as loss increases. Vajda (1962) restricts his study on a variance contracting problem to policies that have this feature of the insurer covering proportionally more for larger losses. Corollary 3.9 uncovers ex post this feature in our optimal policies, *provided* that the insured is prudent.

4   Comparative Statics

Thanks to the semi-analytic results derived in the previous section, we are able to analyze the impacts of the insured's initial wealth and the variance bound on the insurance demand.

We make the following assumptions for our comparative statics analysis:

Assumption 4.1.   (i) F_X *is strictly increasing on* $(0, \mathcal{M})$.

(ii) *The insurance is fairly priced, i.e.,* $\rho = 0$.

Assumption 4.1-(i) is standard in the literature that accommodates most of the used distributions by actuaries, such as (truncated) exponential, lognormal, gamma, and Pareto distributions (Bahnemann, 2015). Assumption 4.1-(ii) is not necessarily plausible in practice, but it is meaningful in theory, as it describes a state in competitive equilibrium, in which insurers break even and insurance policies are actuarially fair for representative insureds (see e.g., Rothschild and Stiglitz, 1976; Viscusi, 1979). It is important to carry out comparative statics analyses in such a "fair" state in order to rule out

any impact emanating from an unfair price. Such an assumption is indeed often imposed when conducting comparative statics in the literature of insurance economics. For example, the comparative statics results of Ehrlich and Becker (1972) and Viscusi (1979) deal exclusively with actuarially fair situations. Many recent studies, such as Eeckhoudt et al. (2003), Huang and Tzeng (2006) and Teh (2017), also impose this assumption for their comparative statics analyses.

Finally, we will assume $\nu < var[X]$ throughout this section, as otherwise the variance constraint is redundant and the optimal solution is trivially full insurance.

4.1 Impact of the insured's initial wealth

In this subsection we examine the impact of the insured's initial wealth on insurance demand. We first recall the notion of one function up-crossing another. A function g_1 is said to *up-cross* a function g_2, both defined on $\mathbb{R}$, if there exists $z_0 \in \mathbb{R}$ such that

$$
\begin{cases}
g_1(x) \leqslant g_2(x), & x < z_0, \\
g_1(x) \geqslant g_2(x), & x \geqslant z_0.
\end{cases}
$$

Moreover, g_1 is said to up-cross g_2 twice if there exist $z_0 < z_1$ such that

$$
\begin{cases}
g_1(x) \leqslant g_2(x), & x < z_0, \\
g_1(x) \geqslant g_2(x), & z_0 \leqslant x < z_1, \\
g_1(x) \leqslant g_2(x), & x \geqslant z_1.
\end{cases}
$$

Consider two initial wealth levels $w_1 < w_2$ and denote the corresponding optimal contracts by I_1^* and I_2^* and the associated parameters by β_1^* and β_2^*, respectively, which are determined by Theorem 3.8. Recall that $\rho = 0$, so the insurer's risk exposure functions are

$$
e_{I_i^*}(x) = I_i^*(x) - \mathbb{E}[I_i^*(X)] = I_i^*(x) - m_i^*, \quad i = 1, 2, \tag{4.1}
$$

where $m_i^* := \mathbb{E}[I_i^*(X)]$. Taking expectations on (3.15) yields

$$
U'(w_i - m_i^*) = \mathbb{E}[U'(w_i - X + e_{I_i^*}(X))] - 2\beta_i^* m_i^*,
$$

which in turn implies, for $i = 1, 2$,

$$
U'(w_i - x + e_{I_i^*}(x)) - 2\beta_i^* e_{I_i^*}(x) - \mathbb{E}[U'(w_i - X + e_{I_i^*}(X))] = 0, \tag{4.2}
$$

$$
\mathbb{E}[e_{I_i^*}(X)] = 0 \quad \text{and} \quad \mathbb{E}[(e_{I_i^*}(X))^2] = var[I_i^*(X)] = \nu. \tag{4.3}
$$

Note that the insurer's profit with the contract I_i is $\mathbb{E}[I_i^*(X)] - I_i^*(X) \equiv -e_{I_i^*}(X)$, $i = 1, 2$. The following theorem establishes the impact of the initial wealth on the insurance contract.

18

Theorem 4.1. *In addition to Assumption 4.1, we assume that $\nu < var[X]$ and the insured's utility function U exhibits strictly DAP. Then, the insurer's risk exposure function with the larger initial wealth, $e_{I_2^*}(x)$, up-crosses the risk exposure function with the smaller initial wealth, $e_{I_1^*}(x)$, twice. Moreover, the insurer's profit, $-e_{I_2^*}(X)$, has less downside risk when contracting with the wealthier insured.*[18]
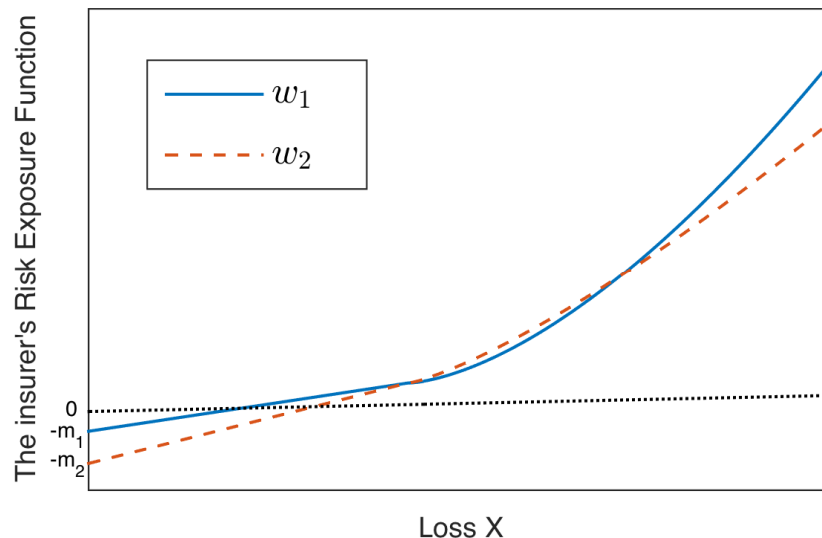

Figure 4.1: A schematic comparison of two insurer's risk exposure functions with $w_1 < w_2$.

Figure 4.1 illustrates graphically the first part of Theorem 4.1. The actuarial implication is that when the insured becomes wealthier, *ceteris paribus,* the insurer's risk exposure is lower for large or small losses and is higher for moderate losses. This can be explained intuitively as follows. Even if the insurance pricing is actuarially fair, the insureds are unable to transfer all the risk to the insurer due to the variance bound. However, the wealthier insured is more tolerant with large losses due to the DAP; hence, the insurer's risk exposure is lower for large losses when contracting with the wealthier insured. Due to the requirement that the insurer's *expected* risk exposure be always zero, the insurer's risk exposure with the wealthier insured must be higher for moderate losses. Now, should the insurer's risk exposure with the wealthier insured also be higher for small losses, then overall the insurer's risk exposure with the wealthier insured would be strictly more spread out than that with the less wealthy one, leading to a smaller variance of the former (see Lemmas A.3 and B.3), which

---

[18]Although the result obtained in this theorem is somewhat similar to Theorem 3.2 in Chi et al. (2020), there are significant differences in the model settings. Under a fixed premium, Chi et al. (2020) analyze the effect of the cedent's wealth on the optimal contract subject to the constraints on the first two moments of indemnity. In this paper, we instead assume that the insurance premium is calculated by the net premium principle, and analyze this effect only with a variance constraint.

would be a contradiction. Hence, the insurer's risk exposure must be *lower* for small losses with the wealthier insured.

The second part of the theorem, on the other hand, suggests that a variance minding insurer prefers to provide insurance to a wealthier insured due to the smaller downside risk.

Corollary 4.2. *Under the assumptions of Theorem 4.1, we have the following conclusions:*

(i) $\mathbb{E}[I_1^*(X)] < \mathbb{E}[I_2^*(X)]$ *and* $\beta_2^* < \beta_1^*$.

(ii) *Either* $I_1^*(x) < I_2^*(x) \; \forall x > 0$, *or* I_1^* *up-crosses* I_2^*.
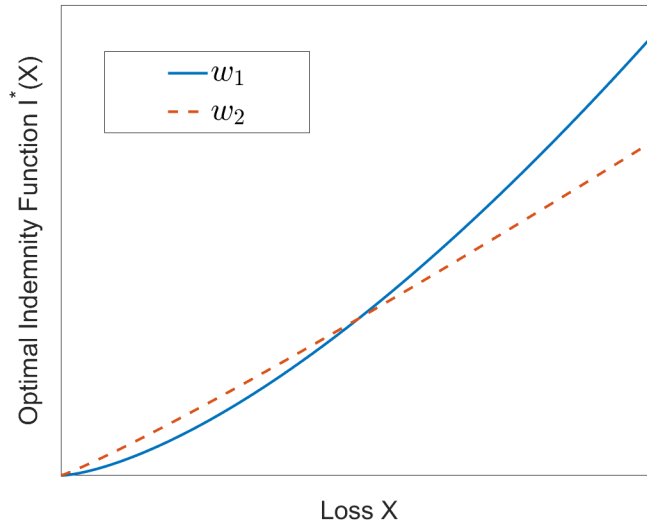

Figure 4.2: A schematic comparison of two optimal indemnity functions with $w_1 < w_2$.

In part (ii) of this corollary, while the case $I_1^*(x) < I_2^*(x)$ for any $x > 0$ is a special case of I_1^* up-crossing I_2^*, we state it separately to highlight its possibility. In the classical Arrow (1963, 1974)'s model, full insurance is optimal when insurance pricing is actuarially fair. This conclusion is *independent* of an insured's wealth. Zhou et al. (2010) show that such a conclusion is intact when there is an *exogenous* upper limit imposed on the insurer's risk exposure. Our result yields that adding a variance bound fundamentally changes the insurance demand – it makes the insured's wealth level relevant, and it changes the way in which the two parties share the risk. Specifically, Corollary 4.2 suggests that a DAP wealthier insured would either demand more coverage across the board or retain more larger risk and cede more smaller risk (see Figure 4.2). Either way, the *expected* coverage is *always* larger for the wealthier insured. Recall that insurance is called a *normal (inferior) good* if wealthier people purchase more (less) insurance converge; see Mossin (1968), Schlesinger (1981) and Gollier (2001). Millo (2016) argues that nonlife insurance is a normal good by empirically testing whether or not income elasticity is significantly greater than one. Armantier et al. (2018) use micro

level survey data on households' insurance coverage to conclude that insurance is a normal good, thereby providing a better understanding of the relationship between insurance demand and economic development. These studies, however, are purely empirical. To the best of our knowledge, ours is the first *theoretical* result regarding insurance as a normal good under the insurer's variance constraint, confirming these empirical findings.[19]

4.2 Impact of the variance bound

In this subsection we keep the insured's initial wealth unchanged and analyze the impact of the variance bound on her demand for insurance. Consider two variance bounds with $0 < \nu_1 < \nu_2 < var[X]$ and denote the corresponding optimal indemnity functions by I_1^* and I_2^* and the parameters by β_1^* and β_2^*, respectively. Thus, the insurer's risk exposures, $e_{I_i^*}(x) = I_i^*(x) - \mathbb{E}[I_i^*(X)]$, $i = 1, 2$, satisfy

$$U'(w_0 - x + e_{I_i^*}(x)) - 2\beta_i^* e_{I_i^*}(x) - \mathbb{E}[U'(w_0 - X + e_{I_i^*}(X))] = 0 \tag{4.4}$$

and

$$\mathbb{E}[e_{I_i^*}(X)] = 0 \quad \text{and} \quad \mathbb{E}[(e_{I_i^*}(X))^2] = var[e_{I_i^*}(X)] = \nu_i. \tag{4.5}$$

The following theorem illustrates how the insurer's risk exposure responds to the change in the variance bound.

Theorem 4.3. *Under Assumption 4.1, the insurer's risk exposure function with the larger variance bound, $e_{I_2^*}$, up-crosses that with the smaller variance bound, $e_{I_1^*}$.*

Under fair insurance pricing, this theorem indicates that, *ceteris paribus,* as the variance bound decreases, the insurer is exposed to less risk for a larger X and to more risk for a smaller X. This result has a rather significant implication in terms of the insurer's tail risk management. A variance constraint by its very definition does not control the tail risk *directly.* However, Theorem 4.3 suggests that the insurer can reduce the risk exposure for larger losses simply by tightening the variance constraint.[20] This further justifies our formulation of the variance contracting model.

Corollary 4.4. *Under the assumption of Theorem 4.3, for any $0 < \nu_1 < \nu_2 < var[X]$, we have the following conclusions:*

(i) $\mathbb{E}[I_1^*(X)] < \mathbb{E}[I_2^*(X)]$ *and* $\beta_2^* < \beta_1^*$;

(ii) *If the insured's utility function satisfies* $U''' \geqslant 0$, *then* $I_1^*(x) < I_2^*(x) \ \forall x > 0$.

---

[19]Mossin (1968), Schlesinger (1981) and Gollier (2001) show that a wealthier insured with a DARA preference will cede less risk under *unfair* insurance pricing; hence, insurance is an inferior good in the corresponding economy. Their results degenerate into full insurance when the pricing is fair, and thus insurance demand is independent of the insured's wealth. Consequently, our results do *not* contradict theirs.

[20]It follows from Lemma A.3 that the insurer with a more relaxed variance constraint suffers more underwriting risk in the sense of convex order, i.e., $e_{I_1^*}(X) \leqslant_{cx} e_{I_2^*}(X)$.
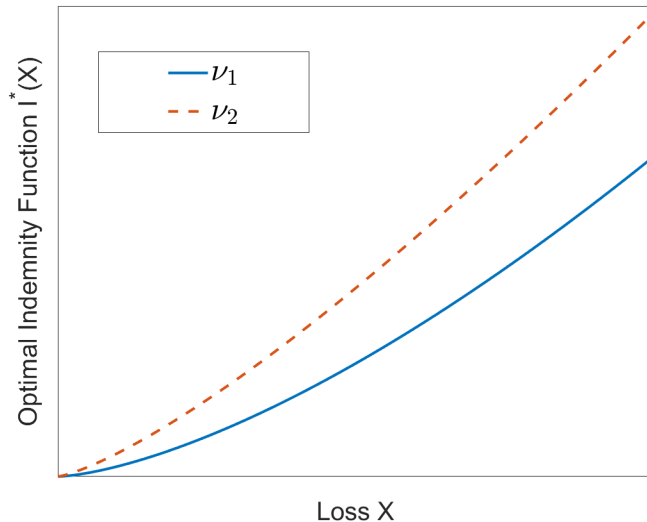
Figure 4.3: A schematic comparison of two optimal indemnity functions with $\nu_1 < \nu_2$.

Corollary 4.4-(i) can be easily interpreted: an insurer with a tighter variance bound offers less expected coverage. As a complement to Theorem 4.3, Corollary 4.4-(ii) establishes a direct characterization of the insured's optimal risk transfer with regard to the change in the variance bound: A prudent insured consistently cedes more losses when the variance bound increases (see Figure 4.3). In other words, if the insurance contract is priced fairly, the insured will transfer as much risk to the insurer as the latter's risk tolerance allows.

5   Concluding Remarks

In this paper, we have revisited the classical Arrow (1963, 1974)'s model by adding a variance limit on the insurer's risk exposure. This constraint is motivated by the insurer's desire to manage underwriting risk; at the same time, it poses considerable technical challenges for solving the problem. We have developed an approach to derive optimal contracts semi-analytically, in the form of coinsurance above a deductible when the variance constraint is active. The final policies automatically satisfy the incentive-compatible condition, thereby eliminating potential ex post moral hazard. We have also conducted a comparative analysis to examine the impact of the insured's wealth and of the variance bound on insurance demand.

Although the model we propose here appears rather ad hoc, and is not in the realm of Pareto efficiency, our results can actually explain some prevalent provisions in insurance policies such as deductible and coinsurance. More importantly, we have shown that insurers can use the variance bound to control the tail risk. This in turn suggests that variance may be useful in designing insurance products or computing capital requirements in insurance practice and regulation, even if it has not

yet been done or, indeed, realized.

This work can be extended in a few directions. First, we have restricted the comparative analysis to actuarially fair insurance. Analyzing for the general unfair case calls for a different approach than the one presented here. Second, although the expected value principle is used in this paper, the main results of Section 3 can be generalized to the premium setting $\pi(Y) = h(\mathbb{E}[Y])$ for any non-negative random variable Y, where $h(\cdot)$ is an increasing convex function with $h(y) \geqslant y$ for any $y \in \mathbb{R}_+$. Third, a model incorporating probability distortion (weighting) is of significant interest, both theoretically and practically. This is because probability distortion, a phenomenon well documented in psychology and behavioral economics, is related to tail events, about which both insurers and insureds have great concerns.

Acknowledgments

Appendices

A   Stochastic Orders

Since the notion of stochastic orders plays an important role in this paper, we present, in this appendix, some useful results in this regard.

A random variable Y is said to be greater than a random variable Z in the sense of stop-loss order, denoted as $Z \leqslant_{sl} Y$, if

$$\mathbb{E}[(Z - d)_+] \leqslant \mathbb{E}[(Y - d)_+] \ \ \forall d \in \mathbb{R},$$

provided that the expectations exist. It follows readily that Y is greater than Z in convex order (i.e., $Z \leqslant_{cx} Y$), if $\mathbb{E}[Y] = \mathbb{E}[Z]$ and $Z \leqslant_{sl} Y$.

A useful way to verify the stop-loss order is the well–known Karlin–Novikoff criterion (Karlin and Novikoff 1963).

Lemma A.1. *Suppose $\mathbb{E}[Z] \leqslant \mathbb{E}[Y] < \infty$. If F_Z up-crosses F_Y, then $Z \leqslant_{sl} Y$.*

If $Z \leqslant_{sl} Y$, then $\mathbb{E}[g(Z)] \leqslant \mathbb{E}[g(Y)]$ holds for all the increasing convex functions g, provided that the expectations exist. Based on the Karlin–Novikoff criterion, Gollier and Schlesinger (1996) obtain the following lemma.

Lemma A.2. *For any $h \in \mathfrak{C}$, we have $X \wedge d \leqslant_{cx} h(X)$, where $d \in [0, \mathcal{M}]$ satisfies $\mathbb{E}[X \wedge d] = \mathbb{E}[h(X)]$.*

The following result with respect to convex order is from Lemma 3 of Ohlin (1969).

Lemma A.3. *Let Y be a random variable and h_i, $i = 1, 2$, be two increasing functions with $\mathbb{E}[h_1(Y)] = \mathbb{E}[h_2(Y)]$. If h_1 up-crosses h_2, then $h_2(Y) \leqslant_{cx} h_1(Y)$.*

B Other Useful Lemmas

This appendix presents some other technical results that are useful in connection with this paper.

It is easy to verify that any sequence of indemnity functions in $\mathcal{IC}$ is uniformly bounded and equicontinuous over $[0, \mathcal{M}]$. Hence, the Arzéla-Ascoli theorem implies

Lemma B.1. *The set $\mathcal{IC}$ is compact under the norm $d(I_1, I_2) = \max_{t \in [0, \mathcal{M}]} | I_1(t) - I_2(t) |$, $I_1, I_2 \in \mathcal{IC}$.*

For the following lemma, one can refer to Komiya (1988) for a proof.

Lemma B.2. *(Sion's Minimax Theorem) Let Y be a compact convex subset of a linear topological space and Z a convex subset of a linear topological space. If Γ is a real-valued function on $Y \times Z$ such that $\Gamma(y, \cdot)$ is continuous and concave on Z for any $y \in Y$ and $\Gamma(\cdot, z)$ is continuous and convex on Y for any $z \in Z$, then*

$$\min_{y \in Y} \max_{z \in Z} \Gamma(y, z) = \max_{z \in Z} \min_{y \in Y} \Gamma(y, z).$$

The following lemmas are needed in the comparative analysis.

Lemma B.3. *If a non-negative increasing function h_1 up-crosses a non-negative increasing function h_2 with $\mathbb{E}[(h_1(X))^2] = \mathbb{E}[(h_2(X))^2]$, then either $h_1(X)$ and $h_2(X)$ have the same distribution or $\mathbb{E}[h_1(X)] < \mathbb{E}[h_2(X)]$.*

Proof. If $\mathbb{E}[h_2(X)] \leqslant \mathbb{E}[h_1(X)]$, then Lemma A.1 implies $h_2(X) \leqslant_{sl} h_1(X)$. Moreover, we have

$$\mathbb{E}[(h_i(X))^2] = 2 \int_0^\infty \mathbb{E}[(h_i(X) - t)_+] \mathrm{d}t.$$

Since it is assumed that $\mathbb{E}[(h_1(X))^2] = \mathbb{E}[(h_2(X))^2]$, we must have $\mathbb{E}[(h_1(X) - t)_+] = \mathbb{E}[(h_2(X) - t)_+]$ for any $t \geqslant 0$. It then follows from the equation $\mathbb{E}[(h_i(X) - t)_+] = \int_t^\infty (1 - F_{h_i(X)}(y)) \mathrm{d}y$ that $h_1(X)$ and $h_2(X)$ have the same distribution. $\qquad\square$

Lemma B.4. *Under the assumption of Theorem 4.1, it is impossible that $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ have the same distribution, and it is impossible either $e_{I_1^*}$ up-crosses $e_{I_2^*}$ or $e_{I_2^*}$ up-crosses $e_{I_1^*}$, where $e_{I_i^*}$ is given in (4.1).*

Proof. First of all, we show that $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ cannot have the same distribution. Define

$$\phi(z) := \frac{-U''(w_1 - z)}{-U''(w_2 - z)}, \quad 0 \leqslant z < w_1 < w_2. \tag{B.1}$$

A direct calculation based on the assumption of strictly DAP shows that

$$\phi'(z) = (\mathcal{P}(w_1 - z) - \mathcal{P}(w_2 - z))\phi(z) > 0,$$

where $\mathcal{P}$ is defined in (2.7). As a result, ϕ is a strictly increasing function.

We now prove the result by contradiction. Assume that $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ are equal in distribution. Noting that $e_{I_i^*}$ is increasing and Lipschitz-continuous and that $F_X(x)$ is strictly increasing, we have

$$e_{I_1^*}(x) = e_{I_2^*}(x), \ \forall x \in [0, \mathcal{M}],$$

which in turn implies $e'_{I_1^*}(x) = e'_{I_2^*}(x)$. It follows from (4.2) that $e'_{I_i^*}(x) = \frac{1}{1 + 2\beta_i^*/(-U''(w_i - x + e_{I_i^*}(x)))}$ for all $x \geqslant 0$; hence,

$$\frac{2\beta_1^*}{-U''(w_1 - x + e_{I_1^*}(x))} = \frac{2\beta_2^*}{-U''(w_2 - x + e_{I_2^*}(x))}.$$

Because $e_{I_1^*}(x) = e_{I_2^*}(x)$ for all $x \in [0, \mathcal{M}]$, we obtain

$$\phi(x - e_{I_1^*}(x)) = \frac{-U''(w_1 - x + e_{I_1^*}(x))}{-U''(w_2 - x + e_{I_1^*}(x))} = \frac{\beta_1^*}{\beta_2^*}, \ \forall x \in [0, \mathcal{M}],$$

which contradicts the fact that $x - e_{I_1^*}(x) \equiv x - I_1^*(x) + \mathbb{E}[I_1^*(X)]$ is strictly increasing in $x \in [0, \mathcal{M}]$ and that ϕ is a strictly increasing function.

Next we show that it is impossible that $e_{I_1^*}$ up-crosses $e_{I_2^*}$. Again we prove the result by contradiction. Since $e_{I_1^*}$ is not always non-negative, we introduce the increasing function

$$\tilde{f}_i^*(x) := e_{I_i^*}(x) + \widetilde{m}, \ \ i = 1, 2,$$

where $\widetilde{m} := \max\{\mathbb{E}[I_1^*(X)], \mathbb{E}[I_2^*(X)]\}$. It then follows that $\tilde{f}_1^*$ up-crosses $\tilde{f}_2^*$, and

$$\tilde{f}_i^*(x) \geqslant 0, \quad \mathbb{E}[\tilde{f}_i^*(X)] = \widetilde{m}, \quad \mathbb{E}[(\tilde{f}_i^*(X))^2] = \mathbb{E}[(e_{I_i^*}(X))^2] + \widetilde{m}^2 = \nu + \widetilde{m}^2, \ \ i = 1, 2,$$

where the second equality follows from the fact that $\mathbb{E}[e_{I_i^*}(X)] = 0$. Lemma B.3 yields that $\tilde{f}_1^*(X)$ and $\tilde{f}_2^*(X)$ have the same distribution, and hence so do $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$. As shown above, $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ cannot have the same distribution, and therefore $e_{I_1^*}$ cannot up-cross $e_{I_2^*}$. A similar analysis shows that it is impossible that $e_{I_2^*}$ up-crosses $e_{I_1^*}$. The proof is thus complete. $\qquad\square$

C   Proofs

Proof of Lemma 3.2:

For any $I \in \mathfrak{C}$ with $\mathbb{E}[I(X)] \leqslant m_L$, it follows from Lemma A.2 that

$$X \wedge d_I \leqslant_{cx} R_I(X),$$

where $d_I \geqslant d_L$ is determined by $\mathbb{E}[(X - d_I)_+] = \mathbb{E}[I(X)]$ (or, equivalently, $\mathbb{E}[X \wedge d_I] = \mathbb{E}[R_I(X)]$). Thus, we have $\mathbb{E}[U(W_I(X))] \leqslant \mathbb{E}[U(W_{(x-d_I)_+}(X))]$. Furthermore, according to the proof of Theorem 4.2 in Chi (2019), $\mathbb{E}[U(W_{(x-d)_+}(X))]$ is a decreasing function of d over $[d^*, \mathcal{M})$, where d^* is defined in (3.2). Recalling that $d_L \geqslant d^*$, we conclude that I is no better than $(x - d_L)_+$.

Proof of Lemma 3.3:

We prove by contradiction. Assume $\mathbb{E}[I(X)] > m_U$ for some indemnity function $I \in \mathfrak{C}$ satisfying $var[I(X)] \leqslant \nu$. Lemma A.2 implies that there exists $K > K_U$ such that

$$X \wedge K \leqslant_{cx} I(X).$$

Noting that $var[X \wedge K]$ is strictly increasing and continuous in $K \in [K_U, \mathcal{M}]$, we obtain

$$\nu = var[X \wedge K_U] < var[X \wedge K] \leqslant var[I(X)] \leqslant \nu,$$

leading to a contradiction.

For any $I \in \mathfrak{C}$ satisfying $var[I(X)] \leqslant \nu$ and $\mathbb{E}[I(X)] = m_U$, it follows from Lemma A.2 that

$$X \wedge K_U \leqslant_{cx} I(X),$$

which in turn implies

$$\nu = var[X \wedge K_U] \leqslant var[I(X)] \leqslant \nu.$$

Because

$$var[Z] = 2 \int_{-\infty}^{\infty} \left\{ \mathbb{E}[(Z - t)_+] - (\mathbb{E}[Z] - t)_+ \right\} \mathrm{d}t$$

for any random variable Z with a finite second moment, we deduce $\mathbb{E}[(X \wedge K_U - t)_+] = \mathbb{E}[(I(X) - t)_+]$ almost everywhere in t. Therefore, $X \wedge K_U$ and $I(X)$ are equally distributed, which implies $\mathbb{P}(I(X) \leqslant K_U) = 1$. Moreover, it follows from $I \in \mathfrak{C}$ that $\mathbb{P}(I(X) \leqslant X \wedge K_U) = 1$. Since $\mathbb{E}[I(X)] = \mathbb{E}[X \wedge K_U]$, we obtain that $I(X) = X \wedge K_U$ almost surely. The proof is thus complete.

Proof of Proposition 3.4:

Because $m_L = m_U$, it follows from Lemma 3.3 that $X \wedge K_U = (X - d_L)_+$ almost surely. Thus, it follows from the fact that $d_L > 0$ that X must follow a Bernoulli distribution with values 0 and $d_L + K_U$. Furthermore, Lemmas 3.2 and 3.3 imply that any admissible insurance policy is no better

26

than $(x - d_L)_+$. Therefore, the optimal indemnity must be K_U at point $d_L + K_U$.

Proof of Proposition 3.5:

We prove the desired result with the help of Propositions 2.1 and 2.2 in Chi et al. (2020). More specifically, denote $\widetilde{W} := w_0 - (1 + \rho)m$, then we can rewrite Problem (3.3) as follows

$$
\begin{aligned}
\max_{I \in \mathfrak{C}} \quad & \mathbb{E}[U(\widetilde{W} - X + I(X))] \\
\text{subject to} \quad & \mathbb{E}\left[I(X)\right] = m; \\
& \mathbb{E}\left[(I(X))^2\right] \leqslant m^2 + \nu^2.
\end{aligned} \tag{C.1}
$$

In view of the fact that $m \in (m_L, m_U)$, the optimal solution of Problem (C.1) cannot boil down to Case A in Chi et al. (2020).

Next, we consider the following optimization problem

$$
\begin{aligned}
\max_{I \in \mathfrak{C}} \quad & \mathbb{E}[U(\widetilde{W} - X + I(X))] \\
\text{subject to} \quad & \mathbb{E}\left[(I(X))^2\right] \leqslant m^2 + \nu^2.
\end{aligned} \tag{C.2}
$$

Using Proposition 2.1 in Chi et al. (2020), we can solve Problem (C.2) and denote the corresponding optimal solution as $I^B(x)$. Let $m^B := \mathbb{E}[I^B(X)]$, then we show that $m^B \geqslant m$ by contradiction. Otherwise, there exists $\alpha \in (0, 1)$ such that $m^2 = \alpha(m^B)^2 + (1 - \alpha)(m_U)^2$. Denote $I_\alpha^*(x) := \alpha I^B(x) + (1 - \alpha)X \wedge K_U$, then we must have $\mathbb{E}[I_\alpha^*(X)] > m^B$ and

$$
\begin{aligned}
\mathbb{E}[(I_\alpha^*(X))^2] &\leqslant \alpha \mathbb{E}[(I^B(X))^2] + (1 - \alpha)\mathbb{E}[(X \wedge K_U)^2] \\
&= \alpha\big((m^B)^2 + \nu^2\big) + (1 - \alpha)\big((m_U)^2 + \nu^2\big) = m^2 + \nu^2.
\end{aligned}
$$

Furthermore, we can deduce

$$
\begin{aligned}
\mathbb{E}[U(\widetilde{W} - X + I_\alpha^*(X))] &\geqslant \alpha \mathbb{E}[U(\widetilde{W} - X + I^B(X))] + (1 - \alpha)\mathbb{E}[U(\widetilde{W} - X + X \wedge K_U)] \\
&> \mathbb{E}[U(\widetilde{W} - X + I^B(X))],
\end{aligned}
$$

where the last inequality follows from the facts that $X \wedge K_U \geqslant X \wedge K_{m^B}$ and $X - X \wedge K_{m^B} \leqslant_{cx} X - I^B(X)$. Here, K_{m^B} is such that $\mathbb{E}[X \wedge K_{m^B}] = m^B$. Therefore, it follows that $I^B(x)$ cannot be optimal to Problem (C.2) leading to a contradiction. To sum up, only $m^B = m$ and Case C in Chi et al. (2020) are possible cases.

Finally, directly applying Propositions 2.1 and 2.2 in Chi et al. (2020), we can get the desired results.

Proof of Proposition 3.7:

Denote

$$L_\beta(I) := \mathbb{E}\big[U\big(w_0 - X + I(X) - (1+\rho)\mathbb{E}[I(X)]\big)\big] - \beta\big(var[I(X)] - \nu\big), \ \beta \geqslant 0, \ I \in \mathcal{IC},$$

which is linear in β and concave in I, because $U''(\cdot) < 0$ and $var[I(X)]$ is convex in I.

We denote by U^* the maximum EU value of the insured's final wealth in Problem (3.11). On one hand, for any given $I \in \mathcal{IC}$ satisfying $var[I(X)] > \nu$, it is easy to show that $\min\limits_{\beta \geqslant 0} L_\beta(I) = -\infty$. On the other hand, for any given $\beta \geqslant 0$ and $I \in \mathcal{IC}$ satisfying $var[e_I(X)] = var[I(X)] \leqslant \nu$, we have $\mathbb{E}[U(W_I(X))] \leqslant L_\beta(I)$. As a result, we can deduce

$$\max_{I \in \mathcal{IC}} \min_{\beta \geqslant 0} L_\beta(I) \leqslant U^* \leqslant \min_{\beta \geqslant 0} \max_{I \in \mathcal{IC}} L_\beta(I),$$

which, together with Lemmas B.1 and B.2, implies

$$U^* = \min_{\beta \geqslant 0} \max_{I \in \mathcal{IC}} L_\beta(I).$$

Denoting $\tilde{\beta} = \frac{U^*+1}{\nu}$, we have

$$\max_{I \in \mathcal{IC}} L_\beta(I) \geqslant L_\beta(0) \geqslant \beta\nu \geqslant U^* + 1, \quad \forall \beta \geqslant \tilde{\beta}.$$

Furthermore, since $\max_{I \in \mathcal{IC}} L_\beta(I)$ is convex in β, there must exist $\beta^* \in [0, \tilde{\beta}]$ such that

$$U^* = \max_{I \in \mathcal{IC}} L_{\beta^*}(I).$$

If $\beta^* = 0$, then

$$U^* = \max_{I \in \mathcal{IC}} \mathbb{E}[U(W_I(X))],$$

in which case the stop-loss insurance $(x - d^*)_+$ is optimal, where d^* is defined in (3.2). However, this is contradicted by Assumption 3.1. So we must have $\beta^* > 0$.

According to Lemma 3.2, Lemma 3.3 and Proposition 3.5, I^*, which solves Problem (2.1) under Assumption 3.1, satisfies $var[I^*(X)] = \nu$ and therefore also solves the maximization problem

$$\max_{I \in \mathcal{IC}} L_{\beta^*}(I).$$

Thus, for any $I \in \mathcal{IC}$, we have

$$\lim_{\alpha \downarrow 0} \frac{L_{\beta^*}\big((1-\alpha)I^*(x) + \alpha I(x)\big) - L_{\beta^*}(I^*(x))}{\alpha} \leqslant 0,$$

leading to

$$0 \geqslant \mathbb{E}\Big[U'(W_{I^*}(X))\big(I(X) - I^*(X) - (1+\rho)\mathbb{E}[I(X) - I^*(X)]\big)\Big] - 2\beta^*\big(cov(I^*(X), I(X)) - var[I^*(X)]\big)$$

$$= \int_0^\infty \Big(\mathbb{E}[U'(W_{I^*}(X))(\mathbb{I}_{\{X>t\}} - (1+\rho)\mathbb{P}\{X > t\})]$$

28

$$- 2\beta^*\big(\mathbb{E}[I^*(X)\mathbb{I}_{\{X>t\}}] - \mathbb{E}[I^*(X)]\mathbb{P}\{X > t\}\big)\Big)\big(I'(t) - I^{*\prime}(t)\big)\mathrm{d}t$$

$$= \int_0^{\mathcal{M}} \Big(\mathbb{E}\big[U'(W_{I^*}(X)) - 2\beta^* I^*(X)|X > t\big]$$

$$- (1+\rho)\mathbb{E}\big[U'(W_{I^*}(X))\big] + 2\beta^*\mathbb{E}[I^*(X)]\Big)\mathbb{P}\{X > t\}\big(I'(t) - I^{*\prime}(t)\big)\mathrm{d}t$$

$$= \int_0^{\mathcal{M}} \Phi_{I^*}(t)\mathbb{P}\{X > t\}\big(I'(t) - I^{*\prime}(t)\big)\mathrm{d}t,$$

where Φ_{I^*} is defined in (3.14) and $\mathbb{I}_A$ is the indicator function of an event A. The arbitrariness of $I \in \mathcal{IC}$ and the fact that $\mathbb{P}\{X > t\} > 0$ for any $t < \mathcal{M}$ yield that $I^{*\prime}$ should be in the form of (3.13). The proof is complete.

Proof of Theorem 3.8:

Let I^* be optimal for Problem (2.1). Then it follows from Proposition 3.7 that

$$\Phi_{I^*}(x) = \mathbb{E}\big[\psi(X)|X > x\big] \tag{C.3}$$

where

$$\psi(x) := U'(W_{I^*}(x)) - 2\beta^* I^*(x) - \Big((1+\rho)\mathbb{E}[U'(W_{I^*}(X))] - 2\beta^*\mathbb{E}[I^*(X)]\Big), \tag{C.4}$$

for some $\beta^* > 0$.

Since F_X is assumed to be strictly increasing on $(0, \mathcal{M})$, it follows from Proposition 3.4 that $m_L < m_U$. Recall that solving Problem (2.1) can be reduced to solving Problem (3.10) under Assumption 3.1. In the following, we proceed to solve Problem (3.10) with the help of Proposition 3.7. We carry out the analysis for three cases:

Case (A) If $I^*(x) = (x - d_L)_+$, then ψ is strictly increasing on $[0, d_L]$ and strictly decreasing on $[d_L, \mathcal{M})$. If there exists $\tilde{x} \in [d_L, \mathcal{M})$ such that $\psi(\tilde{x}) < 0$, then $\psi(x) < 0$ and thus $\Phi_{I^*}(x) < 0$ $\forall x \in [\tilde{x}, \mathcal{M})$, which contradicts Proposition 3.7. Therefore, $\psi(x) \geqslant 0$ $\forall x \in [d_L, \mathcal{M})$ and $\psi(d_L) > 0$. Because $\psi(x)$ is continuous in x and F_X is strictly increasing on $(0, \mathcal{M})$, there must exist $\epsilon > 0$ such that $\Phi_{I^*}(x) > 0$ $\forall x \in [d_L - \epsilon, d_L)$, contradicting Proposition 3.7. So $(x - d_L)_+$ cannot be an optimal solution to Problem (3.10).

Case (B) If $I^*(x) = x \wedge K_U$, then ψ is strictly decreasing on $[0, K_U]$ and strictly increasing on $[K_U, \mathcal{M})$. Using a similar argument as the one for Case (A), we deduce $\psi(x) \leqslant 0$ $\forall x \in [K_U, \mathcal{M})$ and $\psi(K_U) < 0$. Since ψ is a continuous function, there exists $\epsilon > 0$ such that $\Phi_{I^*}(x) < 0$ $\forall x \in [K_U - \epsilon, K_U)$, contradicting Proposition 3.7. As a result, $x \wedge K_U$ cannot be optimal to Problem (3.10) either.

Case (C) Hence, the optimal solution must be of the form $I^*(x) = I_{\lambda_m^*, \beta_m^*}(x)$ for some $m \in (m_L, m_U)$. Noting that $I'_{\lambda_m^*, \beta_m^*}(x) \in (0, 1)$ for sufficiently large x due to $\beta_m^* > 0$ and (3.8), we deduce from

Proposition 3.7 that $\Phi_{I^*}(x) = 0$ for sufficiently large x. Thus, together with (3.7), (C.4) and (C.3), Proposition 3.7 further implies that

$$\beta_m^* = \beta^* \qquad \text{and} \qquad \lambda_m^* = (1+\rho)\mathbb{E}[U'(W_{I^*}(X))] - 2\beta^*\mathbb{E}[I^*(X)]. \tag{C.5}$$

Next, we consider two subcases that depend on the values of λ_m^* and $U'(w_0 - (1+\rho)m)$.

(C.1) If $\lambda_m^* < U'(w_0 - (1+\rho)m)$, then $I^*(x) = x \ \forall 0 \leqslant x \leqslant \hat{x}$ and

$$U'(w_0 - x + I^*(x) - (1+\rho)m) - 2\beta^*I^*(x) - \lambda_m^* \begin{cases} = 0, & x \geqslant \hat{x}, \\ > 0, & x < \hat{x}, \end{cases}$$

where $\hat{x} := \frac{U'(w_0 - (1+\rho)m) - \lambda_m^*}{2\beta^*} > 0$. Therefore, it follows that

$$\begin{aligned}
0 \ &< \ \mathbb{E}[U'(w_0 - X + I^*(X) - (1+\rho)m) - 2\beta^*I^*(X) - \lambda_m^*] \\
&= \ \mathbb{E}[U'(W_{I^*}(X))] - 2\beta^*\mathbb{E}[I^*(X)] - (1+\rho)\mathbb{E}[U'(W_{I^*}(X))] + 2\beta^*\mathbb{E}[I^*(X)] \\
&= \ -\rho\mathbb{E}[U'(W_{I^*}(X))],
\end{aligned}$$

leading to a contradiction.

(C.2) Consequently, we must have $\lambda_m^* \geqslant U'(w_0 - (1+\rho)m)$, in which case I^* is coinsurance above a deductible (i.e., (3.5)). Similarly to Subcase (C.1), we can show that

$$U'(w_0 - x + I^*(x) - (1+\rho)m) - 2\beta^*I^*(x) - \lambda_m^* \begin{cases} = 0, & x \geqslant \tilde{d}, \\ < 0, & x < \tilde{d}, \end{cases}$$

where $\tilde{d} := w_0 - (1+\rho)m - (U')^{-1}(\lambda_m^*) \geqslant 0$. This, together with (C.5), yields

$$\lambda_m^* = U'(w_0 - \tilde{d} - (1+\rho)m) \tag{C.6}$$

and

$$-\rho\mathbb{E}[U'(W_{I^*}(X))] = \mathbb{E}\left[\left(U'(w_0 - X + I^*(X) - (1+\rho)m) - 2\beta^*I^*(X) - \lambda_m^*\right)\mathbb{I}_{\{X < \tilde{d}\}}\right]. \tag{C.7}$$

Therefore, $\rho = 0$ if and only if $\tilde{d} = 0$. Moreover, if $\rho = 0$, the above analysis indicates that I^* solves the equation (3.15). Otherwise, if $\rho > 0$, then it follows from (C.5) and (C.7) that

$$\mathbb{E}[U'(W_{I^*}(X))] = \frac{\mathbb{E}[U'(w_0 - X - (1+\rho)m)\mathbb{I}_{\{X \leqslant \tilde{d}\}}] + 2\beta^*mF_X(\tilde{d})}{1 - (1+\rho)\mathbb{P}\{X > \tilde{d}\}},$$

which in turn implies $\tilde{d} > VaR_{\frac{1}{1+\rho}}(X)$. Plugging the above equation and (C.6) into (C.5) yields

$$2\beta^* = \frac{1}{m\rho}\left(U'(w_0 - \tilde{d} - (1+\rho)m) - (1+\rho)\mathbb{E}[U'(w_0 - X \wedge \tilde{d} - (1+\rho)m)]\right).$$

As a result, the optimal solution I^* must be given by (3.17). The proof is complete.

Proof of Corollary 3.9:

We prove for the case in which $\rho > 0$, but note that the proof to the case of $\rho = 0$ is similar and indeed simpler. It follows from (3.5) and (3.8) that $I_{\lambda_m^*, \beta_m^*}(x) = 0$ for $x \leqslant \tilde{d}$ and $I'_{\lambda_m^*, \beta_m^*}(x) = f'_{\lambda_m^*, \beta_m^*}(x)$ increases in x for $x > \tilde{d}$, where we use the fact that $x - I_{\lambda_m^*, \beta_m^*}(x)$ increases in x, and $\beta_m^* > 0$. Moreover, for $x > \tilde{d}$, taking the derivative of $\frac{I_{\lambda_m^*, \beta_m^*}(x)}{x}$ with respect to x yields

$$\left(\frac{I_{\lambda_m^*, \beta_m^*}(x)}{x}\right)' = \left(\frac{f_{\lambda_m^*, \beta_m^*}(x)}{x}\right)' = \frac{f'_{\lambda_m^*, \beta_m^*}(x)x - f_{\lambda_m^*, \beta_m^*}(x)}{x^2} \geqslant \frac{\int_{\tilde{d}}^x \left(f'_{\lambda_m^*, \beta_m^*}(x) - f'_{\lambda_m^*, \beta_m^*}(y)\right)\mathrm{d}y}{x^2} \geqslant 0.$$

This completes the proof.

Proof of Theorem 4.1:

The proof is a slight modification to that of Theorem 3.2 in Chi et al. (2020). We first prove the result *assuming* that there exists $x_0 \in (0, \mathcal{M}]$ such that

$$y_0 := e_{I_1^*}(x_0) = e_{I_2^*}(x_0) \quad \text{and} \quad \phi(x_0 - y_0) > \frac{\beta_1^*}{\beta_2^*}, \tag{C.8}$$

where ϕ is defined in (B.1). First, we show that $e_{I_1^*}(x)$ up-crosses $e_{I_2^*}(x)$ in a neighborhood of x_0. To this end, we first note

$$e'_{I_1^*}(x_0) - e'_{I_2^*}(x_0) = \frac{1}{1 + \frac{2\beta_1^*}{-U''(w_1 - x_0 + y_0)}} - \frac{1}{1 + \frac{2\beta_2^*}{-U''(w_2 - x_0 + y_0)}} > 0,$$

which in turn implies that there exists an $\epsilon > 0$ such that

$$\begin{cases} e_{I_1^*}(x) < e_{I_2^*}(x), & x \in (x_0 - \epsilon, x_0), \\ e_{I_1^*}(x) > e_{I_2^*}(x), & x \in (x_0, x_0 + \epsilon). \end{cases}$$

Next, we show that there exists no $y > x_0$ such that $e_{I_1^*}(y) = e_{I_2^*}(y)$. Otherwise, if such y existed, then the increasing property of $\phi(z)$ in z, together with the strictly increasing property of $x - e_{I_i^*}(x)$ in x, would imply

$$\phi(y - e_{I_1^*}(y)) = \phi(y - e_{I_2^*}(y)) > \phi(x_0 - e_{I_2^*}(x_0)) > \frac{\beta_1^*}{\beta_2^*},$$

which would in turn yield that $e_{I_1^*}$ up-crosses $e_{I_2^*}$ in a neighborhood of y. This, however, contradicts the fact that $e_{I_1^*}(x) > e_{I_2^*}(x)$ for $x \in (x_0, x_0 + \epsilon)$.

Now define

$$x_1 := \sup \left\{x \in [0, x_0) : e_{I_2^*}(x) < e_{I_1^*}(x)\right\}.$$

If $x_1 = 0$, then $e_{I_1^*}$ up-crosses $e_{I_2^*}$, which contradicts Lemma B.4. Thus, we must have $0 < x_1 < x_0$ because $e_{I_1^*}(x) < e_{I_2^*}(x)$ for all $x \in (x_0 - \epsilon, x_0)$. Moreover, it follows readily that

$$\phi\left(x_1 - e_{I_1^*}(x_1)\right) \leqslant \frac{\beta_1^*}{\beta_2^*}.$$

In the following, we show that there exists no point $x_2 \in (0, x_1)$ such that $e_{I_1^*}(x_2) = e_{I_2^*}(x_2)$. Indeed, if such x_2 existed, then, noting that ϕ is strictly increasing, we would have

$$\phi\left(x_2 - e_{I_1^*}(x_2)\right) < \frac{\beta_1^*}{\beta_2^*},$$

leading to $e'_{I_1^*}(x_2) - e'_{I_2^*}(x_2) < 0$. In other words, $e_{I_2^*}$ up-crosses $e_{I_1^*}$ at x_2. This contradicts the fact that $e_{I_2^*}$ up-crosses $e_{I_1^*}$ at x_1. Therefore, we can now conclude that $e_{I_2^*}$ up-crosses $e_{I_1^*}$ twice when (C.8) is satisfied.

Let us now consider the case in which (C.8) is not satisfied. We study two cases:

Case (A) If there exists no $x \in (0, \mathcal{M}]$ such that $e_{I_1^*}(x) = e_{I_2^*}(x)$, then it is easy to show from $\mathbb{E}[e_{I_i^*}(X)] = 0$ that $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ have the same distribution. This contradicts Lemma B.4.

Case (B) Otherwise, any $x \in (0, \mathcal{M}]$ satisfying $e_{I_1^*}(x) = e_{I_2^*}(x)$ must have

$$\phi(x - e_{I_i^*}(x)) \leqslant \frac{\beta_1^*}{\beta_2^*}.$$

If the above inequality is *always* strict, then the previous analysis shows that $e_{I_2^*}$ up-crosses $e_{I_1^*}$, which contradicts Lemma B.4. Hence, there must exist $x_3 \in (0, \mathcal{M}]$ such that

$$y_3 := e_{I_1^*}(x_3) = e_{I_2^*}(x_3) \quad \text{and} \quad \phi(x_3 - y_3) = \frac{\beta_1^*}{\beta_2^*}.$$

In this case, we further divide our analysis into three subcases.

(B.1) If $e_{I_2^*}$ up-crosses $e_{I_1^*}$ at x_3, then the above analysis would imply that no up-crossing occurs before x_3. A similar argument indicates that $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ would have the same distribution, which would not be possible.

(B.2) Otherwise, if $e_{I_1^*}$ up-crosses $e_{I_2^*}$ at x_3, then the previous analysis shows that $e_{I_2^*}$ up-crosses $e_{I_1^*}$ twice.

(B.3) Finally, if no up-crossing happens at x_3, then we can simply neglect this single point x_3 in the analysis. If there further exists $x_4 \in (0, x_3)$ satisfying $e_{I_1^*}(x_4) = e_{I_2^*}(x_4)$, then we have

$$\phi(x_4 - e_{I_i^*}(x_4)) < \frac{\beta_1^*}{\beta_2^*}.$$

In this case, the previous analysis indicates that $e_{I_2^*}$ up-crosses $e_{I_1^*}$ at x_4, which contradicts Lemma B.4. Otherwise, if there exists no $x \in (0, x_3)$ such that $e_{I_1^*}(x) = e_{I_2^*}(x)$, then it follows from $\mathbb{E}[e_{I_i^*}(X)] = 0$ that $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ have the same distribution. This again contradicts Lemma B.4.

In summary, we have shown that $e_{I_2^*}$ up-crosses $e_{I_1^*}$ twice. Because $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ have the same first two moments, we can easily see that the insurer's profit with the wealthier insured, $-e_{I_2^*}(X)$,

has less downside risk than the counterpart with the less wealthy insured, $-e_{I_1^*}(X)$, when the insurance pricing is actuarially fair. The proof is complete.

Proof of Corollary 4.2:

(i) It follows from the proof of Theorem 4.1 that $e_{I_2^*}(0) < e_{I_1^*}(0)$, which is equivalent to $\mathbb{E}[I_1^*(X)] < \mathbb{E}[I_2^*(X)]$. Suppose that $e_{I_2^*}$ up-crosses $e_{I_1^*}$ at points x_0 and x_1 with $0 < x_0 < x_1$. Then $e_{I_1^*}(x_j) = e_{I_2^*}(x_j)$ for $j = 0, 1$, which, together with (4.2), implies

$$U'(w_1 - x_0 + e_{I_1^*}(x_0)) - U'(w_2 - x_0 + e_{I_1^*}(x_0)) - 2(\beta_1^* - \beta_2^*)e_{I_1^*}(x_0)$$
$$= U'(w_1 - x_1 + e_{I_1^*}(x_1)) - U'(w_2 - x_1 + e_{I_1^*}(x_1)) - 2(\beta_1^* - \beta_2^*)e_{I_1^*}(x_1). \tag{C.9}$$

Denote

$$L(y) := U'(w_1 - y) - U'(w_2 - y), \quad y \in [0, w_1).$$

Then

$$L'(y) = -U''(w_1 - y) + U''(w_2 - y) > 0$$

due to $w_1 < w_2$ and $U''' > 0$. Recalling that $x - e_{I_i^*}(x)$ and $e_{I_i^*}(x)$ are strictly increasing in x, we deduce from (C.9) that $\beta_2^* < \beta_1^*$.

(ii) Let us denote $\widetilde{w}_i := w_i - \mathbb{E}[I_i^*(X)]$, $i = 1, 2$. The following analysis depends on the comparison between $\widetilde{w}_1$ and $\widetilde{w}_2$.

Case (A) If $\widetilde{w}_1 = \widetilde{w}_2$ and there exists $z \in [0, \mathcal{M}]$ such that $I_1^*(z) = I_2^*(z)$, then it follows from (3.15) that $2(\beta_1^* - \beta_2^*)I_i^*(z) = 0$, $i = 1, 2$. Recalling that $\beta_2^* < \beta_1^*$, we have $I_i^*(z) = 0$, which implies $z = 0$. Because $\mathbb{E}[I_1^*(X)] < \mathbb{E}[I_2^*(X)]$, we conclude that $I_1^*(x) < I_2^*(x) \ \forall x \in (0, \mathcal{M}]$.

Case (B) Otherwise, if $\widetilde{w}_1 \neq \widetilde{w}_2$, we can show that either $I_1^*(x) < I_2^*(x) \ \forall x > 0$ or I_1^* up-crosses I_2^*. Indeed, if there exists $z \in (0, \mathcal{M}]$ such that $I_1^*(z) = I_2^*(z)$, then we have

$$U'(\widetilde{w}_1 - z + I_1^*(z)) - U'(\widetilde{w}_1) - (U'(\widetilde{w}_2 - z + I_2^*(z)) - U'(\widetilde{w}_2))$$
$$= 2(\beta_1^* - \beta_2^*)I_2^*(z) > 0.$$

Noting that $U'(\widetilde{w} - y) - U'(\widetilde{w})$ is strictly decreasing in $\widetilde{w}$ for any $y > 0$ because of $U''' > 0$, we must have $\widetilde{w}_1 < \widetilde{w}_2$. Furthermore, we have

$$I_i^{*\prime}(z) = \frac{1}{1 + \frac{2\beta_i^*}{-U''(\widetilde{w}_i - z + I_i^*(z))}} = \frac{1}{1 + \frac{U'(\widetilde{w}_i - z + I_i^*(z)) - U'(\widetilde{w}_i)}{-U''(\widetilde{w}_i - z + I_i^*(z)) \times I_i^*(z)}}, \quad i = 1, 2.$$

Denoting

$$H(w) := \frac{U'(w - y) - U'(w)}{-U''(w - y)}, \quad y \in [0, w),$$

33

we obtain

$$
\begin{aligned}
H'(w) &= \frac{U'(w-y) - U'(w)}{-U''(w-y)} \left[\mathcal{P}_U(w-y) + \frac{U''(w-y) - U''(w)}{U'(w-y) - U'(w)} \right] \\
&= H(w)\left[\mathcal{P}_U(w-y) - \mathcal{P}_U(w-\theta y) \right] > 0,
\end{aligned}
$$

where the second equality is due to the mean-value theorem with $\theta \in (0,1)$ and the last inequality is due to the assumption of strict DAP. The strictly increasing property of $H(w)$, together with the fact that $\widetilde{w}_1 < \widetilde{w}_2$, yields $I_2^{*\prime}(z) < I_1^{*\prime}(z)$, which, in turn, implies that I_1^* up-crosses I_2^* at point z.

Otherwise, if $I_1^*(x) \neq I_2^*(x) \ \forall x \in (0, \mathcal{M})$, then it follows from $\mathbb{E}[I_1^*(X)] < \mathbb{E}[I_2^*(X)]$ that $I_1^*(x) < I_2^*(x) \ \forall x > 0$. The proof is complete.

Proof of Theorem 4.3:

If there exists no point $x \in (0, \mathcal{M})$ such that $e_{I_1^*}(x) = e_{I_2^*}(x)$, then $e_{I_1^*}(X)$ and $e_{I_2^*}(X)$ are equal in distribution due to $\mathbb{E}[e_{I_1^*}(X)] = \mathbb{E}[e_{I_2^*}(X)] = 0$. This contradicts the fact that $var[e_{I_1^*}(X)] = \nu_1 < \nu_2 = var[e_{I_2^*}(X)]$. Therefore, we must have $e_{I_1^*}(z) = e_{I_2^*}(z)$ for some $z \in (0, \mathcal{M})$; and thus (4.4) implies

$$
\mathbb{E}[U'(w_0 - X + e_{I_2^*}(X))] - \mathbb{E}[U'(w_0 - X + e_{I_1^*}(X))] = 2(\beta_1^* - \beta_2^*)e_{I_1^*}(z).
$$

We divide the following proof into two cases by comparing $\mathbb{E}[U'(w_0 - X + e_{I_1^*}(X))]$ with $\mathbb{E}[U'(w_0 - X + e_{I_2^*}(X))]$.

Case (A) If $\mathbb{E}[U'(w_0 - X + e_{I_1^*}(X))] \neq \mathbb{E}[U'(w_0 - X + e_{I_2^*}(X))]$, then $\beta_1^* \neq \beta_2^*$. Since $e_{I_1^*}$ is a strictly increasing function, we have $\mathcal{X} := \{x \in [0, \mathcal{M}] : e_{I_1^*}(x) = e_{I_2^*}(x)\} = \{z\}$. Recalling that $var[e_{I_1^*}(X)] < var[e_{I_2^*}(X)]$, we conclude from Lemma A.3 that $e_{I_2^*}$ up-crosses $e_{I_1^*}$.

Case (B) If $\mathbb{E}[U'(w_0 - X + e_{I_1^*}(X))] = \mathbb{E}[U'(w_0 - X + e_{I_2^*}(X))]$, we have

$$
U'(w_0 - x + e_{I_1^*}(x)) - 2\beta_1^* e_{I_1^*}(x) = U'(w_0 - x + e_{I_2^*}(x)) - 2\beta_2^* e_{I_2^*}(x). \tag{C.10}
$$

In this case, we further divide our analysis into two subcases based on the comparison between β_1^* and β_2^*.

(B.1) If $\beta_1^* \neq \beta_2^*$, then it follows from (C.10) that $e_{I_1^*}(x) = 0 \ \forall x \in \mathcal{X}$. Similar to Case (A), we can show that $e_{I_2^*}$ up-crosses $e_{I_1^*}$.

(B.2) If $\beta_1^* = \beta_2^*$, then (C.10) can be rewritten as

$$
U'(w_0 - x + e_{I_1^*}(x)) + 2\beta_1^*(x - e_{I_1^*}(x)) = U'(w_0 - x + e_{I_2^*}(x)) + 2\beta_1^*(x - e_{I_2^*}(x)).
$$

Note that $U'(w_0 - y) + 2\beta_1^* y$ is strictly increasing in y. Hence, the above equation yields $x - e_{I_1^*}(x) = x - e_{I_2^*}(x)$ for all $x \in [0, \mathcal{M}]$, which contradicts the fact that $var[e_{I_1^*}(X)] < var[e_{I_2^*}(X)]$.

In summary, we conclude that $e_{I_2^*}$ up-crosses $e_{I_1^*}$.

Proof of Corollary 4.4:

(i) From the proof of Theorem 4.3, we know that

$$e_{I_2^*}(x) - e_{I_1^*}(x) \begin{cases} < 0, & x < z; \\ > 0, & x \in (z, \mathcal{M}] \end{cases} \tag{C.11}$$

for some $z \in (0, \mathcal{M})$. Therefore, we have

$$\mathbb{E}[I_1^*(X)] = -e_{I_1^*}(0) < -e_{I_2^*}(0) = \mathbb{E}[I_2^*(X)].$$

Furthermore, $e'_{I_1^*}(z) - e'_{I_2^*}(z) \leqslant 0$. Recalling that

$$e'_{I_1^*}(x) = \frac{1}{1 + \frac{2\beta_1^*}{-U''(w_0 - x + e_{I_1^*}(x))}},$$

we have $\beta_2^* \leqslant \beta_1^*$. As $\beta_1^* = \beta_2^*$ does not hold, as shown in the proof of Theorem 4.3, we obtain $\beta_2^* < \beta_1^*$.

(ii) On the one hand, in view of (C.11), it follows from $\mathbb{E}[I_1^*(X)] < \mathbb{E}[I_2^*(X)]$ that $I_2^*(x) > I_1^*(x)$ $\forall x \geqslant z$. On the other hand, for any $x \in [0, z)$, noting that $e_{I_2^*}(x) < e_{I_1^*}(x)$, we deduce from $U''' \geqslant 0$ that

$$-U''(w_0 - x + e_{I_1^*}(x)) \leqslant -U''(w_0 - x + e_{I_2^*}(x)).$$

Because

$$\beta_2^* < \beta_1^* \quad \text{and} \quad e'_{I_i^*}(x) = \frac{1}{1 + \frac{2\beta_i^*}{-U''(w_0 - x + e_{I_i^*}(x))}}, \quad i = 1, 2,$$

we have $e'_{I_1^*}(x) < e'_{I_2^*}(x)$, which is equivalent to $I_1^{*\prime}(x) < I_2^{*\prime}(x)$ $\forall x \in [0, z)$. Furthermore, as $I_1^*(0) = I_2^*(0) = 0$, it must hold that $I_1^*(x) < I_2^*(x)$ $\forall x \geqslant 0$. The proof is complete.

References

Armantier, O., Foncel, J. and Treich, N. (2018) Insurance and portfolio decisions: A wealth effect puzzle. *Working paper.*

Arrow, K.J. (1963) Uncertainty and the welfare economics of medical care. *American Economic Review,* **53(5)**, 941-973.

Arrow, K.J. (1965) Aspects of the Theory of Risk-bearing. *Helsinki: Yrjo Jahnssonin Saatio.*

Arrow, K.J. (1971) Essays in the Theory of Risk Bearing. *Chicago.*

Arrow, K.J. (1974) Optimal insurance and generalized deductibles. *Scandinavian Actuarial Journal,* **1974(1)**, 1-42.

Asimit, A.V., Cheung, K.C., Chong, W.F. and Hu, J. (2020) Pareto-optimal insurance contracts with premium budget and minimum charge constraints. *Insurance: Mathematics and Economics,* **95**, 17-27.

Bahnemann, D. (2015) Distributions for actuaries. *Arlington, VA: Casualty Actuarial Society.*

Bernard, C., He, X.D., Yan, J.A. and Zhou, X.Y. (2015) Optimal insurance design under rank dependent utility. *Mathematical Finance,* **25(1)**, 154-186.

Birghila, C., Boonen, T.J. and Ghossoub, M. (2023) Optimal insurance under maxmin expected utility. *Finance and Stochastics,* **27(2)**, 467-501.

Borch, K. (1960) An attempt to determine the optimum amount of stop loss reinsurance. In: Transactions of the 16th International Congress of Actuaries, Vol. I, 597-610. *Brussels, Belgium: Georges Thone.*

Boudt, K., Dragun, K. and Vanduffel, S. (2022) The optimal payoff for a Yaari investor. *Quantitative Finance,* **22(10)**, 1839-1852.

Cai, J., Liu, H. and Wang, R. (2017) Pareto-optimal reinsurance arrangements under general model settings. *Insurance: Mathematics and Economics,* **77**, 24-37.

Carlier, G. and Dana, R.-A. (2003) Pareto efficient insurance contracts when the insurer's cost function is discontinuous. *Economic Theory,* **21(4)**, 871-893.

Carlier, G. and Dana, R.-A. (2005) Rearrangement inequalities in non-convex insurance models. *Journal of Mathematical Economics,* **41**, 483-503.

Cheung, K.C., He, W. and Wang, H. (2023) Multi-constrained optimal reinsurance model from the duality perspectives. *Insurance: Mathematics and Economics,* **113**, 199-214.

Chi, Y. (2012) Reinsurance arrangements minimizing the risk-adjusted value of an insurer's liability. *ASTIN Bulletin: The Journal of the IAA,* **42(2)**, 529-557.

Chi, Y. (2019) On the optimality of a straight deductible under belief heterogeneity. *ASTIN Bulletin: The Journal of the IAA,* **49(1)**, 242-263.

Chi, Y., Tan, K.S. and Zhuang, S.C. (2020) A Bowley solution with limited ceded risk for a monopolistic reinsurer. *Insurance: Mathematics and Economics,* **91**, 188-201.

Chi, Y. and Wei, W. (2020) Optimal insurance with background risk: An analysis of general dependence structures. *Finance and Stochastics*, **24**, 903-937.

Cummins, J.D. and Mahul, O. (2004) The demand for insurance with an upper limit on coverage. *The Journal of Risk and Insurance*, **71(2)**, 253-264.

Dionne, G. and St-Michel, P. (1991) Workers' compensation and moral hazard. *The Review of Economics and Statistics*, **73(2)**, 236-244.

Doherty, N.A., Laux, C. and Muermann, A. (2015) Insuring nonverifiable losses. *Review of Finance*, **19(1)**, 283-316.

Doherty, N.A. and Schlesinger, H. (1990) Rational insurance purchasing: Consideration of contract non-performance. *The Quarterly Journal of Economics*, **105(1)**, 243-253.

Eeckhoudt, L. and Kimball, M. (1992) Background risk, prudence, and the demand for insurance. Contributions to Insurance Economics, edited by G. Dionne, 239-254. *New York: Springer.*

Eeckhoudt, L., Mahul, O. and Moran, J. (2003) Fixedreimbursement insurance: Basic properties and comparative statics. *Journal of Risk and Insurance*, **70(2)**, 207-218.

Ehrlich, I. and Becker, G.S. (1972) Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, **80(4)**, 623-648.

Gollier, C. (1996) Optimum insurance of approximate losses. *The Journal of Risk and Insurance*, **63(3)**, 369-380.

Gollier, C. (2001) The Economics of Risk and Time. *London: The MIT Press.*

Gollier, C. (2013) The economics of optimal insurance design. Handbook of Insurance, edited by G. Dionne. *New York: Springer.*

Gollier, C. and Schlesinger, H. (1996). Arrow's theorem on the optimality of deductibles: a stochastic dominance approach. *Economic Theory*, **7(2)**, 359-363.

Gotoh, J.Y. and Konno, H. (2000) Third degree stochastic dominance and mean-risk analysis. *Management Science*, **46(2)**, 289-301.

Huang, R.J. and Tzeng, L.Y. (2006) The design of an optimal insurance contract for irreplaceable commodities. *The Geneva Risk and Insurance Review*, **31(1)**, 11-21.

Hölmstrom, B. (1979) Moral hazard and observability. *The Bell Journal of Economics*, **10(1)**, 74-91.

Huberman, G., Mayers, D. and Smith, C.W. (1983) Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, **14(2)**, 415-426.

Jiang, W., Hong, H. and Ren, J. (2021) Pareto-optimal reinsurance policies with maximal synergy. *Insurance: Mathematics and Economics*, **96**, 185-198.

Kaluszka, M. (2001) Optimal reinsurance under mean-variance premium principles. *Insurance: Mathematics and Economics*, **28(1)**, 61-67.

Kaluszka, M. (2004) An extension of Arrows result on optimality of a stop loss contract. *Insurance: Mathematics and Economics*, **35(3)**, 527-536.

Kaluszka, M. (2005) Optimal reinsurance under convex principles of premium calculation. *Insurance: Mathematics and Economics*, **36(3)**, 375-398.

Karlin, S. and Novikoff, A. (1963) Generalized convex inequalities. *Pacific Journal of Mathematics*, **13(4)**, 1251-1279.

Kaye, P. (2005) Risk measurement in insurance: A guide to risk measurement, capital allocation and related decision support issues. Discussion paper program, *Casualty Actuarial Society.*

Kimball, M.S. (1990) Precautionary saving in the small and in the large. *Econometrica*, **58(1)**, 53-73.

Komiya, H. (1988) Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, **11(1)**, 5-7.

Mandal, P.K. and Thakur, M. (2023) Higher-order moments in portfolio selection problems: A comprehensive literature review. *Expert Systems with Applications*, **in press**. DOI: 10.1016/j.eswa.2023.121625.

Markowitz, H. (1952) Portfolio selection. *Journal of Finance*, **7**, 7791.

Menezes, C., Geiss, C. and Tressler, J. (1980) Increasing downside risk. *The American Economic Review*, **70(5)**, 921-932.

Millo, G. (2016) The income elasticity of nonlife insurance: A reassessment. *Journal of Risk and Insurance*, **83(2)**, 335-362.

Mossin, J. (1968) Aspects of rational insurance purchasing. *Journal of Political Economy*, **76(4)**, 553-568.

Noussair, C.N., Trautmann, S.T. and Van de Kuilen, G. (2014) Higher order risk attitudes, demographics, and financial decisions. *The Review of Economic Studies*, **81(1)**, 325-355.

Ohlin, J. (1969) On a class of measures of dispersion with application to optimal reinsurance. *ASTIN Bulletin: The Journal of the IAA*, **5(2)**, 249-266.

Picard, P. (2000) On the design of optimal insurance policies under manipulation of audit cost. *International Economic Review*, **41(4)**, 1049-1071.

Pratt, J.W. (1964) Risk aversion in the small and in the large. *Econometrica*, **32**, 122-136.

Raviv, A. (1979) The design of an optimal insurance policy. *American Economic Review*, **69(1)**, 84-96.

Rothschild, M. and Stiglitz, J. (1976) Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, **90(4)**, 629-649.

Schlesinger, H. (1981) The optimal level of deductibility in insurance contracts. *Journal of Risk and Insurance*, **48(3)**, 465-481.

Teh, T.L. (2017) Insurance design in the presence of safety nets. *Journal of Public Economics*, **149**, 47-58.

Vajda, S. (1962) Minimum variance reinsurance. *ASTIN Bulletin: The Journal of the IAA*, **2**, 257-260.

Viscusi, W.K. (1979) Insurance and individual incentives in adaptive contexts. *Econometrica*, **47(5)**, 1195-1207.

Whitmore, G.A. (1970) Third-degree stochastic dominance. *The American Economic Review*, **60(3)**, 457-459.

Xu, Z.Q., Zhou, X.Y. and Zhuang, S.C. (2019) Optimal insurance with rank-dependent utility and incentive compatibility. *Mathematical Finance*, **29(2)**, 659-692.

Zhou, C. and Wu, C. (2008) Optimal insurance under the insurer's risk constraint. *Insurance: Mathematics and Economics*, **42(3)**, 992-999.

Zhou, C., Wu, W. and Wu, C. (2010) Optimal insurance in the presence of insurer's loss limit. *Insurance: Mathematics and Economics*, **46(2)**, 300-307.