Received: 1 January 2017 | Accepted: 1 February 2018
DOI: 10.1111/mafi.12185

WILEY MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics

ORIGINAL ARTICLE

Optimal insurance under rank-dependent utility and incentive compatibility

Zuo Quan Xu[1] | Xun Yu Zhou[2] | Sheng Chao Zhuang[3]

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[2]Department of IEOR, Columbia University, New York

[3]Department of Finance, University of Nebraska–Lincoln, Lincoln, Nebraska

Correspondence
Xun Yu Zhou, Department of IEOR, Columbia University, New York, NY 10027.
Email: xz2574@columbia.edu

Funding information
NSFC, Grant Number: 11471276; Hong Kong GRF, Grant Numbers: 15204216 and 15202817; Hong Kong Polytechnic University; Columbia University; Oxford–Nie Financial Big Data Lab.

Abstract
Bernard, He, Yan, and Zhou (*Mathematical Finance*, *25*(1), 154–186) studied an optimal insurance design problem where an individual's preference is of the rank-dependent utility (RDU) type, and show that in general an optimal contract covers both large and small losses. However, their results suffer from the unrealistic assumption that the random loss has no atom, as well as a problem of moral hazard that provides incentives for the insured to falsely report the actual loss. This paper addresses these setbacks by removing the nonatomic assumption, and by exogenously imposing the "incentive compatibility" constraint that both indemnity function and insured's retention function are increasing with respect to the loss. We characterize the optimal solutions via calculus of variations, and then apply the result to obtain explicitly expressed contracts for problems with Yaari's dual criterion and general RDU. Finally, we use numerical examples to compare the results between ours and Bernard et al.

KEYWORDS
incentive compatibility, indemnity function, moral hazard, optimal insurance design, probability weighting function, quantile formulation, rank-dependent utility theory, retention function

1 | INTRODUCTION

Risk sharing is a method of reducing risk exposure by spreading the burden of loss among several parties. Mathematically, risk sharing can be generally formulated as a multioptimization problem in which Pareto optimality is sought with respect to each party's well-being modeled as a preference functional.

In the context of insurance, the primary risk-sharing problem is that of designing an insurance contract between an insurer and an insured who achieves Pareto optimality for the two parties. Specifically, given an upfront premium that the insured pays the insurer, the problem is to determine the amount of loss $I(X)$ covered by the insurer—called indemnity—for a random, typically nonhedgeable loss X. The premium usually includes a safety loading on top of the actuarial value of the contract in order for the insurer to have a sufficient incentive to offer the contract—this is called the participation constraint of the insurer.

Optimal insurance contract design is an important problem, manifested not only in theory but also in insurance and financial practices. In the insurance literature, most of the work assumes that the insurer is risk neutral,[1] while the insured is a risk-averse expected utility (EU) maximizer (see, e.g., Arrow, 1963; Gollier & Schlesinger, 1996; Raviv, 1979). The problem is formulated as one that maximizes the insured's expected *concave* utility function of his net wealth subject to the insurer's participant constraint being satisfied. Technically, it is a constrained convex optimization problem that can be solved by standard optimization techniques. It has been shown in the aforementioned papers that the optimal contract is in general a deductible one that covers part of the loss in excess of a deductible level. This theoretical result is consistent with most of the insurance contracts available in practice. As a result, the problem is reduced to a one-dimensional optimization problem that determines the optimal deductible. Another important implication of this classical result is that the insurer and insured shares of risk are both increasing functions of the risk;[2] in other words, there is no incentive for either party to hide risk, and thus there is a genuine sharing of risk.

However, EU theory has received many criticisms, for it fails to explain numerous experimental observations and theoretical puzzles. For example, it fails to explain the famous Allais paradox or the reason why a same person may buy both lottery and insurance. Other paradoxes/puzzles that EU theory cannot explain include common ratio effect (Allais, 1953), the Friedman and Savage puzzle (Friedman & Savage, 1948), the Ellsberg paradox (Ellsberg, 1961), and the equity premium puzzle (Mehra & Prescott, 1985). In the context of insurance contracting, the classical EU-based models again fail to account for some behaviors in insurance demand. Sydnor (2010) investigated how people choose the deductible decisions between \$100, \$250, \$500, and \$1,000. The major finding is that the households choosing a \$500 deductible pay an average premium of \$715 per year, yet these households all rejected a policy with a \$1,000 deductible whose average premium was just \$615. As the claim rate is about 5%, effectively these households were willing to pay \$100 to protect against a 5% possibility of paying an additional \$500! As explained by Barberis (2013), this choice can only be explained by unreasonably high levels of risk aversion within the EU framework.[3] Another insurance phenomenon that cannot be explained by the EU theory is demand for protection of small losses (e.g., demand for warranties); see Bernard, He, Yan, and Zhou (2015) for a detailed discussion.

In order to overcome this drawback of EU theory, different measures of evaluating uncertain outcomes have been put forward to depict human behavior. A notable measure is rank-dependent utility (RDU) proposed by Quiggin (1982). In this theory, the preference measure of a final (random) wealth $W \geqslant 0$ is defined as

$$V^{rdu}(W) = \int u(W) d(T \circ \mathbb{P}) := \int_{\mathbb{R}^+} u(x) d[-T(1 - F_W(x))], \tag{1}$$

where $u : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is a (usual) utility function, $T : [0, 1] \mapsto [0, 1]$ is called a probability *weighting* function, and $F_W(\cdot)$ is the cumulative distribution function (CDF) of W. Clearly, if $T(x) \equiv x$ then

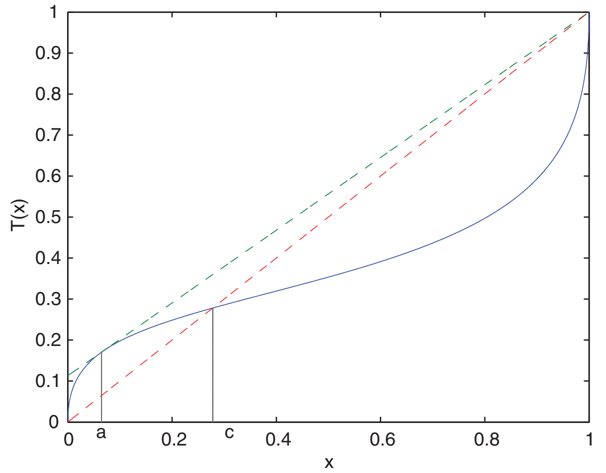

FIGURE 1 An inverse-S-shaped weighting function (solid line) satisfying Assumption 2.3 [Color figure can be viewed at wileyonlinelibrary.com]
Note. The marked points a and c will be explained later.

$V^{rdu}(W) = E[u(W)]$, the classical EU. To see what a nonidentity function T brings about, we rewrite assuming that T is differentiable

$$V^{rdu}(W) = \int_{\mathbb{R}^+} u(x)T'(1 - F_W(x))dF_W(x). \tag{2}$$

Thus, $T'(1 - F_W(x))$ serves as a weight on the outcome x of W when evaluating the EU. This weight depends on $1 - F_W(x)$, the decumulative probability or the *rank* of the outcome x of W, hence the name of the RDU.[4] In particular, if T is inverse-S shaped; that is, it is first concave and then convex (see Figure 1), then $T'(1 - F_W(x)) > 1$ when x is both sufficiently large and sufficiently small. This captures the common observation that people tend to exaggerate small probabilities of extremely good and bad outcomes (hence people buy both insurances and lotteries).

From the optimization point of view, maximizing the RDU preference (1) has a clear challenge: With the presence of a general weighting function T, (1) is no longer concave even if u is concave. With the development of advanced mathematical tools, the RDU preference has been applied to many areas of finance, including portfolio choice and option pricing. In particular, the approach of the so-called quantile formulation has been developed to deal with the nonconvex optimization involved in solving RDU portfolio choice models (He & Zhou, 2011; Jin & Zhou, 2008). The key idea is to change the decision variable from the wealth W to its quantile function, which miraculously leads to a concave optimization problem. On the other hand, Barseghyan, Molinari, O'Donoghue, and Teitelbaum (2013) use data on households' insurance deductible decisions in auto and home insurance to demonstrate the relevance and importance of the probability weighting and suggest the possibility of generalizing their conclusions to other insurance choices.

There have been also studies in the area of insurance contract design within the RDU framework (see, e.g., Carlier & Dana, 2008; Chateauneuf, Dana, & Tallon, 2000; Dana & Scarsini, 2007). However, all these papers assume that the probability weighting function is convex. Bernard et al. (2015) are probably the first to study RDU-based insurance contracting with *inverse-S-shaped* weighting functions, using the quantile formulation. They derive optimal contracts that not only insure large losses above a deductible level but also cover small ones. However, their results suffer from two major problems. One is the assumption that the random loss X has no atom, which is not realistic in the insurance context. The reason is that 0 is typically an atom of X, as it is plausible that $\mathbb{P}(X = 0) > 0$. The second is that their contracts pose a severe problem of moral hazard, as neither the part of the loss to be covered by the insurer (i.e., the indemnity) nor that to be borne by the insured (i.e., the retention or residual loss) are globally increasing with respect to the loss. As a consequence, insured may be motivated to either hide or exaggerate their true losses (see a discussion on Bernard et al., 2015, pp. 175–176).

In one of the earliest papers on moral hazard in the context of insurance contract design, Huberman, Mayers, and Smith (1983) considered an optimal indemnity schedule problem taking the possibility of the insured's bankruptcy into account. They show how the insured may be motivated to misreport the losses if the indemnity or the retention is decreasing in certain loss region, and conclude that the search for optimal indemnity schedules should be confined to the policies under which such moral hazard will not arise. Picard (2000) drew a similar conclusion under manipulation of audit cost. Both Huberman et al. (1983) and Picard (2000) called the increasing condition of indemnity and retention the "incentive compatibility" (IC) constraint for optimal insurance contracting.[5]

An important type of moral hazard particularly relevant to insurance is the so-called ex post moral hazard where an insured can even take some actions to create losses or affect the loss magnitude after losses occur.[6] It has attracted a lot of attention of economists, such as Spence and Zeckhauser (1971), Townsend (1979), and Shavell (1979). In these papers, the magnitude of a claim depends on the insured's own claim report. There are also empirical studies investigating the ex post moral hazard. For instance, Dionne and St-Michel (1991) presented an empirical measure of the level of ex post moral hazard in the workers' compensation market. Butler, Durbin, and Helvacian (1996) found empirical evidence that moral hazard can explain most of the 30% increase in the proportion of soft tissue injury claims. Cummins and Tennyson (1996) showed a strong evidence that ex post moral hazard exists in the automobile insurance market.

This paper aims to address the aforementioned two issues in Bernard et al. (2015). We consider the same insurance model as in Bernard et al. (2015), but remove the nonatomic assumption on the loss, and add an explicit IC constraint that both indemnity function and insured's retention function must be globally increasing with respect to the losses—this latter constraint will rule out completely the moral hazard just discussed. However, mathematically we encounter substantial difficulty. The approach used in Bernard et al. (2015) no longer works. We develop a general approach to overcome this difficulty. Specifically, we first derive the necessary and sufficient conditions for optimal solutions via calculus of variations. Although calculus of variations is a rather standard technique for infinite-dimensional optimization,[7] deducing explicitly expressed optimal contracts based on these conditions requires a fine and involved analysis. This is the main technical contribution of this paper. An interesting finding is that for a good and reasonable range of parameter specifications, there are only two types of optimal contracts, one being the classical deductible contract and the other a "threefold" contract covering both small and large losses.

The remainder of the paper is organized as follows. Section 2 presents the optimal insurance model under the RDU framework including its quantile formulation. Section 3 applies the calculus of variations to derive a general necessary and sufficient condition for optimal solutions. We then derive optimal contracts for Yaari's criterion and the general RDU in Sections 4 and 5, respectively. Section 6 provides a numerical example to illustrate our results. Finally, we conclude with Section 7. All the proofs and some auxiliary results are placed in an Appendix.

2 | THE MODEL

In this section, we present the optimal insurance contracting model in which the insured has RDU type of preferences, as well as its quantile formulation that will facilitate deriving the solutions.

2.1 | Problem formulation

We follow Bernard et al. (2015) for the problem formulation except for two critical differences, which we will highlight. Let $(\Omega, \mathbf{F}, \mathbb{P})$ be a probability space. An insured, endowed with an initial wealth W_0, faces a nonnegative random loss X, *possibly having atoms* and supported in $[0, M]$, where M is a given positive scalar. Here, following the majority of insurance literature, we consider a single type of risk (rather than an aggregate of several types of risk) related to, say, fire, automobile, disability, or travel. He chooses an insurance contract to protect himself from the loss by paying a premium π to the insurer in return for a compensation (or *indemnity*) in the case of a loss. This compensation is to be determined as a function of the loss X, denoted by $I(\cdot)$ throughout this paper. The *retention* function $R(X) := X - I(X)$ is in turn the part of the loss to be borne by the insured.

For a given X, the insured aims to choose an insurance contract that provides the best trade-off between the premium and compensation based on his risk preference. In this paper, we consider the case when the insured's preference on the final random wealth $W > 0$ is dictated by the RDU functional (1), where $u : \mathbb{R}^+ \mapsto \mathbb{R}^+$ and $T : [0, 1] \mapsto [0, 1]$. On the other hand, if the insurer is risk neutral and the cost of offering the compensation is proportional to the expectation of the indemnity, then the premium to be charged for an insurance contract should satisfy the participation constraint

$$\pi \geqslant (1 + \rho)E[I(X)],$$

where the constant ρ is the *safety loading* of the insurer.

It is natural to require an indemnity function to satisfy

$$I(0) = 0, \quad 0 \leqslant I(x) \leqslant x, \quad \forall\, 0 \leqslant x \leqslant M, \tag{3}$$

a constraint that has been imposed in most insurance contracting literature. If the insured's preference is dictated by the classical EU theory, then the optimal contract is typically a deductible contract (i.e., $I(X) = (X - d)^+$), which automatically renders both indemnity function and retention function increasing, a condition called the IC (see, e.g., Arrow, 1971; Raviv, 1979). In practice, three commonly observed insurance provisions are deductible, policy limit (i.e., $I(X) = X \wedge u$), and proportional contract (i.e., $I(X) = \alpha X$, where $\alpha \in (0, 1]$), which all satisfy the IC. However, for the RDU preference specification the resulting optimal insurance schedule may not satisfy this condition, as shown in Bernard et al. (2015). This may potentially cause moral hazard as pointed out earlier. To incorporate the IC constraint on the contract has been an open question.

In this paper, we impose the IC constraint explicitly to our model. Economically speaking, this means the insurer and insured wealths are comonotone, both bearing more when a bigger loss happens. Mathematically speaking, we require

$$I(y) \leqslant I(x), \quad R(y) \leqslant R(x), \quad \forall\, 0 \leqslant y \leqslant x \leqslant M. \tag{4}$$

As $R(x) \equiv x - I(x)$, it is easily seen that the joint constraint of (3) and (4) is equivalent to the following:

$$I(0) = 0, \quad 0 \leqslant I(x) - I(y) \leqslant x - y, \quad \forall\, 0 \leqslant y \leqslant x \leqslant M. \tag{5}$$

We can now formulate our insurance contracting problem as

$$\begin{aligned} \max_{I(\cdot)} \quad & V^{rdu}(W_0 - \pi - X + I(X)) \\ \text{s.t.} \quad & (1+\rho)E[I(X)] \leqslant \pi, \\ & I(\cdot) \in \mathbb{I}, \end{aligned} \tag{6}$$

where

$$\mathbb{I} := \{I(\cdot) : I(0) = 0,\ 0 \leqslant I(x) - I(y) \leqslant x - y,\ \forall\, 0 \leqslant y \leqslant x \leqslant M\}, \tag{7}$$

and W_0 and π are fixed scalars.

For any random variable $Y \geqslant 0$ a.s., define the quantile function of Y as

$$F_Y^{-1}(t) := \inf\{x \in \mathbb{R}^+ : P(Y \leqslant x) \geqslant t\}, \quad t \in [0,1].$$

Note that any quantile function is nonnegative, increasing and left-continuous.

We now introduce the following assumptions that will be used hereafter.

Assumption 2.1. The random loss X has a strictly increasing distribution function F_X. Moreover, F_X^{-1} is absolutely continuous on [0,1].

Assumption 2.2 (**Concave utility**)**.** The utility function $u : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is strictly increasing and continuously differentiable. Furthermore, u' is decreasing.

Assumption 2.3 (**Inverse-S-shaped weighting**)**.** The probability weighting function T is a continuous and strictly increasing mapping from [0,1] onto [0,1] and twice differentiable on (0,1). Moreover, there exists $b \in (0,1)$ such that $T'(\cdot)$ is strictly decreasing on $(0,b)$ and strictly increasing on $(b,1)$. Furthermore, $T'(0+) := \lim_{z\downarrow 0} T'(z) > 1$ and $T'(1-) := \lim_{z\uparrow 1} T'(z) = +\infty$.

The first part of Assumption 2.1 is standard in the insurance literature that accommodates most of the continuous distributions actuaries frequently use to fit sample data such as uniform, exponential, lognormal, gamma, and Pareto distributions (see, e.g., Arrow, 1963; Bernard et al., 2015; Doherty & Eeckhoudt, 1995; Doherty & Posey, 1997; Huberman et al., 1983; Raviv, 1979; Smith, 1968). The second part of the assumption is very mild, and is again satisfied in all the papers just mentioned. Bahnemann (2015) painstakingly explained why "the task of the actuary often is to fit a continuous parametric claim-size distribution to a discrete sample of claim data" and argued "distributions of various derived random variables are neither wholly discrete nor continuous, but of the mixed discrete/continuous type."[8] It can be shown that a mixed discrete/continuous distribution satisfies Assumption 2.1 so long as its continuous component does. On the other hand, as noted earlier, a significant difference from Bernard et al. (2015) is that here we allow X to have atoms.[9] For example, let $F_X(x) = \frac{1-\gamma e^{-\eta x}}{1-\gamma e^{-\eta M}}$ for $x \in [0, M]$, where $\gamma \in (0,1)$ and $\eta > 0$. Then, X satisfies Assumption 2.1, and has an atom at 0 with the probability $\mathbb{P}(X = 0) = \frac{1-\gamma}{1-\gamma e^{-\eta M}} > 0$. This assumption also ensures that $F_X^{-1}(F_X(x)) \equiv x, \forall\, x \in [0, M]$, a fact that will be used often in the subsequent analysis.

Next, Assumption 2.2 is standard for a utility function. Finally, Assumption 2.3 is satisfied for many weighting functions proposed or used in the literature, for example, the one proposed by Tversky and Kahneman (1992; parameterized by θ):

$$T_\theta(x) = \frac{x^\theta}{(x^\theta + (1-x)^\theta)^{\frac{1}{\theta}}}. \tag{8}$$

Figure 1 displays this (inverse-S-shaped) weighting function (solid line) when $\theta = 0.5$.

In practice, most of the insurance contracts are not tailor-made for individual customers. Instead, an insurance company usually has contracts with different premiums to accommodate customers with different needs. Each contract is designed with the best interest of a representative customer in mind so as to stay marketable and competitive, while maintaining the desired profitability (the participation constraint). An insured can then choose one from the menu of contracts to cater for individual needs. Problem (6) is therefore motivated by the insurer's making of this menu.

If the premium $\pi \geqslant (1+\rho)E[X]$, then $I^*(x) \equiv x$ (corresponding to a full coverage) is feasible and maximizes the objective function in the problem (6) pointwise; hence it is optimal. To rule out this trivial case, henceforth we restrict $0 < \pi < (1+\rho)E[X]$. Moreover, we assume

$$W_0 - (1+\rho)E[X] - M \geqslant 0, \tag{9}$$

to ensure that the policyholder will not go bankrupt because $W_0 - \pi - M > 0$ for all $0 < \pi < (1+\rho)E[X]$.

It is more convenient to consider the retention function $R(x) = x - I(x)$ instead of $I(x)$ in our study below. Letting

$$\begin{aligned}
\Delta &:= E[X] - \frac{\pi}{1+\rho} \in (0, E[X]), \\
W &:= W_0 - (1+\rho)E[X] > 0, \\
W_\Delta &:= W + (1+\rho)\Delta \equiv W_0 - \pi,
\end{aligned}$$

one can easily reformulate (6) in terms of $R(\cdot)$

$$\begin{array}{ll}
\max\limits_{R(\cdot)} & V^{rdu}(W_\Delta - R(X)) \\
\text{s.t.} & E[R(X)] \geqslant \Delta, \\
 & R(\cdot) \in \mathcal{R},
\end{array} \tag{10}$$

where

$$\mathcal{R} := \{R(\cdot) : R(0) = 0,\ 0 \leqslant R(x) - R(y) \leqslant x - y,\ \forall\, 0 \leqslant y \leqslant x \leqslant M\}.$$

2.2 | Quantile formulation

The objective function in (10) is not concave in $R(X)$ (due to the nonlinear weighting function T), leading to a major difficulty in solving (10). However, under Assumption 2.3, we have

$$\begin{aligned}
V^{rdu}(W_\Delta - R(X)) &= \int_{\mathbb{R}^+} u(x)d[-T(1 - F_{W_\Delta - R(X)}(x))] \\
&= \int_0^1 u(F^{-1}_{W_\Delta - R(X)}(z))T'(1-z)dz = \int_0^1 u(W_\Delta - F^{-1}_{R(X)}(1-z))T'(1-z)dz \\
&= \int_0^1 u(W_\Delta - F^{-1}_{R(X)}(z))T'(z)dz,
\end{aligned}$$

where the third equality is because

$$F^{-1}_{W_\Delta - R(X)}(z) = W_\Delta - F^{-1}_{R(X)}(1-z)$$

except for an at most countable set of z. Moreover, $E[R(X)] \geqslant \Delta$ is equivalent to $\int_0^1 F_{R(X)}^{-1}(z)dz \geqslant \Delta$.

The above suggests that we may change the decision variable from the random variable $R(X)$ to its quantile function $F_{R(X)}^{-1}$, with which the objective function of (10) becomes concave and the first constraint is linear. It remains to rewrite the monotonicity constraint (represented by the constraint set $\mathcal{R}$) also in terms of $F_{R(X)}^{-1}$. To this end, the next lemma plays an important role.

Lemma 2.4. *Under Assumption 2.1, for any given $R(\cdot) \in \mathcal{R}$, we have*

$$R(x) = F_{R(X)}^{-1}(F_X(x)), \quad \forall\, x \in [0, M].$$

In view of the above results, we can rewrite (10) as the following problem, in which the decision variable is $F_{R(X)}^{-1}(\cdot)$ (denoted by $G(\cdot)$ for simplicity)

$$\begin{array}{ll} \max\limits_{G(\cdot)} & \int_0^1 u(W_\Delta - G(z))T'(z)dz, \\ \text{s.t.} & \int_0^1 G(z)dz \geqslant \Delta, \\ & G(\cdot) \in \mathbb{G}, \end{array} \tag{11}$$

where $\mathbb{G} := \{F_{R(X)}^{-1}(\cdot) : R(\cdot) \in \mathcal{R}\}$.

In the absence of an explicit expression the constraint set $\mathbb{G}$ is hard to deal with. The following result addresses this issue. Note the major technical difficulty arises from the possible existence of the atoms of X.

Lemma 2.5. *Under Assumption 2.1, we have*

$$\mathbb{G} = \{G(\cdot) : G(\cdot) \textit{ is absolutely continuous, } G(0) = 0,\ 0 \leqslant G'(z) \leqslant (F_X^{-1})'(z),\ \textit{a.e. } z \in [0, 1]\}. \tag{12}$$

To solve (11), we apply the Lagrange dual method to remove the constraint $\int_0^1 G(z)dz - \Delta \geqslant 0$ and consider the following auxiliary problem:

$$\begin{array}{ll} \max\limits_{G(\cdot)} & U_\Delta(\lambda, G(\cdot)) := \int_0^1 [u(W_\Delta - G(z))T'(z) + \lambda G(z)]dz - \lambda\Delta, \\ \text{s.t.} & G(\cdot) \in \mathbb{G}. \end{array} \tag{13}$$

The existence of the optimal solutions to (11) and (13) (for each given $\lambda \in \mathbb{R}^+$) is established in Appendix B, while the uniqueness is straightforward when the utility function u is strictly concave.

To derive the optimal solution to (11), we first solve (13) to obtain an optimal solution, denoted by $\widetilde{G}_\lambda(\cdot)$. Then we determine $\lambda^* \in \mathbb{R}^+$ by binding the constraint $\int_0^1 \widetilde{G}_{\lambda^*}(z)dz = \Delta$. A standard duality argument then deduces that $G^*(\cdot) := \widetilde{G}_{\lambda^*}(\cdot)$ is an optimal solution to (11). Finally, an optimal solution to (10) is given by $R^*(z) = G^*(F_X(z))\ \forall z \in [0, M]$ and that to (6) by $I^*(z) = z - R^*(z)\ \forall z \in [0, M]$.

So our problem boils down to solving (13). However, in doing so the *convex* constraint that $0 \leqslant G'(z) \leqslant (F_X^{-1})'(z)$ in $\mathbb{G}$ poses the major difficulty compared with Bernard et al. (2015) in which the constraint is a convex cone.

3 | CHARACTERIZATION OF SOLUTIONS

In this section, we derive a necessary and sufficient condition for a function to be optimal to (13).

Assume $\widetilde{G}_\lambda(\cdot)$ solves (13) with a fixed λ. Let $G(\cdot) \in \mathbb{G}$ be arbitrary and fixed. For any $\varepsilon \in (0, 1)$, set $G^\epsilon(\cdot) = (1 - \epsilon)\widetilde{G}_\lambda(\cdot) + \epsilon G(\cdot)$. Then $G^\epsilon(\cdot) \in \mathbb{G}$. By the optimality of $\widetilde{G}_\lambda(\cdot)$ and the concavity of u, we have

$$
\begin{aligned}
0 \geqslant\ & \frac{1}{\varepsilon}\left\{\int_0^1 [u(W_\Delta - G^\epsilon(z))T'(z) + \lambda G^\epsilon(z)]dz - \int_0^1 [u(W_\Delta - \widetilde{G}_\lambda(z))T'(z) + \lambda \widetilde{G}_\lambda(z)]dz\right\} \\
=\ & \frac{1}{\varepsilon}\left\{\int_0^1 [(u(W_\Delta - G^\epsilon(z)) - u(W_\Delta - \widetilde{G}_\lambda(z)))T'(z) + \lambda(G^\epsilon(z) - \widetilde{G}_\lambda(z))]dz\right\} \\
\geqslant\ & \frac{1}{\varepsilon}\left\{\int_0^1 [(u'(W_\Delta - G^\epsilon(z)))(W_\Delta - G^\epsilon(z) - W_\Delta + \widetilde{G}_\lambda(z))T'(z) + \lambda(G^\epsilon(z) - \widetilde{G}_\lambda(z))]dz\right\} \\
\xrightarrow{\epsilon \downarrow 0}\ & \int_0^1 [(u'(W_\Delta - \widetilde{G}_\lambda(z)))(\widetilde{G}_\lambda(z) - G(z))T'(z) + \lambda(G(z) - \widetilde{G}_\lambda(z))]dz \\
=\ & \int_0^1 [u'(W_\Delta - \widetilde{G}_\lambda(z))T'(z) - \lambda](\widetilde{G}_\lambda(z) - G(z))dz. \qquad (14)
\end{aligned}
$$

Define

$$
N_\lambda(z) := -\int_z^1 [u'(W_\Delta - \widetilde{G}_\lambda(t))T'(t) - \lambda]dt, \quad z \in [0,1]. \qquad (15)
$$

Then (14) yields

$$
\begin{aligned}
0 \geqslant & \int_0^1 [u'(W_\Delta - \widetilde{G}_\lambda(z))T'(z) - \lambda](\widetilde{G}_\lambda(z) - G(z))dz = \int_0^1 \int_0^z (\widetilde{G}'_\lambda(t) - G'(t))dt\, dN_\lambda(z) \\
= & \int_0^1 \int_t^1 (\widetilde{G}'_\lambda(t) - G'(t))dN_\lambda(z)\, dt = \int_0^1 N_\lambda(t)(G'(t) - \widetilde{G}'_\lambda(t))dt,
\end{aligned}
$$

leading to

$$
\int_0^1 N_\lambda(z)G'(z)dz \leqslant \int_0^1 N_\lambda(z)\widetilde{G}'_\lambda(z)dz, \quad \forall\, G(\cdot) \in \mathbb{G}.
$$

In other words, $\widetilde{G}'_\lambda(\cdot)$ maximizes $\int_0^1 N_\lambda(z)G'(z)dz$ over $G(\cdot) \in \mathbb{G}$. Therefore, a necessary condition for $\widetilde{G}_\lambda(\cdot)$ to be optimal for (13) is

$$
\widetilde{G}'_\lambda(z) \overset{a.e.}{=} \begin{cases} 0, & \text{if } N_\lambda(z) = \int_z^1 [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(t))T'(t)]dt < 0, \\ \in [0, (F_X^{-1})'(z)], & \text{if } N_\lambda(z) = \int_z^1 [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(t))T'(t)]dt = 0, \\ (F_X^{-1})'(z), & \text{if } N_\lambda(z) = \int_z^1 [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(t))T'(t)]dt > 0. \end{cases} \qquad (16)
$$

It turns out that (16) completely characterizes the optimal solutions to (13).

Theorem 3.1. *A function $\widetilde{G}_\lambda(\cdot)$ is an optimal solution to (13) if and only if $\widetilde{G}_\lambda(\cdot) \in \mathbb{G}$ and $\widetilde{G}_\lambda(\cdot)$ satisfies (16).*

The above theorem establishes a general characterization result for the optimal solutions of (13). This result, however, is only implicit as an optimal $\widetilde{G}_\lambda(\cdot)$ appears on both sides of (16). Moreover, the derivative of $\widetilde{G}_\lambda(z)$ is undetermined when $N_\lambda(z) = 0$. In the next two sections, we will apply this general result to derive the solutions.

4 | MODEL WITH YAARI'S DUAL CRITERION

When $u(x) \equiv x$, the corresponding V^{rdu} reduces to the so-called Yaari's dual criterion (Yaari, 1987). Doherty and Eeckhoudt (1995) studied an optimal insurance problem under Yaari's dual criterion by considering only coinsurance or deductible contracts. In this section, by utilizing Theorem 3.1, we are able to solve our insurance problem under Yaari's criterion without restricting the set of contracts.

When $u(x) \equiv x$, the condition (16) is greatly simplified. Indeed, (16) reduces to

$$\widetilde{G}'_\lambda(z) \overset{a.e.}{=} \begin{cases} 0, & \text{if } \int_z^1 (\lambda - T'(t))dt = \lambda(1-z) - (1-T(z)) < 0, \\ \in [0, (F_X^{-1})'(z)], & \text{if } \int_z^1 (\lambda - T'(t))dt = \lambda(1-z) - (1-T(z)) = 0, \\ (F_X^{-1})'(z), & \text{if } \int_z^1 (\lambda - T'(t))dt = \lambda(1-z) - (1-T(z)) > 0. \end{cases} \tag{17}$$

It should be noted that although $u(x) \equiv x$ is not strictly concave here, the uniqueness of optimal solution to (13) is implied by the characterizing condition (17).

To apply (17), we need to compare λ and $\frac{1-T(z)}{1-z}$. Define

$$f(z) := \frac{1-T(z)}{1-z}, \quad z \in [0,1).$$

Lemma 4.1. *The function $f(\cdot)$ is a continuous function on [0,1). Moreover, under Assumption 2.3, there exists a unique $a \in (0,b)$ such that $f(\cdot)$ is strictly decreasing on $[0,a]$ and strictly increasing on $[a,1)$.*

Clearly, $f(0) = 1, f(1-) = +\infty$. Let a be defined as in Lemma 4.1. From the proof of Lemma 4.1, it is easily seen that a is uniquely determined by

$$T'(a) = \frac{1-T(a)}{1-a}. \tag{18}$$

Set

$$\hat{\lambda} := f(a) < f(0) = 1.$$

Let $c \in (a,1]$ be the unique scalar such that $f(c) = 1$, or equivalently, $T(c) = c$. See Figure 1 for the locations of the points a and c.

Now, we proceed by considering three cases based on the value of λ.

Case I $\lambda \leqslant \hat{\lambda}$. In this case,

$$N_\lambda(z) = (1-z)(\lambda - f(z)) < 0 \quad \forall z \in [0,a) \cup (a,1].$$

It then follows from (17) that $\widetilde{G}'_\lambda(z) \overset{a.e.}{=} 0$; hence $\widetilde{G}_\lambda(z) = 0 \ \forall \ z \in [0,1]$. Thus, the corresponding retention $\widetilde{R}_\lambda(z) = 0 \ \forall z \in [0,M]$ and indemnity $\widetilde{I}_\lambda(z) = z \ \forall z \in [0,M]$; namely, the optimal contract is a full insurance contract.

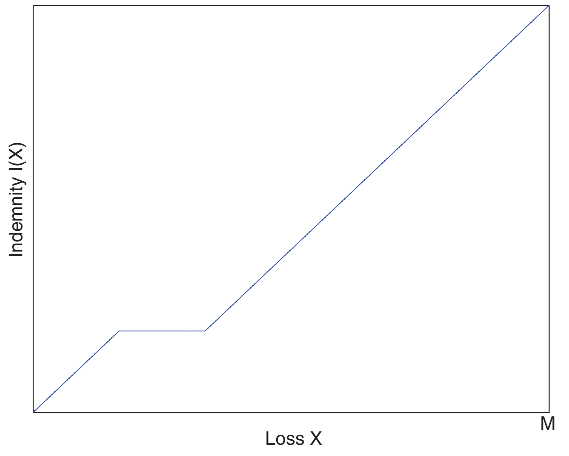

FIGURE 2 A schematic illustration of a threefold contract: it covers small losses as well as large losses in excess of a deductible [Color figure can be viewed at wileyonlinelibrary.com]

Case II $\hat{\lambda} < \lambda < 1$. By Lemma 4.1, there exist unique $x_0 \in (0, a)$ and $y_0 \in (a, c)$ such that $f(x_0) = f(y_0) = \lambda$. Accordingly, we have

$$N_\lambda(z) = \begin{cases} <0, & \text{if } 0 < z < x_0, \\ >0, & \text{if } x_0 < z < y_0, \\ <0, & \text{if } y_0 < z < 1. \end{cases}$$

Hence, (17) leads to the following function:

$$\widetilde{G}_\lambda(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < x_0, \\ F_X^{-1}(z) - F_X^{-1}(x_0), & \text{if } x_0 \leqslant z < y_0, \\ F_X^{-1}(y_0) - F_X^{-1}(x_0), & \text{if } y_0 \leqslant z \leqslant 1. \end{cases} \tag{19}$$

The corresponding retention and indemnity functions are, respectively,

$$\widetilde{R}_\lambda(z) \equiv \widetilde{G}_\lambda(F_X(z)) = \begin{cases} 0, & \text{if } 0 \leqslant z < F_X^{-1}(x_0), \\ z - F_X^{-1}(x_0), & \text{if } F_X^{-1}(x_0) \leqslant z < F_X^{-1}(y_0), \\ F_X^{-1}(y_0) - F_X^{-1}(x_0), & \text{if } F_X^{-1}(y_0) \leqslant z \leqslant M, \end{cases}$$

and

$$\widetilde{I}_\lambda(z) \equiv z - \widetilde{R}_\lambda(z) = \begin{cases} z, & \text{if } 0 \leqslant z < F_X^{-1}(x_0), \\ F_X^{-1}(x_0), & \text{if } F_X^{-1}(x_0) \leqslant z < F_X^{-1}(y_0), \\ z - F_X^{-1}(y_0) + F_X^{-1}(x_0), & \text{if } F_X^{-1}(y_0) \leqslant z \leqslant M. \end{cases} \tag{20}$$

The corresponding indemnity function is schematically illustrated by Figure 2. Qualitatively, the insurance covers not only large losses (when $z \geqslant F_X^{-1}(y_0)$) but also *small* losses (when $z < F_X^{-1}(x_0)$), and the compensation is a constant for the median range of losses. We term such a contract a *threefold* contract. The need for small loss coverage along with its connection to the probability weighting are amply discussed in Bernard et al. (2015). However, in Bernard et al. (2015), the IC constraint is violated in some range of the loss. Such a contract may incentivize the insured to misreport losses in order to get more compensations. In contrast, both our indemnity and retention are increasing functions of the loss, which will rule out this sort of moral hazard.

The threefold contract is nonlinear, and its flat part (from $X = F_X^{-1}(x_0)$ to $X = F_X^{-1}(y_0)$) implies that the insured needs to pay all expenses after the amount $F_X^{-1}(x_0)$ until the amount $F_X^{-1}(y_0)$ is reached. Hence, this part captures the so-called "coverage gap" or "donut hole" in medical drug insurance if X is interpreted as the cumulative drug expense in a year.[10] In recent years, many studies have investigated the impact of the nonlinear medical insurance contracts featuring coverage gaps. For example, Einav, Finkelstein, and Schrimpf (2015) studied how individuals' drug expenditures are influenced by the kink created by the donut hole. They further present a simple dynamic model to optimize the insured's decision on prescription drugs within a single-year period under such contracts. The donut hole feature also exists in other reimbursement models such as organ transplants; see Bajari, Hong, Park, and Town (2017) for a study on how such a feature can impact health providers' services. These nonlinear contracts are not exactly of the threefold type we have derived in this paper, yet the analogy in the part of the coverage gap is interesting enough to suggest that RDU may serve as a rationale to explain and/or justify the formers.

On the other hand, although an *exact* threefold contract has not yet been regularly seen in practice, it is actually a linear combination of the two more common types of contract: deductible and policy limit ($X \wedge F_X^{-1}(x_0)$). Any insured needing to have a threefold contract can use an insurance portfolio to produce it, even if it is not directly available in the market.

Case III $1 \leqslant \lambda < +\infty$. By Lemma 4.1, there exists a unique $z_0 \in [c, 1]$ such that $f(z_0) = \lambda$. Thus

$$N_\lambda(z) = \begin{cases} >0, & \text{if } 0 < z < z_0, \\ <0, & \text{if } z_0 < z < 1. \end{cases}$$

By (17), we have

$$\widetilde{G}_\lambda(z) = \begin{cases} F_X^{-1}(z), & \text{if } 0 \leqslant z < z_0, \\ F_X^{-1}(z_0), & \text{if } z_0 \leqslant z \leqslant 1. \end{cases} \tag{21}$$

So

$$\widetilde{I}_\lambda(z) \equiv z - \widetilde{R}_\lambda(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < F_X^{-1}(z_0), \\ z - F_X^{-1}(z_0), & \text{if } F_X^{-1}(z_0) \leqslant z \leqslant M. \end{cases} \tag{22}$$

This contract is a standard deductible contract in which only losses above a deductible point will be covered.

Define

$$\bar{G}(z) = \begin{cases} F_X^{-1}(z), & \text{if } 0 \leqslant z < c, \\ F_X^{-1}(c), & \text{if } c \leqslant z \leqslant 1, \end{cases} \tag{23}$$

and let

$$K_c := \int_0^1 \bar{G}(z)dz,$$

and

$$\pi_c := (1+\rho)(E[X]-K_c).$$

Clearly $K_c \leqslant \int_0^1 F_X^{-1}(z)dz = E[X]$.

We are now in the position to state our main result in terms of the premium π and the indemnity function $I(\cdot)$.

Theorem 4.2. *Under Yaari's criterion, $u(x) \equiv x$, and Assumptions 2.1 and 2.3, the optimal indemnity function $I^*(\cdot)$ to the problem (6) is given as follows.*

(i) *If $\pi = (1+\rho)E[X]$, then $I^*(z) = z\ \forall z \in [0, M]$.*

(ii) *If $\pi_c < \pi < (1+\rho)E[X]$, then*

$$I^*(z) = \begin{cases} z, & \text{if } 0 \leqslant z < F_X^{-1}(d), \\ F_X^{-1}(d), & \text{if } F_X^{-1}(d) \leqslant z < F_X^{-1}(e), \\ z - F_X^{-1}(e) + F_X^{-1}(d), & \text{if } F_X^{-1}(e) \leqslant z \leqslant M, \end{cases} \tag{24}$$

where (d, e) is the unique pair satisfying $0 \leqslant d < a < e \leqslant c$, $f(d) = f(e)$ and $E[I^*(X)] = \frac{\pi}{1+\rho}$.

(iii) *If $0 \leqslant \pi \leqslant \pi_c$, then*

$$I^*(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < F_X^{-1}(q), \\ z - F_X^{-1}(q), & \text{if } F_X^{-1}(q) \leqslant z \leqslant M, \end{cases} \tag{25}$$

where q is the unique scalar satisfying $c \leqslant q$ and $E[I^*(X)] = \frac{\pi}{1+\rho}$.

The economic interpretation of this result is clear. When the premium is small ($0 \leqslant \pi \leqslant \pi_c$), the contract only compensates large losses in excess of certain amount. When the premium is in middle range ($\pi_c < \pi < (1+\rho)E[X]$), the contract is a threefold one, covering both small and large losses. When the premium is sufficiently large ($\pi \geqslant (1+\rho)E[X]$), it is a full coverage.

It is interesting to investigate the comparative statics of the point π_c (in terms of c) that triggers the coverage for small losses. In fact, as

$$K_c = \int_0^c F_X^{-1}(z)dz + F_X^{-1}(c)(1-c),$$

we have

$$\frac{\partial K_c}{\partial c} = (1-c)(F_X^{-1})'(c).$$

However, $\pi_c = (1+\rho)(E[X]-K_c)$; hence

$$\frac{\partial \pi_c}{\partial c} = (1+\rho)(c-1)(F_X^{-1})'(c) < 0.$$

This implies that the insured is more willing to be protected against small losses if his weighting function has a bigger c. This is consistent with the fact that a bigger c renders a larger concave domain of the probability weighting that overweighs small losses (refer to Figure 1).

5 | MODEL WITH THE RDU CRITERION

In this section, we study the general RDU model in which the utility function is strictly concave. Compared with the Yaari model, solving the corresponding insurance problem calls for a more delicate analysis.

For any twice differentiable function f with $f'(x) \neq 0$, define its *Arrow-Pratt measure of absolute risk aversion*

$$A_f(x) := -\frac{f''(x)}{f'(x)}.$$

We now introduce the following assumptions.

Assumption 5.1 (Strictly concave utility). The utility function $u : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is strictly increasing and twice differentiable. Furthermore, u' is strictly decreasing.

Assumption 5.2.

(i) The function $A_u(z)$ is decreasing on $(0, \infty)$.

(ii) $A_T(z) > A_u(W - F_X^{-1}(z))(F_X^{-1})'(z), \quad \forall z \in (0, a]$.

Assumption 5.1 is to replace Assumption 2.2, ensuring a genuine RDU criterion. Assumption 5.2(i) requires that the absolute risk aversion measure of the utility function u is decreasing, which holds true for many frequently used utility functions including logarithmic, power, and exponential utilities. In general, experimental and empirical evidence is consistent with the decreasing absolute risk aversion (see, e.g., Friend & Blume, 1975). On the other hand, $A_T(z)$, $z \in (0, a]$ measures the level of probability weighting for small losses. The economical interpretation of Assumption 5.2(ii) is, therefore, that the degree of the insured's concern for small losses is sufficiently large relative to the absolute risk aversion of the utility function. Note that Assumption 5.2(ii) is automatically satisfied when $F_X^{-1}(z) = 0$, $\forall z \in [0, a]$, which is equivalent to $\mathbb{P}(X = 0) \geqslant a$. In practice, $\mathbb{P}(X = 0) \geqslant 0.5$ is a plausible assumption for many insurance products such as automobile and house insurance. On the other hand, a is very small for many commonly used inverse-S-shaped weighting functions. Take Tversky and Kahneman's weighting function (8) as an example, $a \approx 0.013$ when $\theta = 0.3$, $a \approx 0.07$ when $\theta = 0.5$, and $a \approx 0.166$ when $\theta = 0.8$. In these cases, Assumption 5.2(ii) holds automatically.

Problem (11) has trivial solutions in the following two cases. When $\Delta = 0$, the optimal solution is $G^*(z) = 0$ $\forall z \in [0, 1]$, corresponding to a full coverage. When $\Delta = E[X]$, the optimal solution is $G^*(z) = F_X^{-1}(z)$ $\forall z \in [0, 1]$ as it is the only feasible solution, corresponding to no coverage.

So we are interested in only the case $0 < \Delta < E[X]$. It follows from Proposition C.2 in Appendix C that there exists λ^* such that $\widetilde{G}_{\lambda^*}(\cdot)$ is an optimal solution to (13) under λ^* and $\int_0^1 \widetilde{G}_{\lambda^*}(z)dz = \Delta$. Furthermore, recall that we have proved that (13) has a unique solution when u is strictly concave and (16) provides the necessary and sufficient condition for the optimal solution.

Lemma 5.3. *For any $G(\cdot) \in \mathbb{G}$, if there exists $z \in (0, 1)$ such that*

$$\lambda - u'(W_\Delta - G(z))T'(z) = \int_z^1 [\lambda - u'(W_\Delta - G(t))T'(t)]dt = 0,$$

then $z \leqslant a$.

Lemma 5.4. *Under Assumption 5.2, for any $G(\cdot) \in \mathbb{G}$, $u'(W_\Delta - G(z))T'(z)$ is a strictly decreasing function of z on $[0, a]$.*

Now, for any $\lambda \leqslant \hat{\lambda} u'(W_\Delta)$, we have

$$\int_z^1 [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(t))T'(t)]dt \leqslant \int_z^1 [\hat{\lambda} u'(W_\Delta) - u'(W_\Delta)T'(t)]dt$$

$$= u'(W_\Delta)\int_z^1 [\hat{\lambda} - T'(t)]dt = u'(W_\Delta)(1-z)\left[\hat{\lambda} - \frac{1-T(z)}{1-z}\right] < 0,$$

where the last inequality is due to Lemma 4.1. Hence $\widetilde{G}_\lambda(z) = 0\ \forall z \in [0,1]$ is the only solution satisfying (16). However, $\int_0^1 \widetilde{G}_\lambda(z)dz = 0 < \Delta$, a contradiction. Therefore, only when $\lambda > \hat{\lambda} u'(W_\Delta)$ it is possible for (16) to hold.

Fixing $\lambda > \hat{\lambda} u'(W_\Delta)$, we now analyze the shape of the function $\widetilde{G}_\lambda(\cdot)$ that satisfies (16). Suppose that $\widetilde{G}_\lambda(1) = k < W_\Delta$. We then have $N_\lambda(1) = 0$ and $\lambda - u'(W_\Delta - k)T'(1-) < 0$ as $T'(1-) = +\infty$. So, $\widetilde{G}'_\lambda(z) = 0$ when z is close to 1 as $N_\lambda(z) < 0$ for such z. Hence, $\widetilde{G}_\lambda(z) \equiv k\ \forall z \in [z_1, 1]$ for some $z_1 \in [0,1)$, at which $N_\lambda(z_1) = 0$ and $N_\lambda(z) < 0$ for $\forall z \in (z_1, 1)$. Next, we consider three cases, respectively, depending on the value of k and location of z_1.

Case (A) $k > W_\Delta - (u')^{-1}(\frac{\lambda}{\hat{\lambda}})$, that is, $\lambda < \hat{\lambda} u'(W_\Delta - k)$. In this case, we have, for any $z \in [0,1)$,

$$\int_z^1 [\lambda - u'(W_\Delta - k)T'(t)]dt < \int_z^1 [\hat{\lambda} u'(W_\Delta - k) - u'(W_\Delta - k)T'(t)]dt$$

$$= u'(W_\Delta - k)(1-z)\left[\hat{\lambda} - \frac{1-T(z)}{1-z}\right] \leqslant 0.$$

It then follows from (16) that $\widetilde{G}_\lambda(z) \equiv k = \widetilde{G}_\lambda(0) = 0$. However, $0 = k > W_\Delta - (u')^{-1}(\frac{\lambda}{\hat{\lambda}})$, or $\lambda \leqslant \hat{\lambda} u'(W_\Delta)$, leading to a contradiction. So, this case in fact will not take place.

Case (B) $k = W_\Delta - (u')^{-1}(\frac{\lambda}{\hat{\lambda}})$. In this situation, z_1 should be a. This is because $\int_a^1 [\lambda - u'(W_\Delta - k)T'(t)]dt = 0$ and

$$\int_z^1 [\lambda - u'(W_\Delta - k)T'(t)]dt = \frac{\lambda}{\hat{\lambda}}(1-z)(\hat{\lambda} - \frac{1-T(z)}{1-z}) < 0$$

for $z \in (a,1)$ by Lemma 4.1. Moreover, $\lambda - u'(W_\Delta - k)T'(a) = 0$. By Lemma 5.4, $\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z))T'(z)$ strictly increases with respect to $z \in [0,a]$. It follows that

$$\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z))T'(z) < 0$$

for $z \in [0,a)$. Then (16) implies $\widetilde{G}'_\lambda(z) = 0$ for $z \in (0,a)$. As a result, $k = \widetilde{G}_\lambda(a) = \widetilde{G}_\lambda(0) = 0$, or $\lambda = \hat{\lambda} u'(W_\Delta)$, which is a contradiction. So, again, this case will not occur.

Case (C) $k < W_\Delta - (u')^{-1}(\frac{\lambda}{\hat{\lambda}})$. In this case, $z_1 \in (a,1)$ exists. By Lemma 5.3, we have $\lambda - u'(W_\Delta - k)T'(z_1) > 0$. Hence, there may or may not exist $z_2 \in (0,1)$ such that $N_\lambda(z_2) = 0$ and $N_\lambda(z) > 0$ for $z \in (z_2, z_1)$. We now discuss four subcases depending on the existence and location of z_2.

(C.1) If z_2 does not exist or $z_2 = 0$ (i.e., $N_\lambda(z) > 0$ for $z \in (0, z_1)$), then by (16), $\widetilde{G}'_\lambda(z) = (F_X^{-1})'(z)$ for $z \in (0, z_1)$. Combined with the fact that $\widetilde{G}_\lambda(0) = 0$, we have

$$\widetilde{G}_\lambda(z) = \begin{cases} F_X^{-1}(z), & \text{if } 0 \leqslant z < z_1, \\ F_X^{-1}(z_1), & \text{if } z_1 \leqslant z \leqslant 1. \end{cases}$$

This corresponds to a deductible contract.

(C.2) If z_2 exists and $z_2 \in (0, a]$, then $\widetilde{G}'_\lambda(z) = (F_X^{-1})'(z)$ for $z \in (z_2, z_1)$ in view of (16). Combining the property of z_1 and z_2, we deduce

$$\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z_2) \leqslant 0.$$

Then, using Lemma 5.4, we have

$$\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z))T'(z) < 0$$

for $z \in [0, z_2)$. It follows from (16) that $\widetilde{G}'_\lambda(z) = 0$ for $z \in (0, z_2)$. In this case, we can express $\widetilde{G}_\lambda(\cdot)$ as follows:

$$\widetilde{G}_\lambda(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < z_2, \\ F_X^{-1}(z) - F_X^{-1}(z_2), & \text{if } z_2 \leqslant z < z_1, \\ F_X^{-1}(z_1) - F_X^{-1}(z_2), & \text{if } z_1 \leqslant z \leqslant 1. \end{cases}$$

This is the threefold contract, schematically depicted in Figure 2.

(C.3) If z_2 exists and $z_2 \in (b, 1)$ (recall that b is the turning point where the weighting function $T(\cdot)$ changes from being concave to convex), then a similar analysis as in Case (B) shows that $\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_1))T'(z_1) > 0$ and $\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z_2) < 0$. This means

$$u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z_2) > u'(W_\Delta - \widetilde{G}_\lambda(z_1))T'(z_1).$$

However, $u'(W_\Delta - \widetilde{G}_\lambda(z_1)) \geqslant u'(W_\Delta - \widetilde{G}_\lambda(z_2)) > 0$ and $T'(z_1) > T'(z_2) > 0$, which is a contradiction. So, this case is not feasible.

(C.4) If z_2 exists and $z_2 \in (a, b]$, then $\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z_2) < 0$. We prove $\widetilde{G}_\lambda(z) \equiv \widetilde{G}_\lambda(z_2)$ $\forall z \in [0, z_2]$. In fact, if it is false, then there exists z_3 such that

$$\int_{z_3}^{z_2} [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(t)]dt = 0 \quad \text{and} \quad \int_{z}^{z_2} [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(t)]dt < 0$$

for $z \in (z_3, z_2)$. However,

$$\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z) < \lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z_2) < 0$$

for $z \in (z_3, z_2)$ as $z_2 \in (a, b]$. So,

$$\int_{z_3}^{z_2} [\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(t)]dt < (\lambda - u'(W_\Delta - \widetilde{G}_\lambda(z_2))T'(z_2))(z_2 - z_3) < 0,$$

arriving at a contradiction. Therefore, $k = F_X^{-1}(z_1) - F_X^{-1}(z_2)$. From $\int_{z_1}^{1}[\lambda - u'(W_\Delta - k)T'(t)]dt = 0$, it follows that

$$\lambda = u'(W_\Delta - k)\frac{1-T(z_1)}{1-z_1} = u'(W_\Delta + F_X^{-1}(z_2) - F_X^{-1}(z_1))\frac{1-T(z_1)}{1-z_1}.$$

However,

$$\begin{aligned}
&\int_{z_2}^{z_1}\left[u'(W_\Delta + F_X^{-1}(z_2) - F_X^{-1}(z_1))\frac{1-T(z_1)}{1-z_1} - u'(W_\Delta + F_X^{-1}(z_2) - F_X^{-1}(t))T'(t)\right]dt\\
>&\int_{z_2}^{z_1}\left[u'(W_\Delta + F_X^{-1}(z_2) - F_X^{-1}(z_1))\frac{1-T(z_1)}{1-z_1} - u'(W_\Delta + F_X^{-1}(z_2) - F_X^{-1}(z_1))T'(t)\right]dt\\
=&\ u'(W_\Delta + F_X^{-1}(z_2) - F_X^{-1}(z_1))(z_1-z_2)\left[\frac{1-T(z_1)}{1-z_1} - \frac{T(z_1)-T(z_2)}{z_1-z_2}\right] > 0,
\end{aligned}$$

where the last inequality follows from Lemma A.1(ii) in Appendix A. This is a contradiction. So, the current case will not occur either.

To summarize, for any $\lambda > \hat{\lambda}u'(W_\Delta)$, only deductible and threefold contracts are possibly optimal, stipulated by (C.1) and (C.2). Next, we investigate these two cases more closely.

Define a function $h_\Delta(\cdot)$ on $[a, c]$ as follows:

$$h_\Delta(z) := \int_0^z \left[\frac{u'(W_\Delta - F_X^{-1}(z))(1-T(z))}{1-z} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right]dt. \tag{26}$$

Then, by (18) and using Lemma 5.4, we have

$$h_\Delta(a) = \int_0^a \left[u'\left(W_\Delta - F_X^{-1}(a)\right)T'(a) - u'\left(W_\Delta - F_X^{-1}(t)\right)T'(t)\right]dt < 0.$$

Recalling that $T(c) = c$, we have

$$\begin{aligned}
h_\Delta(c) &= \int_0^c \left[\frac{u'(W_\Delta - F_X^{-1}(c))(1-T(c))}{1-c} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right]dt\\
&= \int_0^c \left[u'(W_\Delta - F_X^{-1}(c)) - u'(W_\Delta - F_X^{-1}(t))T'(t)\right]dt\\
&> u'(W_\Delta - F_X^{-1}(c))c - \int_0^c \left[u'(W_\Delta - F_X^{-1}(c))T'(t)\right]dt = 0.
\end{aligned}$$

Moreover, we take the derivative of $h_\Delta(z)$ with respect to $z \in [a, c]$ to obtain

$$\begin{aligned}
h'_\Delta(z) =& -u'\left(W_\Delta - F_X^{-1}(z)\right)T'(z) + u'\left(W_\Delta - F_X^{-1}(z)\right)\frac{1-T(z)}{1-z}\\
&- u''\left(W_\Delta - F_X^{-1}(z)\right)\frac{1-T(z)}{1-z}z\left(F_X^{-1}\right)'(z) + u'\left(W_\Delta - F_X^{-1}(z)\right)z\frac{\frac{1-T(z)}{1-z} - T'(z)}{1-z}\\
=&\ u'\left(W_\Delta - F_X^{-1}(z)\right)\left(\frac{1-T(z)}{1-z} - T'(z)\right) - u''\left(W_\Delta - F_X^{-1}(z)\right)\frac{1-T(z)}{1-z}z\left(F_X^{-1}\right)'(z)
\end{aligned}$$

$$+\,u'\left(W_\Delta - F_X^{-1}(z)\right) z\frac{\frac{1-T(z)}{1-z} - T'(z)}{1-z} > 0.$$

Hence, there exists a unique point $l_\Delta \in (a, c)$ such that

$$h_\Delta(z)\begin{cases} < 0, & \text{if } a \leqslant z < l_\Delta, \\ = 0, & \text{if } z = l_\Delta, \\ > 0, & \text{if } l_\Delta < z \leqslant c. \end{cases} \tag{27}$$

Define

$$\underline{G}(z) = \begin{cases} F_X^{-1}(z), & \text{if } 0 \leqslant z < l_\Delta, \\ F_X^{-1}(l_\Delta), & \text{if } l_\Delta \leqslant z \leqslant 1, \end{cases} \tag{28}$$

and $K_\Delta := \int_0^1 \underline{G}(z)dz$.

Proposition 5.5. *If $K_\Delta \leqslant \Delta < E[X]$, then the optimal solution to (11) is*

$$G^*(z) = \begin{cases} F_X^{-1}(z), & \text{if } 0 \leqslant z < f, \\ F_X^{-1}(f), & \text{if } f \leqslant z \leqslant 1, \end{cases} \tag{29}$$

where f is the unique scalar such that $f \geqslant l_\Delta$ and $\int_0^1 G^*(z)dz = \Delta$.

Lemma 5.6. *If $0 < \Delta < K_\Delta$, then the corresponding optimal contract is not deductible.*

It follows from Lemma 5.6 that if $0 < \Delta < K_\Delta$, the optimal contract (which always exists) can only be threefold, corresponding to (C.2). We are now led to the following proposition.

Proposition 5.7. *If $0 < \Delta < K_\Delta$, then the optimal solution to (11) is given as*

$$G^*(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < z_2, \\ F_X^{-1}(z) - F_X^{-1}(z_2), & \text{if } z_2 \leqslant z < z_1, \\ F_X^{-1}(z_1) - F_X^{-1}(z_2), & \text{if } z_1 \leqslant z \leqslant 1, \end{cases}$$

where z_1, z_2 satisfy $z_2 \leqslant a \leqslant z_1$,

$$\int_{z_2}^{z_1} \left[\frac{u'(W_\Delta - F_X^{-1}(z_1) + F_X^{-1}(z_2))(1 - T(z_1))}{1 - z_1} - u'(W_\Delta - F_X^{-1}(t) + F_X^{-1}(z_2))T'(t)\right] dt = 0$$

and

$$\int_0^1 G^*(z)dz = \Delta.$$

Note that any pair (z_2, z_1) satisfying the requirements in Proposition 5.7 leads to an optimal solution to (11). Therefore such a pair (z_2, z_1) is unique as the optimal solution to (11) is unique.

Propositions 5.5 and 5.7 give two qualitatively distinct optimal contracts for any given $0 < \Delta < E[X]$, and the two cases are divided depending on whether or not $\Delta < K_\Delta$. However, K_Δ in general depends on Δ in an implicit and complicated way, so it is hard to compare Δ and K_Δ.

Nevertheless, we are able to treat at least two cases where $A_u(z)$ is either a constant or strictly decreasing in z.

First, assume that the utility function exhibits constant absolute risk aversion, that is, $u(z) = 1 - e^{-\alpha z}$ $\forall z \in \mathbb{R}^+$. Then it is easy to see from (26) that l_Δ is independent of Δ, and hence so is K_Δ. In this case, denote $K \equiv K_\Delta$ and $\hat{\pi} = (1+\rho)(E[X] - K)$. Then we have the following result.

Theorem 5.8. *Assume that Assumptions 2.1, 2.3, and 5.2 hold, and that $u(\cdot)$ exhibits constant absolute risk aversion. Then the optimal indemnity function $I^*(\cdot)$ to the problem (6) is given as follows.*

(i) *If $\pi = (1+\rho)E[X]$, then $I^*(z) = z$ for $z \in [0, M]$.*

(ii) *If $\hat{\pi} < \pi < (1+\rho)E[X]$, then*

$$I^*(z) = \begin{cases} z, & \text{if } 0 \leqslant z < F_X^{-1}(z_2), \\ F_X^{-1}(z_2), & \text{if } F_X^{-1}(z_2) \leqslant z < F_X^{-1}(z_1), \\ z - F_X^{-1}(z_1) + F_X^{-1}(z_2), & \text{if } F_X^{-1}(z_1) \leqslant z \leqslant M, \end{cases}$$

where (z_2, z_1) is the unique pair satisfying $z_2 \leqslant a \leqslant z_1$,

$$\int_{z_2}^{z_1} \left[\frac{u'(W_\Delta - F_X^{-1}(z_1) + F_X^{-1}(z_2))(1 - T(z_1))}{1 - z_1} - u'\left(W_\Delta - F_X^{-1}(t) + F_X^{-1}(z_2)\right) T'(t) \right] dt = 0,$$

and $E[I^*(X)] = \frac{\pi}{1+\rho}$.

(iii) *If $0 \leqslant \pi \leqslant \hat{\pi}$, then*

$$I^*(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < F_X^{-1}(f), \\ z - F_X^{-1}(f), & \text{if } F_X^{-1}(f) \leqslant z \leqslant M, \end{cases}$$

where f is the unique scalar satisfying $E[I^*(X)] = \frac{\pi}{1+\rho}$.

Now, we study the case in which $A_u(z)$ is strictly decreasing. We need the following lemma.

Lemma 5.9. *If $0 < \Delta_1 < \Delta_2 < E[X]$, then $a < l_{\Delta_1} < l_{\Delta_2} < c$.*

Define

$$\Delta(d) := \int_0^d F_X^{-1}(z)dz + \int_d^1 F_X^{-1}(d)dz = \int_0^d F_X^{-1}(z)dz + F_X^{-1}(d)(1-d)$$

on $d \in [a, c]$. Then $\Delta'(d) = (1-d)(F_X^{-1})'(d) > 0$. Hence, $\Delta(\cdot)$ is a continuous and strictly increasing function. Determine $l_{\Delta(a)}$ and $l_{\Delta(c)}$ by $h_{\Delta(a)}(l_{\Delta(a)}) = 0$ and $h_{\Delta(c)}(l_{\Delta(c)}) = 0$, and set $\widetilde{\Delta} := \Delta(l_{\Delta(a)})$ and $\overline{\Delta} := \Delta(l_{\Delta(c)})$. Finally, define a function $g(\cdot)$ on $[a, c]$ as follows:

$$g(z) := \int_0^z \left[\frac{u'(W_0 + (1+\rho)\Delta(z) - F_X^{-1}(z))(1 - T(z))}{1 - z} - u'\left(W_0 + (1+\rho)\Delta(z) - F_X^{-1}(t)\right) T'(t) \right] dt.$$

Now, we are ready to give the main result in terms of the premium π and the indemnity function $I(\cdot)$.

Theorem 5.10. *Assume that Assumptions 2.1, 2.3, and 5.2 hold, and that $A_u(\cdot)$ is strictly decreasing. Then the optimal indemnity function $I^*(\cdot)$ to the problem (6) is given as*

(i) *If* $\pi = (1+\rho)E[X]$, *then* $I^*(z) = z \; \forall z \in [0, M]$.

(ii) *If* $(1+\rho)(E[X] - \widetilde{\Delta}) \leqslant \pi < (1+\rho)E[X]$, *then*

$$I^*(z) = \begin{cases} z, & \text{if } 0 \leqslant z < F_X^{-1}(z_2), \\ F_X^{-1}(z_2), & \text{if } F_X^{-1}(z_2) \leqslant z < F_X^{-1}(z_1), \\ z - F_X^{-1}(z_1) + F_X^{-1}(z_2), & \text{if } F_X^{-1}(z_1) \leqslant z \leqslant M, \end{cases}$$

where (z_2, z_1) *is the unique pair satisfying* $z_2 \leqslant a \leqslant z_1$,

$$\int_{z_2}^{z_1} \left[\frac{u'(W_\Delta - F_X^{-1}(z_1) + F_X^{-1}(z_2))(1 - T(z_1))}{1 - z_1} - u'\left(W_\Delta - F_X^{-1}(t) + F_X^{-1}(z_2)\right) T'(t) \right] dt = 0,$$

and $E[I^*(X)] = \frac{\pi}{1+\rho}$.

(iii) *If* $(1+\rho)(E[X] - \overline{\Delta}) < \pi < (1+\rho)(E[X] - \widetilde{\Delta})$, *then let* $p \in (l_{\Delta(a)}, l_{\Delta(c)})$ *such that* $\Delta(p) = E[X] - \frac{\pi}{1+\rho}$. *If* $g(p) < 0$, *then*

$$I^*(z) = \begin{cases} z, & \text{if } 0 \leqslant z < F_X^{-1}(z_2), \\ F_X^{-1}(z_2), & \text{if } F_X^{-1}(z_2) \leqslant z < F_X^{-1}(z_1), \\ z - F_X^{-1}(z_1) + F_X^{-1}(z_2), & \text{if } F_X^{-1}(z_1) \leqslant z \leqslant M, \end{cases}$$

where (z_2, z_1) *is the unique pair satisfying* $z_2 \leqslant a \leqslant z_1$,

$$\int_{z_2}^{z_1} \left[\frac{u'\left(W_\Delta - F_X^{-1}(z_1) + F_X^{-1}(z_2)\right)(1 - T(z_1))}{1 - z_1} - u'\left(W_\Delta - F_X^{-1}(t) + F_X^{-1}(z_2)\right) T'(t) \right] dt = 0,$$

and $E[I^*(X)] = \frac{\pi}{1+\rho}$. *If* $g(p) \geqslant 0$, *then*

$$I^*(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < F_X^{-1}(f), \\ z - F_X^{-1}(f), & \text{if } F_X^{-1}(f) \leqslant z \leqslant M, \end{cases}$$

where q *is the unique number satisfying* $f < l_{\Delta(c)}$ *and* $E[I^*(X)] = \frac{\pi}{1+\rho}$.

(iv) *If* $0 \leqslant \pi \leqslant (1+\rho)(E[X] - \overline{\Delta})$, *then*

$$I^*(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < F_X^{-1}(f), \\ z - F_X^{-1}(f), & \text{if } F_X^{-1}(f) \leqslant z \leqslant M, \end{cases}$$

where q *is the unique number satisfying* $f \geqslant l_{\Delta(c)}$ *and* $E[I^*(X)] = \frac{\pi}{1+\rho}$.

6 | NUMERICAL EXAMPLES

In this section, we use numerical examples to illustrate our result with varying levels of the premium. We take the same numerical setting as in Bernard et al. (2015) for the comparison purpose (except

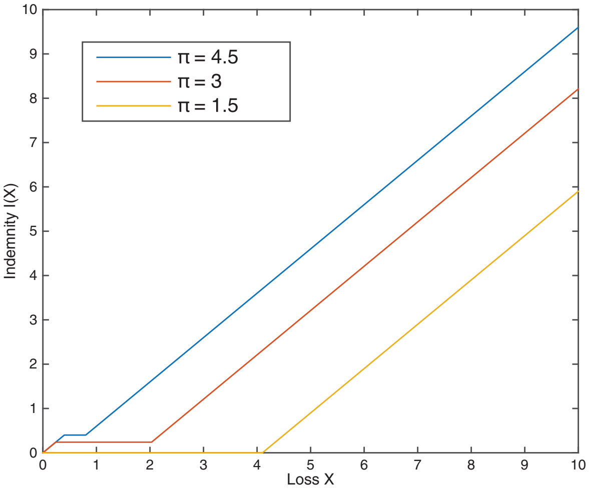

FIGURE 3 Optimal contracts under different premiums: The higher the premiums, the lower the deductibles and the more smaller losses covered [Color figure can be viewed at wileyonlinelibrary.com]

for the value of the premium; see next). The loss X follows a truncated exponential distribution with density function

$$f(x) = \frac{me^{-mx}}{1 - e^{-mM}},$$

where the intensity parameter $m = 0.1$, and $M = 10$. The initial wealth $W_0 = 15$, and $u(x) = 1 - e^{-\gamma x}$ with $\gamma = 0.02$. Moreover, the safety loading of the insurer ρ is 0.2. Finally, the weighting function is

$$T_\theta(x) = \frac{x^\theta}{(x^\theta + (1-x)^\theta)^{\frac{1}{\theta}}}$$

with $\theta = 0.5$. We can verify that the assumptions of Theorem 5.8 are satisfied under this setting. In Bernard et al. (2015), the premium is fixed at $\pi = 3$, but here we compute the optimal indemnities under $\pi = 1.5, 3$, and 4.5, respectively. These are plotted in Figure 3. The contract corresponding to $\pi = 1.5$ is a deductible one, whereas those corresponding to the two higher premiums are threefold covering smaller as well larger losses. For the latter two contracts, the one with the higher premium covers more smaller losses and has a lower deductible. Clearly, these features are all intuitive and sensible.

When $\pi = 3$, the optimal indemnity obtained by Bernard et al. (2015) (without the monotone constraint) is plotted in Figure 4 (dashed line). We note that in the result of Bernard et al. (2015), in some range of the loss, the insured has the incentive to hide part of the loss in order to be paid with a larger compensation. By contrast, our indemnity function (solid line) is increasing and any increment in compensations is always less than or equal to the increment in losses. It effectively rules out the aforementioned behavior of moral hazard. In addition, in this example, we calculate the optimal RDU value of Bernard et al. (2015) to be 0.19 and that of our model to be 0.187. The difference, 0.003, is the "cost" of the additional monotonicity constraint, that is, the loss in RDU value compared to the unconstrained case.

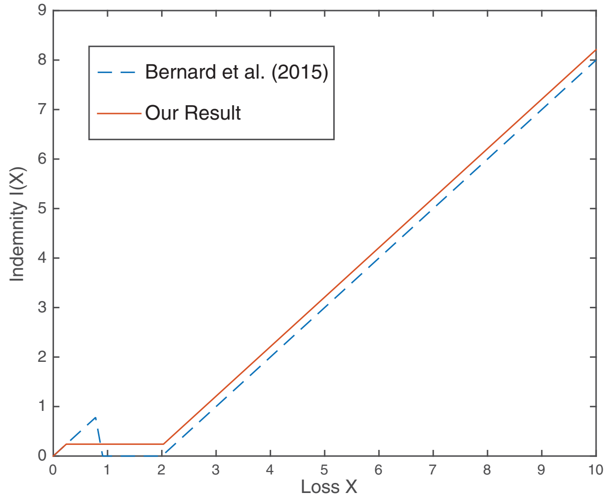

FIGURE 4 A comparison between our contract and Bernard et al. (2015): Our contract is a monotone, threefold contract, whereas theirs has a decreasing part causing potential moral hazard [Color figure can be viewed at wileyonlinelibrary.com]

7 | CONCLUSION

In this paper, we have studied an optimal insurance design problem where the insured uses the RDU preference. There is documented evidence proving that this preference captures human behaviors better than the EU preference. The main contribution of our work is that our optimal contracts are monotone with respect to losses, thereby eliminating the potential problem of moral hazard associated with the existing results.

An interesting conclusion from our results is that, under our assumptions (in particular, Assumption 5.2(ii)), there are only two types of nontrivial optimal contracts possible, one being the classical deductible and the other the threefold contract covering both small and large losses. On the other hand, while we have demonstrated that Assumption 5.2(ii) holds for many economically interesting cases, removing this assumption remains a mathematically outstanding open problem.

ACKNOWLEDGMENTS

We thank the two referees for constructive comments that have led to a substantially improved version. Xu acknowledges financial supports from NSFC (No. 11471276), Hong Kong GRF (Nos. 15204216 and 15202817), and the Hong Kong Polytechnic University. Zhou acknowledges supports from a start-up grant at Columbia University and from Oxford–Nie Financial Big Data Lab.

ENDNOTES

[1] This assumption is motivated by the fact that an insurer typically has many independent insured as its clients; hence, its risk is adequately diversified.

[2] Throughout this paper, by an “increasing” function we mean a “nondecreasing” function; namely, f is increasing if $f(x) \geqslant f(y)$ whenever $x > y$. We say f is “strictly increasing” if $f(x) > f(y)$ whenever $x > y$. Similar conventions are used for “decreasing” and “strictly decreasing” functions.

[3] A more recent paper, Bhargava, Loewenstein, and Sydnor (2017), found that in the context of health insurance, factors beyond preference specification such as behavioral aspects, insurance contracts complexity, or even low health literacy drive consumers to choose dominated policies.

[4] On the other hand, the RDU preference reduces to Yaari's dual criterion (Yaari, 1987) when the utility function is $u(x) = x$.

[5] This constraint is closely related to the general "revelation principle" in economics (see, e.g., Dasgupta, Hammond, & Maskin, 1979; Harris & Townsend, 1981; Myerson, 1979, for the study of this principle in different contexts). This principle dictates that a mechanism designer should design a mechanism in such a way that agents are willing to reveal their hidden information. Winter (2013) discussed the application of the revelation principle to insurance design problems.

[6] Picard (2000) gave several examples of ex post moral hazard, one example being a firm lets stocks in warehouse burn so as to pocket the insurance compensation. In many cases, a loss creation or increase can be made thanks to the help of a middleman. For instance, a physician increases the prescription drugs or the hospital charges just to exceed the threshold of the insurance policy of his or her patient. Car repairers or attorneys also may be in a position that allows them to inflate the cost of road accidents or casualties.

[7] Calculus of variations has also been applied in the insurance context. For example, Spence and Zeckhauser (1971) employed calculus of variations to solve an insurance contracting problem in the setting of EU theory.

[8] A mixed discrete/continuous distribution is the one whose CDF is a linear combination of those of a continuous distribution and a discrete distribution.

[9] Smith (1968) pointed out that it would be more interesting to consider X having atoms especially at 0 and M.

[10] For example, the 2008 U.S. government defined standard benefit design is a contract that contains the donut hole. In this plan, an individual initially pays all drug expenses up to $275 and then pay 25% of all expenses after $275 up to $2,510. After $2,510, there is a coverage gap in which the individual needs to pay all expenses until $5,726 is reached. Above $5,726, the individual just pays 7% of the cost. Such a contract has a nonlinear, "fourfold" structure. See also https://www.medicare.gov/part-d/costs/coverage-gap/part-d-coverage-gap.html

REFERENCES

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, *21*(4), 503–546.

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

Arrow, K. (1971). *Essays in the theory of risk-bearing*. Amsterdam, Netherlands: North-Holland.

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

Bajari, P., Hong, H., Park, M., & Town, R. (2017). Estimating price sensitivity of economic agents using discontinuity in nonlinear contracts. *Quantitative Economics*, *8*(2), 397–433.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, *27*(1), 173–195.

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, *103*(6), 2499–2529.

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

Bhargava, S., Loewenstein, G., & Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *Quarterly Journal of Economics*, *132*(3), 1319–1372.

Butler, R. J., Durbin, D. L., & Helvacian, N. M. (1996). Increasing claims for soft tissue injuries in workers' compensation: Cost shifting and moral hazard. *Journal of Risk and Uncertainty*, *13*(1), 73–87.

Carlier, G., & Dana, R.-A. (2008). Two-persons efficient risk-sharing and equilibria for concave law-invariant utilities. *Economic Theory*, *36*(2), 189–223.

Chateauneuf, A., Dana, R.-A., & Tallon, J.-M. (2000). Optimal risk-sharing rules and equilibria with choquet-expected-utility. *Journal of Mathematical Economics*, *34*(2), 191–214.

Cummins, J. D., & Tennyson, S. (1996). Moral hazard in insurance claiming: Evidence from automobile insurance. *Journal of Risk and Uncertainty*, *12*(1), 29–50.

Dana, R.-A., & Scarsini, M. (2007). Optimal risk sharing with background risk. *Journal of Economic Theory*, *133*(1), 152–176.

Dasgupta, P., Hammond, P., & Maskin, E. (1979). The implementation of social choice rules: Some general results on incentive compatibility. *Review of Economic Studies*, *46*(2), 185–216.

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

Doherty, N. A., & Eeckhoudt, L. (1995). Optimal insurance without expected utility: The dual theory and the linearity of insurance contracts. *Journal of Risk and Uncertainty*, *10*(2), 157–179.

Doherty, N. A., & Posey, L. (1997). Availability crises in insurance markets: Optimal contracts with asymmetric information and capacity constraints. *Journal of Risk and Uncertainty*, *15*(1), 55–80.

Einav, L., Finkelstein, A., & Schrimpf, P. (2015). The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D. *Quarterly Journal of Economics*, *130*(2), 841–899.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*(4), 643–669.

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, *56*(4), 279–304.

Friend, I., & Blume, M. (1975). The demand for risky assets. *American Economic Review*, *65*(5), 900–922.

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

Harris, M., & Townsend, R. M. (1981). Resource allocation under asymmetric information. *Econometrica: Journal of the Econometric Society*, *49*(1), 33–64.

He, X. D., & Zhou, X. Y. (2011). Portfolio choice via quantiles. *Mathematical Finance*, *21*(2), 203–231.

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

Jin, H., & Zhou, X. (2008). Behavioral portfolio selection in continuous time. *Mathematical Finance*, *18*(3), 385–426.

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

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, *15*(2), 145–161.

Myerson, R. B. (1979). Incentive compatibility and the bargaining problem. *Econometrica: Journal of the Econometric Society*, *47*(1), 61–73.

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

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, *3*(4), 323–343.

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

Shavell, S. (1979). On moral hazard and insurance. *Quarterly Journal of Economics*, *93*(4), 280–301.

Smith, V. L. (1968). Optimal insurance coverage. *Journal of Political Economy*, *76*(1), 68–77.

Spence, M., & Zeckhauser, R. (1971). Insurance, information, and individual action. *American Economic Review*, *61*(2), 380–387.

Sydnor, J. (2010). (Over) insuring modest risks. *American Economic Journal: Applied Economics*, *2*(4), 177–199.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, *21*(2), 265–293.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

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

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, *55*(1), 95–115.

How to cite this article: Xu ZQ, Zhou XY, Zhuang SC. Optimal insurance under rank-dependent utility and incentive compatibility. *Mathematical Finance*. 2018;1–34. https://doi.org/10.1111/mafi.12185

APPENDIX A: TWO LEMMAS

In this part, we prove some lemmas which have been used in Section 5.

Lemma A.1. *Assume* $T(\cdot) : [0,1] \mapsto [0,1]$ *satisfies Assumption 2.3. Then*

(i) *if* $a < z$, *then* $T'(z) < \frac{1-T(z)}{1-z}$;

(ii) *if* $a \leqslant z_2 < z_1 < 1$, *then* $\frac{1-T(z_1)}{1-z_1} > \frac{T(z_1)-T(z_2)}{z_1-z_2}$.

Proof.

(i) If $a < z \leqslant b$, then $T'(z) < T'(a) < \frac{1-T(z)}{1-z}$. If $b < z$, then $T'(z) < \frac{1-T(z)}{1-z}$ as $T(\cdot)$ is convex and strictly increasing on $[b, 1]$.

(ii) As $1 - T(z_1)$, $1 - z_1$, $T(z_1) - T(z_2)$, and $z_1 - z_2$ are all strictly positive, we have

$$\frac{1-T(z_1)}{1-z_1} > \frac{T(z_1)-T(z_2)}{z_1-z_2} \Longleftrightarrow \frac{1-T(z_1)}{1-z_1} > \frac{(1-T(z_1))+(T(z_1)-T(z_2))}{(1-z_1)+(z_1-z_2)}$$
$$\Longleftrightarrow \frac{1-T(z_1)}{1-z_1} > \frac{1-T(z_2)}{1-z_2}.$$

However, $\frac{1-T(z_1)}{1-z_1} > \frac{1-T(z_2)}{1-z_2}$ follows from Lemma 4.1. □

For fixed $x > 0$, define $q(z) := u'(x+z)u'(x-z)$ on $z \in (0, x)$.

Lemma A.2. *If* $-\frac{u''(z)}{u'(z)}$ *is strictly decreasing, then* $q(z)$ *is a strictly increasing function on* $z \in (0, x)$.

Proof. We take derivative

$$q'(z) = u''(x+z)u'(x-z) - u'(x+z)u''(x-z)$$
$$= u'(x+z)u'(x-z)\left[\left(-\frac{u''(x-z)}{u'(x-z)}\right) - \left(-\frac{u''(x+z)}{u'(x+z)}\right)\right] > 0.$$

Hence, we get the result. □

APPENDIX B: EXISTENCE OF OPTIMAL SOLUTIONS TO (11) AND (13)

We first prove that the constraint set $\mathbb{G}$ is compact under some norm. We consider all the continuous functions on [0,1], denoted as $C[0,1]$. Define a metric between $x(\cdot), y(\cdot) \in C[0,1]$ as

$$\rho(x(\cdot), y(\cdot)) = \max_{0 \leqslant t \leqslant 1} |x(t) - y(t)|.$$

Clearly, $C[0,1]$ is a metric space under ρ. By Arzela–Ascoli's theorem, for any sequence $(G_n(\cdot))_{n\in\mathbb{N}}$ in $\mathbb{G}$, there exists a subsequence $G_{n_k}(\cdot)$ that converges in $C[0,1]$ under ρ.

Lemma B.1. *The feasible set* $\mathbb{G}$ *is compact under* ρ.

Proof. For any sequence $(G_n(\cdot))_{n\in\mathbb{N}}$ in $\mathbb{G}$, there exists a subsequence $G_{n_k}(\cdot)$ that uniformly converges in $G^*(\cdot) \in C[0,1]$. We now prove that $G^*(\cdot) \in \mathbb{G}$. If there exist $a > b$ such that $G^*(b) - G^*(a) = \eta > 0$, then take $\varepsilon := \frac{1}{3}\eta$. If follows from the uniform convergence that there exists K such that

$$\rho(G_{n_k}(\cdot), G^*(\cdot)) \leqslant \varepsilon$$

for any $k \geqslant K$. Hence,

$$0 < \eta = G^*(b) - G^*(a) = G^*(b) - G_{n_k}(b) + G_{n_k}(b) - G_{n_k}(a) + G_{n_k}(a) - G^*(a) \leqslant \varepsilon + 0 + \varepsilon = \frac{2}{3}\eta$$

for any $k \geqslant K$, which is a contradiction. This proves that $0 \leqslant G^*(a) - G^*(b)$, for all $a > b$. Similarly, we can prove that $G^*(a) - G^*(b) \leqslant F_X^{-1}(a) - F_X^{-1}(b)$. □

The existence of optimal solutions to (11) and (13) can be established now. For example, for (13), let $v_\lambda(\Delta)$ be the optimal value of (13) under given λ and Δ. We can take a sequence $(G_n(\cdot))_{n\in\mathbb{N}}$ in $\mathbb{G}$ such that $v_\lambda(\Delta) = \lim_{n\uparrow+\infty} U_\Delta(\lambda, G_n(\cdot))$. Then, according to Lemma B.1, there exists a subsequence $G_{n_k}(\cdot)$ converging to $G^*(\cdot)$ in $\mathbb{G}$ and $G^*(\cdot)$ is optimal solution to (13). For (11), the proof is similar.

APPENDIX C: EXISTENCE OF LAGRANGIAN MULTIPLIER TO (11)

For the following lemma, refer to Komiya (1988) for an elementary proof.

Lemma C.1 (Sion's minimax theorem). *Let X be a compact convex subset of a linear topological space and Y a convex subset of a linear topological space. If f is a real-valued function on $X \times Y$ such that $f(x,\cdot)$ is continuous and concave on Y $\forall x \in X$, and $f(\cdot, y)$ is continuous and convex on X $\forall y \in Y$, then,* $\min_{x\in X} \max_{y\in Y} f(x,y) = \max_{y\in Y} \min_{x\in X} f(x,y)$.

Proposition C.2. *For any* $0 < \Delta < E[X]$*, there is λ^* such that $\widetilde{G}_{\lambda^*}(\cdot)$ is optimal solution to (13) under λ^* and* $\int_0^1 \widetilde{G}_{\lambda^*}(z)dz = \Delta$.

Proof. Let Δ be given with $0 < \Delta < E[X]$. Denote by $G^*(\cdot)$ the optimal solution to (11) under Δ (it is easy to show $\int_0^1 G^*(z)dz = \Delta$) and by $\widetilde{G}_\lambda(\cdot)$ the optimal solution to (13) under λ and Δ. Denote by $v(\Delta)$ and $v(\lambda, \Delta)$ be, respectively, the optimal values of (11) and (13).

We first prove that $v(\lambda, \Delta)$ is a convex function in λ for given Δ. Noting that $U_\Delta(\lambda, G(\cdot))$ is linear in λ for any given $G(\cdot)$, we have

$$\begin{aligned} v(\alpha\lambda_1 + (1-\alpha)\lambda_2, \Delta) &= \max_{G(\cdot)} U_\Delta(\alpha\lambda_1 + (1-\alpha)\lambda_2, G(\cdot)) \\ &= \max_{G(\cdot)} \{\alpha U_\Delta(\lambda_1, G(\cdot)) + (1-\alpha) U_\Delta(\lambda_2, G(\cdot))\} \\ &\leqslant \max_{G(\cdot)} \{\alpha U_\Delta(\lambda_1, G(\cdot))\} + \max_{G(\cdot)} \{(1-\alpha) U_\Delta(\lambda_2, G(\cdot))\} \end{aligned}$$

$$= \alpha \max_{G(\cdot)}\{U_\Delta(\lambda_1, G(\cdot))\} + (1-\alpha)\max_{G(\cdot)}\{U_\Delta(\lambda_2, G(\cdot))\}$$

$$= \alpha v(\lambda_1, \Delta) + (1-\alpha) v(\lambda_2, \Delta).$$

Moreover, by Sion's minimax theorem, the following equality holds

$$\max_{0\leqslant\lambda}\min_{G(\cdot)\in\mathbb{G}} -U_\Delta(\lambda, G(\cdot)) = \min_{G(\cdot)\in\mathbb{G}}\max_{0\leqslant\lambda} -U_\Delta(\lambda, G(\cdot));$$

hence

$$\min_{0\leqslant\lambda}\max_{G(\cdot)\in\mathbb{G}} U_\Delta(\lambda, G(\cdot)) = \max_{G(\cdot)\in\mathbb{G}}\min_{0\leqslant\lambda} U_\Delta(\lambda, G(\cdot)).$$

Finally, we have

$$v(\Delta) = \inf_{0\leqslant\lambda} v(\lambda, \Delta),$$

namely,

$$\min_{0\leqslant\lambda}\max_{G(\cdot)\in\mathbb{G}} U_\Delta(\lambda, G(\cdot)) = U_\Delta(G^*(\cdot)).$$

Let us denote

$$\overline{\lambda} := \frac{v(\Delta)+1}{\int_0^1 F_X^{-1}(z)dz - \Delta} = \frac{U_\Delta(G^*(\cdot))+1}{E[X]-\Delta}.$$

For any $\lambda \geqslant \overline{\lambda}$, we have

$$\begin{aligned}
v(\lambda, \Delta) &= \max_{G(\cdot)\in\mathbb{G}} U_\Delta(\lambda, G(\cdot)) \geqslant U_\Delta(\lambda, F_X^{-1}(z))) \\
&= \int_0^1 u\left(W_\Delta - F_X^{-1}(z)\right) T'(z)dz + \lambda\left(\int_0^1 F_X^{-1}(z)dz - \Delta\right) \\
&\geqslant \lambda\left(\int_0^1 F_X^{-1}(z)dz - \Delta\right) \\
&\geqslant \overline{\lambda}\left(\int_0^1 F_X^{-1}(z)dz - \Delta\right) \quad \left(\text{as} \int_0^1 F_X^{-1}(z)dz > \Delta\right) \\
&= v(\Delta) + 1,
\end{aligned}$$

which yields

$$v(\Delta) = \inf_{0\leqslant\lambda} v(\lambda, \Delta) = \inf_{0\leqslant\lambda\leqslant\overline{\lambda}} v(\lambda, \Delta).$$

Therefore, by the convexity of $v(\lambda, \Delta)$, we can find the optimal $\lambda^* \in [0, \overline{\lambda}]$ minimizes the right part, and satisfies that $v(\Delta) = v(\lambda^*, \Delta)$. Moreover,

$$v(\lambda^*, \Delta) \geqslant U_\Delta(\lambda^*, G^*(\cdot)) = \int_0^1 u(W_\Delta - G^*(z))T'(z)dz + \lambda^*\left(\int_0^1 G^*(z)dz - \Delta\right)$$

$$= \int_0^1 [u(W_\Delta - G^*(z))T'(z)]dz = U_\Delta(G^*(\cdot)) = v(\Delta).$$

The second equality comes from the fact that $G^*(\cdot)$ is the optimal solution to (11) under Δ; hence $\int_0^1 G^*(z)dz = \Delta$. By $v(\Delta) = v(\lambda^*, \Delta)$ and $v(\lambda^*, \Delta) \geqslant U_\Delta(\lambda^*, G^*(\cdot)) = v(\Delta)$, we have $G^*(\cdot)$ is optimal solution to (13) under given λ^*. And, by uniqueness of optimal solutions to (13), we know that $G^*(\cdot)$ is the unique optimal solution to (13) under given λ^* and satisfying $\int_0^1 G^*(z)dz = \Delta$. □

APPENDIX D: PROOFS

Proof of Lemma 2.4. First, by the monotonicity of $R(\cdot)$, we have

$$\mathbb{P}(R(X) \leqslant R(x)) \geqslant \mathbb{P}(X \leqslant x) = F_X(x),$$

so by the definition of $F^{-1}_{R(X)}(F_X(x))$, we conclude that

$$F^{-1}_{R(X)}(F_X(x)) \leqslant R(x).$$

It suffices to prove the reverse inequality. There are two possible cases.

- $R(x) = 0$. In this case, we have $F^{-1}_{R(X)}(F_X(x)) = 0$ as quantile functions are always nonnegative by definition.
- $R(x) > 0$. It suffices to prove that $\mathbb{P}(R(X) \leqslant z) < F_X(x)$ for any $z < R(x)$. Take z_1 such that $z < z_1 < R(x)$. By the continuity and monotonicity of $R(\cdot)$, there exists $y < x$ such that $R(y) = z_1$. Then,

$$\mathbb{P}(R(X) \leqslant z) \leqslant \mathbb{P}(R(X) < z_1) = \mathbb{P}(R(X) < R(y)) \leqslant \mathbb{P}(X \leqslant y) = F_X(y) < F_X(x),$$

  where we have used the fact that F_X is strictly increasing under Assumption 2.1.

The claim is thus proved. □

Proof of Lemma 2.5. We denote the right hand side of (12) by $\mathbb{G}_1$. For any $G(\cdot) \in \mathbb{G}$, there exists $R(\cdot) \in \mathcal{R}$ such that $G(\cdot) = F^{-1}_{R(X)}(\cdot)$. For any $0 \leqslant b < a \leqslant 1$, define

$$\underline{a} = \inf\{x \in [0, M] : R(x) = G(a)\},$$
$$\overline{a} = \sup\{x \in [0, M] : R(x) = G(a)\},$$

define $\underline{b}$ and $\overline{b}$ similarly. Let us show that $\underline{a} \leqslant F_X^{-1}(a) \leqslant \overline{a}$. In fact, by definition,

$$\begin{aligned} F_X^{-1}(a) &= \inf\{x \in \mathbb{R}^+ : F_X(x) \geqslant a\} \geqslant \inf\{x \in \mathbb{R}^+ : G(F_X(x)) \geqslant G(a)\} \\ &= \inf\{x \in \mathbb{R}^+ : R(x) \geqslant G(a)\} = \underline{a}. \end{aligned}$$

Suppose $F_X^{-1}(a) - \varepsilon > \overline{a}$ for some $\varepsilon > 0$. Then by monotonicity,

$$G(a) = R(\overline{a}) < R\left(F_X^{-1}(a) - \varepsilon\right) = G\left(F_X(F_X^{-1}(a) - \varepsilon)\right) \leqslant G(a),$$

where we have used the fact that $F_X(F_X^{-1}(a)-\varepsilon) < a$ to get the last inequality. This leads to a contradiction; hence it must hold that $F_X^{-1}(a) \leqslant \overline{a}$. Similarly, we can prove $\underline{b} \leqslant F_X^{-1}(b) \leqslant \overline{b}$. Then we have

$$0 \leqslant G(a) - G(b) = R(\underline{a}) - R(\overline{b}) \leqslant \underline{a} - \overline{b} \leqslant F_X^{-1}(a) - F_X^{-1}(b).$$

This inequality shows that G is absolutely continuous as F_X^{-1} is an absolutely continuous function under Assumption 2.1. Furthermore, it also implies

$$0 \leqslant G'(z) \leqslant (F_X^{-1})'(z),$$

a.e. $z \in [0,1]$. So we have established that $\mathbb{G} \subseteq \mathbb{G}_1$.

To prove the reverse inclusion, take any $G(\cdot) \in \mathbb{G}_1$ and define $R(\cdot) = G(F_X(\cdot))$. It follows from Assumption 2.1 that

$$0 \leqslant R(0) = G(F_X(0)) - G(0) \leqslant F_X^{-1}(F_X(0)) - F_X^{-1}(0) = 0$$

and

$$0 \leqslant R(a) - R(b) = G(F_X(a)) - G(F_X(b)) \leqslant F_X^{-1}(F_X(a)) - F_X^{-1}(F_X(b)) = a - b, \quad \forall\, 0 \leqslant b < a \leqslant 1.$$

Hence $R(\cdot) \in \mathcal{R}$. It now suffices to show $G(a) = F_{R(X)}^{-1}(a)$ for any $0 \leqslant a \leqslant 1$. If $G(a) = 0$, then $G(a) \leqslant F_{R(X)}^{-1}(a)$ holds. Otherwise, for any $s < G(a)$, there exists y such that $s < R(y) = G(F_X(y)) < G(a)$ by the continuity of $R(\cdot)$. Then by the monotonicity of $R(\cdot)$ and $G(\cdot)$, we have

$$\mathbb{P}(R(X) \leqslant s) \leqslant \mathbb{P}(R(X) < R(y)) \leqslant \mathbb{P}(X \leqslant y) = F_X(y) < a,$$

which means $G(a) \leqslant F_{R(X)}^{-1}(a)$. Using the same notation, $\overline{a}$, as above, and noting that $G(a) = R(\overline{a}) = G(F_X(\overline{a}))$, we have $a \leqslant F_X(\overline{a})$ by the definition of $\overline{a}$ and the continuity of $R(\cdot)$. Moreover, it follows from

$$\mathbb{P}(R(X) \leqslant G(a)) = \mathbb{P}(R(X) \leqslant R(\overline{a})) = \mathbb{P}(X \leqslant \overline{a}) = F_X(\overline{a})$$

that $F_{R(X)}^{-1}(F_X(\overline{a})) \leqslant G(a)$. Therefore,

$$G(a) \leqslant F_{R(X)}^{-1}(a) \leqslant F_{R(X)}^{-1}(F_X(\overline{a})) \leqslant G(a)$$

holds by monotonicity. The desired result follows. □

Proof of Theorem 3.1. We only need to prove the "if" part. For any feasible $G(\cdot)$ in $\mathbb{G}$, we have

$$\begin{aligned}
&U_\Delta(\lambda, \widetilde{G}_\lambda(\cdot)) - U_\Delta(\lambda, G(\cdot)) \\
&= \int_0^1 [u(W_\Delta - \widetilde{G}_\lambda(z)) - u(W_\Delta - G(z))]T'(z)dz + \int_0^1 \lambda(\widetilde{G}_\lambda(z) - G(z))dz \\
&\geqslant \int_0^1 u'(W_\Delta - \widetilde{G}_\lambda(z))(G(z) - \widetilde{G}_\lambda(z))T'(z)dz - \int_0^1 \lambda(G(z) - \widetilde{G}_\lambda(z))dz \\
&= \int_0^1 N_\lambda'(z)(G(z) - \widetilde{G}_\lambda(z))dz = \int_0^1 N_\lambda(t)(\widetilde{G}_\lambda'(t) - G'(t))dt \geqslant 0.
\end{aligned}$$

Hence, $\widetilde{G}_\lambda(\cdot)$ is optimal for (13). □

Proof of Lemma 4.1. We have

$$f'(z) = \frac{(1-T(z)) - T'(z)(1-z)}{(1-z)^2} = \frac{p(z)}{(1-z)^2},$$

where

$$p(z) := (1-T(z)) - T'(z)(1-z).$$

As

$$p'(z) = -T'(z) + T'(z) - T''(z)(1-z) = -T''(z)(1-z),$$

it follows from Assumption 2.3 that $p'(z) > 0$ for $z \in (0,b)$ and $p'(z) < 0$ for $z \in (b,1)$. Moreover, $p(0+) = 1 - T'(0+) < 0$,

$$p(b) = (1-T(b)) - T'(b)(1-b) = \left(\frac{1-T(b)}{1-b} - T'(b)\right)(1-b) > 0,$$

and

$$p(1-) = \lim_{z\uparrow 1}\left(\frac{1-T(z)}{1-z} - T'(z)\right)(1-z) \geqslant 0,$$

as $T(\cdot)$ is strictly convex on $[b,1]$. So, there exists $a \in (0,b)$ such that $p(z) < 0$ for $z \in [0,a)$ and $p(z) > 0$ for $z \in (a,1)$. The desired result follows. □

Proof of Theorem 4.2. We note that $\Delta = E[X] - \frac{\pi}{1+\rho}$ and the binding constraint $E[R(X)] \equiv \int_0^1 G_{R(X)}(z)dz = \Delta$ is equivalent to that $E[I(X)] = \frac{\pi}{1+\rho}$.

(i) If $\pi = (1+\rho)E[X]$, then $\Delta = 0$. Therefore, the optimal solution to (11) is trivially $G^*(z) = 0\ \forall z \in [0,1]$, or $I^*(z) = z\ \forall z \in [0,M]$.

(ii) If $\pi_c < \pi < (1+\rho)E[X]$, then $0 < \Delta < K_c$. In this case, there exists a unique pair (d,e) such that $0 \leqslant d < a < e \leqslant c$, $f(d) = f(e)$ and $\int_0^1 G^*(z)dz = \Delta$ where G^* is defined as follows:

$$G^*(z) = \begin{cases} 0, & \text{if } 0 \leqslant z < d, \\ F_X^{-1}(z) - F_X^{-1}(d), & \text{if } d \leqslant z < e, \\ F_X^{-1}(e) - F_X^{-1}(d), & \text{if } e \leqslant z \leqslant 1. \end{cases}$$

The existence of this pair follows from the condition that $\Delta < K_c$ and the definition of K_c, whereas the uniqueness comes from the requirement that $f(d) = f(e)$ and $\int_0^1 G^*(z)dz = \Delta$. Letting $\lambda = f(d)$, it is easy to show that $G^*(\cdot)$ satisfies (17) under λ, corresponding to the aforementioned Case II. This implies that $G^*(\cdot)$ is optimal for (11) under Δ. The optimal indemnity function is therefore $I^*(z) = z - G^*(F_X(z))$ for $z \in [0,M]$, leading to the desired expression.

(iii) If $0 \leqslant \pi \leqslant \pi_c$, then $K_c \leqslant \Delta \leqslant E[X]$, a case corresponding to Case III. The desired result can be derived similarly as in (ii) where $\lambda = f(q)$.

The proof is completed. □

Proof of Lemma 5.3. From $\lambda - u'(W_\Delta - G(z))T'(z) = 0$, it follows

$$u'(W_\Delta - G(z)) = \frac{\lambda}{T'(z)}.$$

Hence, if $z > a$, then

$$\begin{aligned} 0 &= \int_z^1 [\lambda - u'(W_\Delta - G(t))T'(t)]dt \\ &\leqslant \int_z^1 [\lambda - u'(W_\Delta - G(z))T'(t)]dt = \frac{\lambda}{T'(z)}(1-z)\left[T'(z) - \frac{1-T(z)}{1-z}\right] < 0, \end{aligned}$$

where the last inequality is due to Lemma A.1(i) in Appendix A. This is a contradiction. □

Proof of Lemma 5.5. Noting $W_\Delta \geqslant W$, it follows from Assumption 5.2 that

$$A_T(z) > A_u\left(W - F_X^{-1}(z)\right)\left(F_X^{-1}\right)'(z) \geqslant A_u\left(W_\Delta - F_X^{-1}(z)\right)\left(F_X^{-1}\right)'(z), \quad \forall z \in (0,a].$$

This leads to

$$\begin{aligned} &\frac{d}{dz}(u'(W_\Delta - G(z))T'(z)) \\ =\ & u'(W_\Delta - G(z))T'(z)\left[A_u(W_\Delta - G(z))G'(z) - A_T(z)\right] \\ <\ & u'(W_\Delta - G(z))T'(z)\left[A_u(W_\Delta - G(z))(F_X^{-1})'(z) - A_u(W_\Delta - F_X^{-1}(z))(F_X^{-1})'(z)\right] \\ =\ & u'(W_\Delta - G(z))T'(z)\left[A_u(W_\Delta - G(z)) - A_u(W_\Delta - F_X^{-1}(z))\right](F_X^{-1})'(z) \\ \leqslant\ & 0, \end{aligned}$$

where the last inequality is due to the fact that A_u is decreasing and $G(z) \leqslant F_X^{-1}(z)$. The proof is complete. □

Proof of Proposition 5.6. The existence of f follows from the monotonicity of G^* with respect to f immediately. Denoting

$$\lambda_\Delta := u'(W_\Delta - F_X^{-1}(f))\frac{1-T(f)}{1-f},$$

we need to show that $G^*(\cdot)$ satisfies (16) with $\lambda = \lambda_\Delta$. First, it is straightforward that

$$\int_f^1 \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(f))T'(t)\right] dt = 0.$$

Next, we are to prove that

$$\int_z^f \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt > 0, \quad \forall z \in (0,f).$$

We divide the proof into three cases.

- If $z \in [a, f)$, then

$$\int_z^f \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt$$

$$\geqslant \int_z^f \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(f))T'(t)\right] dt$$

$$= u'(W_\Delta - F_X^{-1}(f))(f-z)\left[\frac{1-T(f)}{1-f} - \frac{T(f)-T(z)}{f-z}\right] > 0,$$

where the last inequality is due to Lemma A.1(ii).

- If $z \in (0, a)$ and $u'(W_\Delta - F_X^{-1}(z))T'(z) \leqslant \frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f}$, then by Lemma 5.4 and the result above, we have

$$\int_z^f \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt$$

$$> \int_z^a \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt > 0.$$

- If $z \in (0, a)$ and $u'(W_\Delta - F_X^{-1}(z))T'(z) > \frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f}$, then, using $h_\Delta(l_\Delta) = 0$,

$$\int_z^f \left[\frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt$$

$$\geqslant \int_z^{l_\Delta} \left[\frac{u'(W_\Delta - F_X^{-1}(l_\Delta))(1-T(l_\Delta))}{1-l_\Delta} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt$$

$$= -\int_0^z \left[\frac{u'(W_\Delta - F_X^{-1}(l_\Delta))(1-T(l_\Delta))}{1-l_\Delta} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt > 0,$$

where the last inequality is due to

$$u'(W_\Delta - F_X^{-1}(z))T'(z) > \frac{u'(W_\Delta - F_X^{-1}(f))(1-T(f))}{1-f} \geqslant \frac{u'(W_\Delta - F_X^{-1}(l_\Delta))(1-T(l_\Delta))}{1-l_\Delta},$$

as $a < l_\Delta \leqslant f$ and the fact that $u'(W_\Delta - F_X^{-1}(z))T'(z)$ is strictly decreasing on $[0, a]$.

The claim follows now. □

Proof of Lemma 5.7. There exists λ^* such that $\widetilde{G}_{\lambda^*}(\cdot)$ satisfies (16) under λ^* and $\int_0^1 \widetilde{G}_{\lambda^*}(z)dz = \Delta$ (see Appendix C). If $\widetilde{G}_{\lambda^*}(\cdot)$ corresponds to a deductible contract, then there exists $\overline{z}$ (as $\Delta < K_\Delta$, we have $\overline{z} < l_\Delta$) such that

$$\widetilde{G}_{\lambda^*}(z) = \begin{cases} F_X^{-1}(z), & \text{if } 0 \leqslant z < \overline{z}, \\ F_X^{-1}(\overline{z}), & \text{if } \overline{z} \leqslant z \leqslant 1. \end{cases}$$

As $\widetilde{G}_{\lambda^*}(\cdot)$ satisfies (16), we have

$$\int_{\overline{z}}^1 [\lambda^* - u'(W_\Delta - F_X^{-1}(\overline{z}))T'(t)]dt = 0,$$

or

$$\lambda^* = u'(W_\Delta - F_X^{-1}(\overline{z}))\frac{1 - T'(\overline{z})}{1 - \overline{z}}.$$

On the other hand,

$$M(z) = \int_z^{\overline{z}} \left[u'(W_\Delta - F_X^{-1}(\overline{z}))\frac{1 - T'(\overline{z})}{1 - \overline{z}} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt \geqslant 0$$

for $z \in [0, \overline{z}]$. However, by the definition of l_Δ,

$$h_\Delta(\overline{z}) \equiv M(0) = \int_0^{\overline{z}} \left[\frac{u'(W_\Delta - F_X^{-1}(\overline{z}))(1 - T(\overline{z}))}{1 - \overline{z}} - u'(W_\Delta - F_X^{-1}(t))T'(t)\right] dt < 0$$

as $\overline{z} < l_\Delta$. As $M(\cdot)$ is a continuous function, a contradiction arises. □

Proof of Proposition 5.8. The conclusion is a direct consequence of Lemma 5.6. □

Proof of Theorem 5.9. The result follows from Propositions 5.5, 5.7, and the fact that K_Δ is a constant for any $0 < \Delta < E[X]$. □

Proof of Lemma 5.10. According to the definition of l_{Δ_1}, we have

$$h_{\Delta_1}(l_{\Delta_1}) = \int_0^{l_{\Delta_1}} \left[\frac{u'(W_{\Delta_1} - F_X^{-1}(l_{\Delta_1}))(1 - T(l_{\Delta_1}))}{1 - l_{\Delta_1}} - u'(W_{\Delta_1} - F_X^{-1}(t))T'(t)\right] dt = 0.$$

As $W_{\Delta_1} < W_{\Delta_2}$, we have

$$\frac{u'(W_{\Delta_2} - F_X^{-1}(l_{\Delta_1}))}{u'(W_{\Delta_1} - F_X^{-1}(l_{\Delta_1}))} < \frac{u'(W_{\Delta_2} - F_X^{-1}(t))}{u'(W_{\Delta_1} - F_X^{-1}(t))}$$

for $t \in [0, l_{\Delta_1})$ by Lemma A.2 in Appendix A. Hence

$$h_{\Delta_2}(l_{\Delta_1}) = \int_0^{l_{\Delta_1}} \left[\frac{u'(W_{\Delta_2} - F_X^{-1}(l_{\Delta_1}))(1 - T(l_{\Delta_1}))}{1 - l_{\Delta_1}} - u'(W_{\Delta_2} - F_X^{-1}(t))T'(t)\right] dt < 0.$$

As a result $h_{\Delta_2}(l_{\Delta_1}) < 0$, $h_{\Delta_2}(c) > 0$. As $h'_{\Delta_2}(z) > 0$ for $z \in [l_{\Delta_1}, c)$, we get $l_{\Delta_2} \in (l_{\Delta_1}, c)$. □

Proof of Theorem 5.11. According to the fact that $\Delta = E[X] - \frac{\pi}{1+\rho}$, (i), (ii), and (iv) are direct consequences of Propositions 5.5, 5.7, and Lemma 5.9. For (iii), if $(1+\rho)(E[X] - \overline{\Delta}) < \pi < (1+\rho)(E[X] - \widetilde{\Delta})$, then $\widetilde{\Delta} < \Delta < \overline{\Delta}$ and there is a unique $p \in (l_{\Delta(a)}, l_{\Delta(c)})$ such that $\Delta(p) = \Delta$, which follows from the definition of $\widetilde{\Delta}$, $\overline{\Delta}$ and the fact that $\Delta(\cdot)$ is a continuous and strictly increasing function. If $g(p) < 0$, then $h_\Delta(p) < 0$; hence $l_\Delta > p$. Therefore, $\Delta < K_\Delta$. The desired result follows from Proposition 5.7. The proof for $g(p) \geqslant 0$ is similar. $\square$