# Mean-variance insurance design with counterparty risk and incentive compatibility

Tim J. Boonen[1] and Wenjun Jiang[2]

[1] *Amsterdam School of Economics, University of Amsterdam, Amsterdam, 1001 NJ Amsterdam, The Netherlands. E-mail: t.j.boonen@uva.nl*

[2] *Department of Mathematics and Statistics, University of Calgary, Calgary, AB, T2N 1N4, Canada. E-mail: wenjun.jiang@ucalgary.ca*

**Abstract**

This paper studies the optimal insurance design from the perspective of an insured when there is possibility for the insurer to default on its promised indemnity. Default of the insurer leads to limited liability, and the promised indemnity is only partially recovered in case of a default. To alleviate the potential *ex post* moral hazard, an incentive compatibility condition is added to restrict the permissible indemnity function. Assuming that the premium is determined as a function of the expected coverage and under the mean-variance preference of the insured, we derive the explicit structure of the optimal indemnity function through the marginal indemnity function formulation of the problem. It is shown that the optimal indemnity function depends on the first and second order expectations of the random recovery rate conditioned on the realized insurable loss. The methodology and results in this paper complement the literature regarding the optimal insurance subject to the default risk and provide new insights on problems of similar types.

**Keywords**: Optimal insurance, mean-variance optimization, counterparty risk, marginal indemnity function.

**JEL code**: C70, G22.

## 1 Introduction

Insurance is an efficient and popular risk hedging tool. An insurance contract is usually composed of an indemnity function and a premium, where the premium needs to be paid *ex ante* by the insured to the insurer while the insurance indemnity is provided *ex post* from the insurer to the insured. Traditional indemnity functions include, for example, the stop-loss and proportional functions. Other forms of indemnity function are also possible and their optimality has been studied since the seminal works of Borch (1960) and Arrow (1973). Under various objective functions and premium principles and taking into account of more sophisticated economic factors, considerable advancements have been achieved in the literature. We refer to Chi and Tan (2011); Chi (2012); Cheung et al. (2019); Ghossoub (2019b); Boonen and Ghossoub (2019) for recent developments.

It is realistic that in addition to the insurable loss, the insured is also exposed to uninsurable risk such as inflation or catastrophe risk. These risk classes are jointly denoted as background risk and have received considerable attention since the work of Doherty and Schlesinger (1983). They show that the presence of background risk may affect the insured's demand of insurance, which is particularly the case if the background risk is correlated with the insurable loss. Several directions of introducing background risk in the optimal insurance problem have been proposed in the literature. One direction is on additive background risk, where the insured's total risk is the sum of insurable risk and background risk. In this way, Gollier (1996) shows that when the uninsurable loss increases with respect to the insurable loss, the optimal indemnity function is of disappearing deductible form. Dana and Scarsini (2007) study the Pareto-optimal insurance contract in the presence of background risk, where the qualitative properties of the optimal contract were derived under the assumption of stochastic increasingness. Chi and Wei (2018) investigate the optimal insurance under background risk and with higher order risk attitude. They also establish the optimality of the stop-loss insurance under specific dependence structures between the insurable risk and background risk. An extension is given by Chi and Wei (2020) for general dependence structures where the authors impose a constraint to avoid the moral hazard and present a very general result for the optimal insurance. Although the general solution to their model is implicit, many specific cases are analyzed and the corresponding optimal indemnity functions are derived. The aforementioned studies are all within the expected utility framework. Recently, Chi and Tan (2021) extend the study into the mean-variance framework. Their methodology allows the derivation of the optimal indemnity function for a very general dependence structure between the insurable risk and background risk.

The main focus of this paper is on multiplicative background risk. To gain a comprehensive view of this type of problem, we refer the interested readers to Franke et al. (2006). Multiplicative background risk may arise from the possibility of the insurer to default. The coverage is then the product of the promised indemnity and some random variable distributed on $[0, 1]$. This type of background risk is then interpreted as counterparty risk or default risk.

To model counterparty risk, there are roughly two streams in the literature. First, if there is only one insured in the market, default of the insurer can be modelled explicitly as a function of the indemnity function. In this way, optimal indemnities are derived by Cai et al. (2014), and the pricing is studied by Filipović et al. (2015). The latter study is extended by Boonen (2019) to the case with multiple policyholders, but under the assumption of exchangeable multivariate risk. Second, one assumes that the market with policyholders is "large", so that individual insurance transactions do not impact the likelihood of default. This assumption is imposed by Cummins and Mahul (2003), Bernard and Ludkovski (2012), and Li and Li (2018), and in this paper we also impose this assumption. Cummins and Mahul (2003) study the optimal insurance problem when the insurer has a positive probability to default and the insured and insurer have divergent beliefs about this probability. Their study is probably closest to Bernard and Ludkovski (2012), who investigate the impact of counterparty risk on optimal insurance when the insurer's default probability depends on the loss incurred by the insured. The optimal indemnity function is derived in an implicit way for the case where the insurer is risk neutral. Finally, Li and Li (2018) derive the optimal indemnity function in a Pareto-optimal insurance problem when the loss and recovery rate are negatively correlated for the cases where the information is symmetric and where the information is asymmetric. All these papers focus primarily on expected utility preferences of the insured, while our focus is on a mean-variance objective of the insured.

In this paper, we explicitly focus on the set of indemnity functions that are incentive compatible, which plays key role in alleviating the *ex post* moral hazard. For example, if the slope of indemnity function is larger than 1 at some point or the indemnity function has a discontinuous upward jump, the insured would be incentivized to create an increase in the insurable loss. Such behavior is called moral hazard. To remove the moral hazard issue, Ghossoub (2019a) introduces a state-verification cost such that the insurer could verify the state of real world by paying to a third party some extra cost. Another popular way to mitigate the moral hazard issue, as proposed by Huberman et al. (1983), is by requiring both the insured and insurer to pay more when the loss becomes larger. This is also named as the no-sabotage condition by Carlier and Dana (2003). Note that such condition is also considered by Chi and Wei (2020) and Chi and Tan (2021) where the additive background risk is the focus.

We summarize the main contributions of this paper as follows:

- First, to the authors' best knowledge, this paper is the first one that studies the counterparty-risk-based optimal insurance problem under mean-variance preferences of the insured and the incentive-compatibility condition. Through the marginal indemnity function formulation we present a general result which characterizes the optimal indemnity function implicitly to the problem under counterparty risk within a mean-variance framework. It is shown that the optimal indemnity function depends on the first and second order expectations of the counterparty risk conditioned on the realized insurable loss.

- Second, without assuming any specific dependence structure between the counterparty risk and the insurable loss, we derive the explicit structure of the optimal indemnity function based on the element-wise minimizer to our problem under some mild assumptions on the counterparty risk. We point out that the problem in Chi and Tan (2021) could also be solved using the marginal indemnity function approach, which provides an alternative approach to study the additive background risk model within the mean-variance framework.

The rest of this paper is structured as follow. Section 2 sets up the problem. Section 3 first characterizes the solution to the main problem in an implicit way, and then unveils the explicit structure of the optimal indemnity function based on its implicit characterization. Section 4 studies two special cases of the main problem. Section 5 presents some numerical examples illustrating the main result of this paper. Section 6 concludes the paper and gives directions for future research. All the proofs are delegated to Appendix A.

## 2 Problem formulation

We confine ourselves to a one-period economy. Suppose there is a decision maker (DM, also called insured) who is faced with a non-negative, bounded random loss $X$ whose support is $[0, M]$ (i.e., the set of numbers which have non-zero probability densities). The cumulative distribution function and density function of $X$ are given by $F(x)$ and $f(x)$ respectively. The DM would like to purchase an insurance contract $(I, \pi)$ where $I$ is the indemnity function and $\pi(I)$ is the premium principle that is used by the insurer.

There exists counterparty risk in the sense that the insurer may fail to pay its promised indemnity as per the contract at the end of period. We assume that the coverage received by the DM is given by $I(X) \cdot Y$, where $Y$ is a random variable distributed over $[0, 1]$ and may be correlated with $X$. In this paper, we follow Cummins and Mahul (2003) and Bernard and Ludkovski (2012) by assuming that the default event is exogenous in the sense that it is not affected by the DM's transactions. To avoid some trivial cases of the default event, we adopt the following assumption throughout the paper.

**Assumption 1.** *(i).* $\mathbb{P}(Y = 1) < 1$.

*(ii).* $\mathbb{P}(Y = 0|X = x) < 1$ *for all* $x \in [0, M]$.

Assumption 1 (ii) states that regardless of the DM's loss, the probability of that the insurer recovers some partial liability (e.g., through selling its remaining assets) is positive. Under this setting, the end-of-period loss of the DM is

$$L(I, \pi) = X - I(X) \cdot Y + \pi(I). \tag{2.1}$$

In insurance or reinsurance, some policies provide the DM an incentive to misreport the loss. For example, a franchise deductible indemnity function, i.e. $I(x) = x \cdot \mathbb{1}_{[d,\infty)}(x)$ for some $d \geq 0$, may incentivize the DM to over-report the loss or create an increment in the loss; a truncated excess-of-loss indemnity function, i.e. $I(x) = (x - d_1)_+ \cdot \mathbb{1}_{[0,d_2]}(x)$ for some $d_2 \geq d_1 \geq 0$, may incentivize the DM to under-report the loss if it exceeds $d_2$. Such behavior is called *ex post* moral hazard and should be seriously treated. A popular way in the literature to handle this issue is to restrict the indemnity functions to the following class:

$$\mathcal{I} = \left\{ I : [0, M] \mapsto [0, M] \;\middle|\; I(0) = 0,\ 0 \leq I(x_2) - I(x_1) \leq x_2 - x_1,\ \forall\, 0 \leq x_1 \leq x_2 \leq M \right\}.$$

The indemnity functions belonging to the set $\mathcal{I}$ are said to satisfy the *incentive-compatibility* or *no-sabotage* condition (Huberman et al., 1983; Carlier and Dana, 2003). The advantages of limiting the indemnity function to the class $\mathcal{I}$ are twofold. First, the indemnity function is always non-decreasing, i.e. $I'(x) \geq 0$, and satisfies $0 \leq I(x) \leq x$. That means, the indemnity is increasing with respect to the loss and can never be negative nor exceed the loss $x$ generated by the DM's loss. Second, one unit increment of loss cannot be compensated by more than one unit of indemnity. For any $I \in \mathcal{I}$, $I(X)$ and $X - I(X)$ are comonotonic with respect to the loss $X$. In particular, if the insurable loss $X$ increases, then both the indemnity $I(X)$ and the retained loss $X - I(X)$ increase.

For the premium, we assume that it is based on the coverage $I(X) \cdot Y$ rather than the promised indemnity $I(X)$. We further assume that the premium is determined as a function of the expected coverage:

$$\pi(I) = h(\mathbb{E}[I(X) \cdot Y]), \tag{2.2}$$

where $h(\cdot)$ is some differentiable function satisfying $h(0) = 0$ and $h'(x) > 1$ for $x \geq 0$. A special case of $h$ is $h(x) = (1 + \theta)x$ with $\theta > 0$, which leads to the expectation premium principle.

We follow Chi and Tan (2021) and study the optimal insurance problem within a mean-variance framework. We focus on the following general problem with counterparty risk.

**Problem 1** (Main problem)**.**

$$\min_{I \in \mathcal{I}} \ \mathcal{M}\left(\mathbb{E}[L(I, \pi)],\, Var(L(I, \pi))\right)$$

*where $\mathcal{M}(z_1, z_2)$ is increasing with respect to both $z_1$ and $z_2$.*

Problem 1 accommodates a wide range of mean-variance problems. For example, if let $\mathcal{M}(z_1, z_2) = z_1 + \frac{\mathcal{B}}{2} z_2$ for some $\mathcal{B} \geq 0$, the traditional mean-variance criterion is recovered, where $\mathcal{B}$ measures the DM's aversion towards volatility.

We remark that the optimal indemnity function within the mean-variance framework for additive background risk model has been derived by Chi and Tan (2021) through a constructive approach with stochastic ordering techniques. The marginal indemnity function formulation, which is obtained through the calculus of variations, not only helps us to solve the multiplicative background risk (or counterparty risk) model but also provides an alternative way to solve the additive background risk model.

## 3 Optimal indemnity function

Solving problem 1 usually takes a two-step procedure. In the first step we fix the mean (i.e., the first argument of $\mathcal{M}(\cdot,\cdot)$) and minimize the variance (i.e., the second argument of $\mathcal{M}(\cdot,\cdot)$):

$$\min_{I\in\mathcal{I}} \ \mathrm{Var}(L(I,\pi)), \quad \text{s.t. } \mathbb{E}[L(I,\pi)] = c, \tag{3.1}$$

where $c$ is a constant. As problem (3.1) depends on $c$, we denote its solution by $I_c$. In the second step, we search for the optimal $c$ such that $\mathcal{M}(c, \mathrm{Var}(L(I_c,\pi)))$ reaches its minimum. The second step is a one-dimensional problem, and can be solved using standard techniques. The first step plays a vital role as it identifies the optimal indemnity function with $c$ as its parameter. In the sequel, we focus on this first step, and thus on the problem (3.1).

Let $\psi_1(x) = \mathbb{E}[Y|X = x]$ and $\psi_2(x) = \mathbb{E}[Y^2|X = x]$. Under Assumption 1, it follows that $\psi_1(x) > 0$ and $\psi_2(x) > 0$ for all $x \in [0, M]$. Apply the conditional variance formula to the objective function of (3.1) gives

$$\begin{aligned}
\mathrm{Var}(L(I,\pi)) =& \mathbb{E}\left[\mathrm{Var}(L(I,\pi)\big|X)\right] + \mathrm{Var}\left(\mathbb{E}[L(I,\pi)\big|X]\right)\\
=& \mathbb{E}[I(X)^2 \cdot \mathrm{Var}(Y\big|X)] + \mathrm{Var}(X - I(X)\cdot \mathbb{E}[Y\big|X])\\
=& \mathbb{E}[I(X)^2 \cdot \psi_2(X)] - \mathbb{E}[I(X)^2 \cdot \psi_1(X)^2] + \mathbb{E}\left[(X - I(X)\cdot\psi_1(X))^2\right] - \left(\mathbb{E}[X - I(X)\cdot\psi_1(X)]\right)^2\\
=& \mathbb{E}[I(X)^2 \cdot \psi_2(X)] - \mathbb{E}[I(X)^2 \cdot \psi_1(X)^2] + \mathbb{E}[X^2] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)]\\
&+ \mathbb{E}[I(X)^2 \cdot \psi_1(X)^2] - \mathbb{E}[X]^2 + 2\mathbb{E}[X]\cdot\mathbb{E}[I(X)\cdot\psi_1(X)] - \mathbb{E}[I(X)\cdot\psi_1(X)]^2\\
=& \mathbb{E}[I(X)^2 \cdot \psi_2(X)] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)] + \mathrm{Var}(X)\\
&+ 2\mathbb{E}[X]\cdot\mathbb{E}[I(X)\cdot\psi_1(X)] - \mathbb{E}[I(X)\cdot\psi_1(X)]^2.
\end{aligned}$$

The constraint of (3.1) reduces to

$$\begin{aligned}
&\mathbb{E}[X - I(X)\cdot Y + \pi] = \mathbb{E}[X] - \mathbb{E}[I(X)\cdot Y] + \pi = c\\
\Longrightarrow\ & h(\mathbb{E}[I(X)\cdot Y]) - \mathbb{E}[I(X)\cdot Y] = c - \mathbb{E}[X]\\
\Longrightarrow\ & h(\mathbb{E}[I(X)\cdot\psi_1(X)]) - \mathbb{E}[I(X)\cdot\psi_1(X)] = c - \mathbb{E}[X].
\end{aligned}$$

Since $\big(h(x) - x\big)' = h'(x) - 1 > 0$ and $h(0) = 0$, for any $c \geq \mathbb{E}[X]$, the equation $h(x) - x = c - \mathbb{E}[X]$ has only one solution, which is denoted as $x^*$. Based on the above simplifications, problem (3.1) reduces to

$$\begin{aligned}
&\min_{I\in\mathcal{I}} \mathbb{E}[I(X)^2\cdot\psi_2(X)] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)]\\
&\text{s.t. } \mathbb{E}[I(X)\cdot\psi_1(X)] = x^*.
\end{aligned} \tag{3.2}$$

Solving problem (3.2) is equivalent to solving its Lagrangian dual problem:

$$\min_{I\in\mathcal{I}} \ \mathbb{E}[I(X)^2\cdot\psi_2(X)] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)] + \lambda\cdot\mathbb{E}[I(X)\cdot\psi_1(X)] \tag{3.3}$$

where $\lambda \in \mathbb{R}$ is the Lagrangian coefficient.

The following lemma characterizes the optimal indemnity function to problem (3.3). In this lemma, $\mathbb{1}_A(t)$ is defined as the indicator function: $\mathbb{1}_A(t) = 1$ if $t \in A$ and $\mathbb{1}_A(t) = 0$ otherwise.

**Lemma 3.1.** *Let Assumption 1 hold, and*

$$L(t; I^*, \lambda) = \int_t^M \psi_2(x) \cdot \left( I^*(x) - \frac{\psi_1(x)}{\psi_2(x)}(x - \frac{\lambda}{2}) \right) dF_X(x).$$

*Then,* $I^*(x) = \int_0^x \eta^*(t)dt$ *is an optimal solution to problem* (3.3) *if and only if*

$$\eta^*(t) = \mathbb{1}_{D_\lambda}(t) + \xi(t) \cdot \mathbb{1}_{E_\lambda}(t),$$

*where*

$$D_\lambda = \left\{ t : L(t; I^*, \lambda) < 0 \right\}, \quad E_\lambda = \left\{ t : L(t; I^*, \lambda) = 0 \right\},$$

*and* $\xi(t) \in [0, 1]$ *is such that* $I^* \in \mathcal{I}$.

Lemma 3.1 characterizes the optimal marginal indemnity functions. However, the optimal indemnity functions given by Lemma 3.1 are implicit since $I^*$ also appears in $L(t; I^*, \lambda)$. Nevertheless, Lemma 3.1 provides insights about the explicit structure of $I^*$, which will be derived in detail in the rest of this section. We remark that in deriving the above result, the calculus of variations plays important role. In recent years, such technique has been widely applied to obtain an implicit characterization of the optimal indemnity function under other preference functionals. See, for example, Chi and Wei (2020) and Chi and Zhuang (2020).

For now, let $\lambda$ be fixed and

$$\phi_\lambda(x) = \frac{\psi_1(x)}{\psi_2(x)}(x - \frac{\lambda}{2}).$$

It is easy to verify that $\phi_\lambda(x)$ is the element-wise minimizer to problem (3.3)[1]. The following assumption is needed to proceed.

**Assumption 2.** *The mapping* $x \mapsto \frac{\psi_1(x)}{\psi_2(x)}$ *is continuously differentiable.*

Under Assumption 2, $\phi_\lambda(x)$ is also continuously differentiable over $[0, M]$. As such, the whole domain $[0, M]$ could be partitioned as per the first order derivative of $\phi_\lambda(x)$ such that

$$[0, M] = \bigcup_{i=1}^{m} S_{i,j_i}, \tag{3.4}$$

where

$$j_i = \begin{cases} 1, & \text{if } \phi'_\lambda(x) \in (1, \infty), \\ 2, & \text{if } \phi'_\lambda(x) \in [0, 1], \\ 3, & \text{if } \phi'_\lambda(x) \in (-\infty, 0), \end{cases} \tag{3.5}$$

[1]The objective function of (3.3) could be written as

$$\int_0^M \left\{ I(x)^2 \psi_2(x) - 2x \cdot I(x) \cdot \psi_1(x) + \lambda I(x) \cdot \psi_1(x) \right\} dF(x).$$

An element-wise minimizer $\tilde{I}(x)$ is such that

$$\tilde{I}(x) = \arg\min_{z \in \mathbb{R}} \; Q(z) := z^2 \psi_2(x) - 2x \cdot z \cdot \psi_1(x) + \lambda z \cdot \psi_1(x).$$

As $Q(z)$ is a convex function, it is easy to derive from the first-order condition that $\tilde{I}(x) = \phi_\lambda(x) = \frac{\psi_1(x)}{\psi_2(x)} \left( x - \frac{\lambda}{2} \right)$.

and $m$ is the smallest positive integer required for such a partition. Throughout the rest of this paper, $m$ is assumed to be finite. Under this partition rule, we have $j_{i+1} \neq j_i$ and $|j_{i+1} - j_i| = 1$ for $i = 1, 2, \dots, m-1$. Let $x_{i-1} = \inf \left\{ x : x \in S_{i,j_i} \right\}$ for $i = 1, 2, \dots, m$, we have $0 = x_0 \leq x_1 \leq \cdots \leq x_{m-1} \leq x_m = M$. Intuitively, $x_i = \sup \left\{ x : x \in S_{i,j_i} \right\}$ for $i = 1, 2, \dots, m$. These points $\{x_i\}_{i=1,2,\dots,m-1}$ are referred to as the change points (see also Chi and Tan, 2021). Our goal is to obtain the explicit structure of the optimal indemnity function over each piece $S_{i,j_i}$. For the ease of presentation, the following layer-type indemnity function is defined:

$$I_{(a,b]}(x) = (x-a)_+ - (x-b)_+ \quad \text{where} \quad 0 \leq a \leq b \leq M.$$

The following theorem gives the optimal parametric indemnity function over each $S_{i,j_i}$ for $i = 1, 2, \dots, m$.

**Theorem 3.1.** *Let Assumptions 1 and 2 hold. For problem* (3.3)*, the optimal indemnity function is given by $I^*(x)$ such that, for $x \in S_{m,j_m}$,*

*(1). if $j_m = 1$, then $I^*(x) = I^*(x_{m-1}) + (x - \gamma_{m,1})_+$ for some $\gamma_{m,1} \in [x_{m-1}, M]$,*

*(2). if $j_m = 2$, then $I^*(x) = \min \left\{ \max \left\{ \phi_\lambda(x), I^*(x_{m-1}) \right\}, I^*(x_{m-1}) + x - x_{m-1} \right\}$,*

*(3). if $j_m = 3$, then $I^*(x) = I^*(x_{m-1}) + I_{(x_{m-1}, \gamma_{m,3}]}(x)$ for some $\gamma_{m,3} \in [x_{m-1}, M]$,*

*and for $x \in S_{i,j_i}$, $i = 1, 2, \dots, m-1$,*

*(4). if $j_i = 1$, then $I^*(x) = I^*(x_{i-1}) + I_{(\gamma_{i,1}, \gamma_{i,1} + I^*(x_i) - I^*(x_{i-1})]}(x)$ for some $\gamma_{i,1} \in [x_{i-1}, x_i]$,*

*(5). if $j_i = 2$, then $I^*(x) = \min \left\{ \max \left\{ \phi_\lambda(x), I^*(x_i) + x - x_i, I^*(x_{i-1}) \right\}, I^*(x_{i-1}) + x - x_{i-1}, I^*(x_i) \right\}$,*

*(6). if $j_i = 3$, then $I^*(x) = I^*(x_{i-1}) + x - x_{i-1} - I_{(\gamma_{i,3}, \gamma_{i,3} + x_i - x_{i-1} - (I^*(x_i) - I^*(x_{i-1}))]}(x)$ for some $\gamma_{i,3} \in [x_{i-1}, x_i]$.*

An illustration of the optimal indemnity function $I^*$ is shown in Fig. 1, where the red line denotes $I^*$ and the green dashed line denotes the element-wise minimizer $\phi_\lambda$ to problem (3.3). We can see that $\phi_\lambda$ does not satisfy the incentive compatibility condition, so the optimal indemnity function $I^*$ is obtained based on the slope of $\phi_\lambda$. In Fig. 1, there are two change points, i.e. $x_1$ and $x_2$, and the domain of loss is partitioned into three pieces. We obtain the parametric form of the optimal indemnity function on each piece by using Theorem 3.1.

- Over $[0, x_1]$, $\phi'_\lambda(x) < 0$, so $j_1 = 3$, and

$$I^*(x) = x - I_{(\gamma_{1,3}, \gamma_{1,3} + x_1 - I^*(x_1)]}(x).$$

- Over $[x_1, x_2]$, $0 \leq \phi'_\lambda(x) \leq 1$, so $j_2 = 2$, and

$$I^*(x) = \min \left\{ \max \left\{ \phi_\lambda(x), I^*(x_2) + x - x_2, I^*(x_1) \right\}, I^*(x_1) + x - x_1, I^*(x_2) \right\}.$$

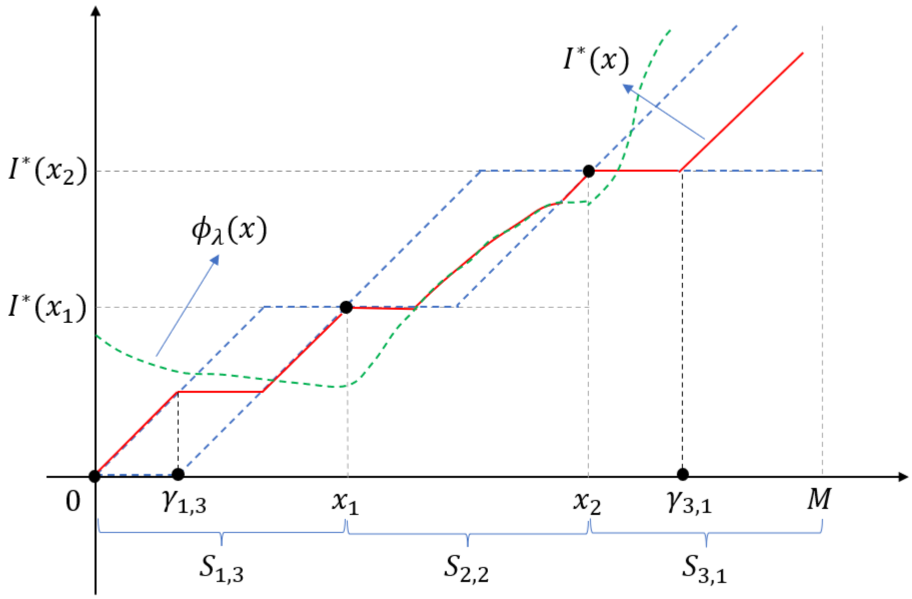


Figure 1: An illustration of an indemnity function $I^*$ that solves problem (3.3), when $m = 3$, $j_1 = 3$, $j_2 = 2$ and $j_3 = 1$.

- Over $[x_2, M]$, $\phi'_\lambda(x) > 1$, so $j_3 = 1$, and

$$I^*(x) = I^*(x_2) + (x - \gamma_{3,1})_+.$$

Theorem 3.1 shows the applicability of Lemma 3.1 in practice and gives the explicit structure of the optimal indemnity function. It reduces the dimension of the original optimization problem from $\infty$ to at most $2m-1$ (i.e., $I^*(x_i)$ for $i = 1, 2, \ldots, m-1$, $\gamma_{i,1}$ and $\gamma_{i,3}$ for $i = 1, 2, \ldots, m$, and $\lambda$).

We remark that the partition (3.4) varies with respect to the Lagrangian coefficient $\lambda$, which is a key parameter pertaining to the premium $\pi$. Therefore, if the number of change points (i.e., $m-1$) is large, numerically optimizing the parameters in Theorem 3.1 is still computationally expensive. To slightly simplify the computation, the following proposition is given.

**Proposition 3.1.** *Let Assumption 1 hold. If $\lambda > 0$, then for problem* (3.3) *the optimal indemnity function over $[0, \frac{\lambda}{2}]$ is given by*

$$I^*(x) = (x - d)_+$$

*for some $d \in [0, \frac{\lambda}{2}]$.*

Proposition 3.1 shows that the optimal indemnity function is of the stop-loss form in a neighborhood of 0 if $\lambda > 0$. To derive the general solution, only the interval $[\frac{\lambda}{2}, M]$ needs to be partitioned as per (3.4) and (3.5).

**Remark 3.1.** *The marginal indemnity function approach applied in this paper is also applicable to the additive-background-risk-based optimal insurance problem within the mean-variance framework.*

*The only difference is that for the additive-background-risk-based problem the partition does not rely on the Lagrangian coefficient. In Chi and Tan (2021), a constructive approach, together with some stochastic ordering technique, was applied to identify the parametric form of the optimal indemnity function on each piece of the domain, which yields different Lagrangian coefficients to be optimized over different pieces. Through Theorem 3.1, we are able to show that at the optimum all the Lagrangian coefficients in their result are equal, which is a supplementary finding to their study.*

## 4 Two special dependence structures

In this section, we study two special cases of our problem, i.e. when the counterparty risk $Y$ is independent of $X$ and when $Y$ is a decreasing function of $X$. In both special cases, we will derive a much simpler solution.

### 4.1 $Y$ is independent of $X$

In this section, we study the indemnity function for the case where $Y$ is independent of $X$. This happens when the DM's loss does not affect the solvency status of the insurer. In such a case, $\frac{\psi_1(x)}{\psi_2(x)} = \frac{\mathbb{E}[Y|X=x]}{\mathbb{E}[Y^2|X=x]} = \frac{\mathbb{E}[Y]}{\mathbb{E}[Y^2]}$ is a constant. Then $\phi'_\lambda(x) = \frac{\mathbb{E}[Y]}{\mathbb{E}[Y^2]} > 1$ for any $x \in [0, M]$. This implies that $m = 1$ and there is no change point. Applying Theorem 3.1 leads to the following corollary.

**Corollary 4.1.** *If $Y$ is independent of $X$, then the optimal indemnity function to Problem 1 is given by*

$$I^*(x) = (x - d)_+$$

*for some $d \geq 0$.*

The optimality of stop-loss indemnity function is also verified in the expected utility framework. For example, Cummins and Mahul (2003) show the optimality of a stop-loss function in a situation where the recovery rate can only take 1 or 0 and both the DM and insurer have the same belief about the default probability. Bernard and Ludkovski (2012) extend the result of Cummins and Mahul (2003) by considering a budget constraint, where the stop-loss function is proved to be optimal again. In the above-mentioned works, the recovery rate does not need to be independent of the insurable loss.

### 4.2 $Y$ is a decreasing function of $X$

In the literature, the recovery rate $Y$ is generally assumed to be negatively correlated with $X$ (see Bernard and Ludkovski (2012); Li and Li (2018)). This is intuitive, as a larger loss would make the insurer more likely to default, which results in a smaller recovery rate. In this section, we analyze a special case where $Y$ is a decreasing function $X$, i.e. $Y = g(X)$ where $g$ is a decreasing function.[2]

To simplify our discussion, we focus on the following situation.

[2]In this paper, we do not distinguish between "decreasing" and "non-increasing" here.

**Assumption 3.** *(i) there exists an* $x_1 \in [0, M]$ *such that* $g(x) = 1$ *for all* $x \in (0, x_1]$*, and* $g(x) < 1$ *for all* $x \in (x_1, M]$*;*

*(ii)* $g(M) > 0$.

Assumption 3 (i) states that the insurer will not default when the DM's loss is less than some threshold. In the rare case where the realized loss is the largest, the insurer may default, but is able to sell its remaining assets and recover part of the indemnity to the DM. This leads Assumption 3 (ii) to hold. Assumption 3 is for instance related to the model of Cai et al. (2014), when the insurer sells an insurance contract to only one DM. In such case, default happens if and only if the insurable loss exceeds a certain threshold.

Note that $\frac{\psi_1(x)}{\psi_2(x)} = \frac{\mathbb{E}[g(X)|X=x]}{\mathbb{E}[g(X)^2|X=x]} = \frac{g(x)}{g(x)^2} = \frac{1}{g(x)}$. Thus, under Assumption 3

$$\phi_\lambda(x) = \begin{cases} x - \dfrac{\lambda}{2}, & x \in [0, x_1], \\ \dfrac{x - \frac{\lambda}{2}}{g(x)}, & x \in (x_1, M]. \end{cases}$$

For $x > \max\left\{x_1, \frac{\lambda}{2}\right\}$, we have

$$\phi_\lambda'(x) = \left(\frac{x - \frac{\lambda}{2}}{g(x)}\right)' = -\frac{g'(x)}{g(x)^2}(x - \frac{\lambda}{2}) + \frac{1}{g(x)} \geq \frac{1}{g(x)} > 1.$$

Depending on the value of $\lambda$, we have the following two sub-cases.

- *Case 1:* $\lambda \leq 2x_1$

  In this case, we have $\phi_\lambda'(x) = 1$ for $x \in [0, x_1]$ and $\phi_\lambda'(x) > 1$ for $x \in (x_1, M]$. Therefore, $x_1$ is the only change point. Applying Theorem 3.1 and Proposition 3.1 gives

  $$I^*(x) = (\min\{x, d_2\} - d_1)_+ + (x - d_3)_+,$$

  where $0 \leq d_1 \leq \frac{\lambda}{2} \leq d_2 \leq x_1 \leq d_2 \leq M$.

- *Case 2:* $\lambda > 2x_1$

  In this case, as per Proposition 3.1, for any $x \in [0, \frac{\lambda}{2}]$ we have $I^*(x) = (x - d_1)_+$ for some $d_1 \in [0, \frac{\lambda}{2}]$. For $x \in (\frac{\lambda}{2}, M]$, applying Theorem 3.1 leads to $I^*(x) = I^*(\frac{\lambda}{2}) + (x - d_2)_+$ for some $d_2 \in [\frac{\lambda}{2}, M]$. In conclusion, the optimal indemnity function for this case is given by

  $$I^*(x) = (\min\left\{x, \frac{\lambda}{2}\right\} - d_1)_+ + (x - d_2)_+,$$

  where $0 \leq d_1 \leq \frac{\lambda}{2} \leq d_2$.

Note that the solutions for Case 1 and 2 are of similar formats. The above discussions are summarized in the next corollary.

**Corollary 4.2.** *If $Y$ is a decreasing function of $X$, under Assumption 3, the optimal indemnity function to Problem 1 is given by*

$$I^*(x) = (\min\{x, a_2\} - a_1)_+ + (x - a_3)_+,$$

*where* $0 \le a_1 \le a_2 \le a_3 \le M$.

## 5 Numerical illustrations

In this section, we present a numerical example illustrating the main result of this paper. For the ease of discussion, we assume that the DM uses the traditional mean-variance criterion: $\mathcal{M}(z_1, z_2) = z_1 + \frac{\beta}{2} z_2$. The premium is determined by the expectation premium principle, i.e. $\pi(I) = (1+\theta)\mathbb{E}[I(X)Y]$ for some $\theta > 0$. Moreover, we assume the following structure of $Y|X = x$ to depict the dependence between the recovery rate and loss:

$$\begin{cases} Y = 1, & \text{with probability } p(x) \\ Y \sim U(0,1), & \text{with probability } 1 - p(x), \end{cases} \tag{5.1}$$

where $x$ is the realization of $X$ and $U(0,1)$ denotes the uniform distribution over $[0,1)$. In this case, the random variable $Y|Y < 1$ is independent of $X$. Such structure is inspired by Bernard and Ludkovski (2012), who study the case where $Y$ takes one out of two values: $Y \in \{y_0, 1\}$ for some $y_0 \in [0,1]$. As $Y$ is usually negatively correlated with $X$, $p(\cdot)$ is usually a decreasing function. Then, the larger the loss is, the smaller is the probability of full recovery. Moreover, for any increasing function $g$, it holds that

$$\mathbb{E}[g(Y)|X = x] = g(1)p(x) + (1 - p(x)) \int_0^1 g(y)dy = \int_0^1 g(y)dy + p(x)(g(1) - \int_0^1 g(y)dy),$$

which is decreasing with respect to $x$. This implies that $Y$ is stochastically decreasing with respect to $X$. For simplicity, we assume that $p(x)$ is subject to exponential decay, i.e. $p(x) = e^{-ax}$ for some $a > 0$.

Under these assumptions, we can easily get

$$\psi_1(x) = \mathbb{E}[Y|X = x] = \frac{1}{2} + \frac{1}{2}p(x), \quad \psi_2(x) = \mathbb{E}[Y^2|X = x] = \frac{1}{3} + \frac{2}{3}p(x).$$

Furthermore,

$$\left(\frac{\psi_1(x)}{\psi_2(x)}\right)' = \frac{3}{2} \cdot \frac{ae^{-ax}}{(1 + 2e^{-ax})^2} > 0.$$

Therefore $\frac{\psi_1(x)}{\psi_2(x)} \ge \frac{\psi_1(0)}{\psi_2(0)} = 1$.

Based on the value of $\lambda$, we have the following cases.

- If $\lambda > 0$, applying Proposition 3.1 leads to

$$I^*(x) = (x - d_1)_+$$

on $[0, \frac{\lambda}{2}]$ for some $d_1 \in [0, \frac{\lambda}{2}]$. When $x > \frac{\lambda}{2}$,

$$\phi_\lambda'(x) = \left(\frac{\psi_1(x)}{\psi_2(x)}\right)'(x - \frac{\lambda}{2}) + \frac{\psi_1(x)}{\psi_2(x)} > 1.$$

Applying Theorem 3.1 leads to

$$I^*(x) = I^*(\frac{\lambda}{2}) + (x - d_2)_+$$

for some $d_2 \geq \frac{\lambda}{2}$.

- If $\lambda \leq 0$, then since $\phi_\lambda'(x) > 1$ on $[0, M]$, applying Theorem 3.1 leads to

$$I^*(x) = (x - d_3)_+$$

over the whole domain, where $d_3 \in [0, M]$.

Summarizing the above discussions leads to the general solution

$$I^*(x) = I_{a_1,a_2,a_3}(x) := \left(\min\{x, a_2\} - a_1\right)_+ + (x - a_3)_+,$$

where $0 \leq a_1 \leq a_2 \leq a_3 \leq M$. Note that if $a_1 = a_2$ or $a_2 = a_3$, this solution reduces to a stop-loss function. An illustrative $I^*$ is given by Fig. 2.

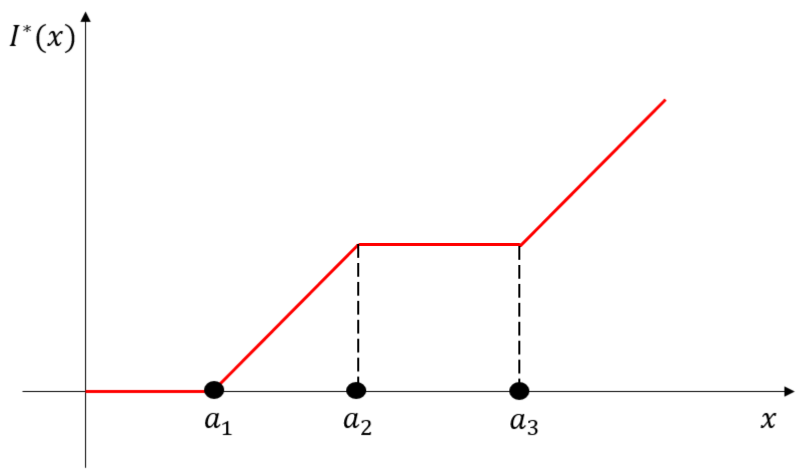


Figure 2: Illustrative $I^*$.

Under the traditional mean-variance criterion, our goal is to minimize

$$\begin{aligned}
&\mathcal{M}(\mathbb{E}[L(I,\pi)], \mathrm{Var}(L(I,\pi))) \\
=&\mathbb{E}[X] + \theta\mathbb{E}[I(X)Y] + \frac{\mathcal{B}}{2}\left\{\mathbb{E}[(X - I(X)Y)^2] - \left(\mathbb{E}[X - I(X)Y]\right)^2\right\} \\
=&\mathbb{E}[X] + \theta\mathbb{E}[I(X)\psi_1(X)] + \frac{\mathcal{B}}{2}\Big\{\mathrm{Var}(X) - 2\mathbb{E}[XI(X)\psi_1(X)] + \mathbb{E}[I(X)^2\psi_2(X)] \\
&+2\mathbb{E}[X]\mathbb{E}[I(X)\psi_1(X)] - \mathbb{E}[I(X)\psi_1(X)]^2\Big\}.
\end{aligned}$$

With the optimal indemnity function $I_{a_1,a_2,a_3}$, the above problem reduces to

$$\min_{0\le a_1\le a_2\le a_3\le M} \theta\mathbb{E}[I_{a_1,a_2,a_3}(X)\psi_1(X)] + \frac{\mathcal{B}}{2}\Big\{\mathbb{E}[I^2_{a_1,a_2,a_3}(X)\psi_2(X)] + 2\mathbb{E}[X]\mathbb{E}[I_{a_1,a_2,a_3}(X)\psi_1(X)] \\ -2\mathbb{E}[XI_{a_1,a_2,a_3}(X)\psi_1(X)] - \mathbb{E}[I_{a_1,a_2,a_3}(X)\psi_1(X)]^2\Big\}. \tag{5.2}$$

To optimize $a_1, a_2$ and $a_3$, we use the "fmincon" or "patternsearch" function in MATLAB. Under different loss distributions and volatility aversion parameters, the optimal $a_1, a_2$ and $a_3$ are presented in Tables 1 and 2.

| | $\mu = 250$ | $\mu = 500$ | $\mu = 750$ | $\mu = 1000$ |
|---|---|---|---|---|
| $\mathcal{B} = 0.001$ | $(115.3, 334.2, 334.2)$ | $(218.3, 218.3, 223.9)$ | $(137.5, 137.5, 324.0)$ | $(120.0, 120.0, 416.9)$ |
| $\mathcal{B} = 0.005$ | $(98.5, 342.1, 342.1)$ | $(208.8, 227.8, 227.8)$ | $(143.7, 143.7, 310.8)$ | $(116.4, 116.4, 404.5)$ |
| $\mathcal{B} = 0.01$ | $(96.3, 342.4, 342.4)$ | $(206.9, 227.1, 227.1)$ | $(114.9, 114.9, 309.3)$ | $(116.8, 116.8, 402.9)$ |
| $\mathcal{B} = 0.02$ | $(95.2, 342.5, 342.5)$ | $(205.7, 226.7, 226.7)$ | $(144.1, 144.1, 308.6)$ | $(114.9, 114.9, 402.2)$ |
| $\mathcal{B} = 0.04$ | $(94.7, 342.8, 342.8)$ | $(205.6, 225.6, 225.6)$ | $(142.2, 142.2, 308.2)$ | $(115.2, 115.2, 401.5)$ |

Table 1: The effects of the mean loss and $\mathcal{B}$ on $(a_1, a_2, a_3)$ for exponentially distributed loss, $\theta = 0.01$ and $a = 0.001$.

| | $\theta = 0.01$ | $\theta = 0.02$ | $\theta = 0.05$ | $\theta = 0.1$ |
|---|---|---|---|---|
| $a = 0.001$ | $(208.8, 227.8, 227.8)$ | $(212.5, 226.2, 226.2)$ | $(221.6, 222.0, 223.4)$ | $(212.1, 212.1, 240.2)$ |
| $a = 0.002$ | $(203.7, 229.7, 229.7)$ | $(206.9, 227.9, 227.9)$ | $(216.0, 222.4, 222.4)$ | $(213.9, 213.9, 230.9)$ |
| $a = 0.003$ | $(192.8, 234.4, 234.4)$ | $(195.9, 233.9, 233.9)$ | $(205.6, 228.2, 228.2)$ | $(218.3, 220.9, 221.2)$ |
| $a = 0.005$ | $(176.5, 247.5, 247.5)$ | $(179.2, 242.8, 242.8)$ | $(187.6, 238.7, 238.7)$ | $(201.5, 232.7, 232.7)$ |
| $a = 0.01$ | $(158.8, 244.9, 244.9)$ | $(161.6, 243.0, 243.0)$ | $(170.2, 237.5, 237.5)$ | $(184.7, 241.2, 241.2)$ |

Table 2: The effects of $\theta$ and $a$ on $(a_1, a_2, a_3)$ for exponentially distributed loss with mean 500 and $\mathcal{B} = 0.005$.

Interestingly, in most of the cases in Tables 1 and 2, we get either $a_1 = a_2$ or $a_2 = a_3$. This implies that the optimal indemnity function for the considered model settings are of the stop-loss type. For the sensitivity of the retention point with respect to model parameters (e.g., the mean loss, $\mathcal{B}$, $\theta$ and $a$), we have the following observations.

- The retention point increases with respect to the mean loss. Usually a larger mean loss results in a larger premium. The DM may choose to increase the retention point to keep the premium affordable.

- A larger volatility aversion level leads to a smaller retention point. This is as expected since a smaller retention point implies that the DM cedes more risk to the insurer so that its retained risk is subject to less uncertainty.

- A larger safety loading parameter leads to a larger retention point. Relatively speaking, a larger safety loading parameter corresponds to a smaller volatility aversion level by looking at problem (5.2). Then as per the second bullet point, a smaller volatility aversion level leads to a larger retention point. From the perspective of premium, larger safety loading increases the premium. So the DM increases the retention point to maintain the affordable premium.

- A larger $a$ leads to a smaller retention point. It is straightforward that a larger $a$ leads to a smaller expected recovery rate, then the DM tends to cede out more risk to reduce the uncertainty of its retained risk.

## 6 Conclusions and future research

In this paper, we re-visit the optimal insurance problem from an insured's perspective under counterparty risk but within a mean-variance framework. Compared with the existing literature on this topic, the incentive compatibility is imposed in this paper to alleviate the possible *ex post* moral hazard issue. We assume that the insured is informative of the risk. As such the premium is calculated based on the coverage instead of the promised indemnity. Under incentive compatibility, the problem could be re-formulated as the one in terms of the marginal indemnity function. By applying the calculus of variations, the optimal marginal indemnity function, or the optimal indemnity function, could be characterized in an implicit manner. It is shown that the optimal indemnity function depends on both the first and second order conditional expectations of the random recovery rate. To make the implicit characterization applicable in practice, we deeply analyze the format of representation and unveil the explicit structure of optimal indemnity function implied by it. Two special cases are studied in detail: the case that the recovery rate is independent of the underlying loss and the case that the recovery rate is a decreasing function of the insurable loss. For both cases, we derive the optimal indemnity functions explicitly.

However, a major drawback of our main result is that the computation cost may be very high if the number of change points is large. This situation may happen when the dependence structure between the counterparty risk $Y$ and the insurable risk $X$ becomes rather complex. Future research is needed to further reduce the complexity of the optimization.

Our study could be extended by replacing the mean-variance criterion with other preference measures, such as expected utility or distortion risk measures. However, switching to the expected utility framework would make the implicit characterization of the optimal indemnity function more complicated, which brings non-trivial technical difficulties to the derivation of the explicit optimal indemnity function. In view of this, we decide to leave such extension for future research.

Besides the specific cases studied in this paper, another realistic case is that the default of insurer is driven by a portfolio of policies, and the decision maker in this paper is one of the claimants. In such a case, the recovery rate $Y$ can be modelled via an *ex post* proportional bankruptcy rule (Ibragimov et al., 2010; Boonen, 2019). The recovery rate $Y$ then depends on the indemnity functions of all policyholders. In this paper, we assume that $Y$ is independent of the indemnity function, and so we do not study this case. We leave this problem as a suggestion for future research.


## Acknowledgment

We are indebted to two anonymous reviewers for their comments which help improving the paper substantially. W.J. Jiang acknowledges the financial support received from the Natural Sciences and Engineering Research Council (RGPIN-2020-04204, DGECR-2020-00332) of Canada. W.J. Jiang is also grateful to the start-up grant received from the University of Calgary.


## Appendix

## A Proofs of the main results

### A.1 Proof of Lemma 3.1

The function $I \in \mathcal{I}$ is 1-Lipschitz continuous, and thus admits the following integral representation

$$I(x) = \int_0^x \eta(t)dt, \quad x \in [0, M],$$

where $\eta$ is called the marginal indemnity function (MIF) as per, for example, Assa (2015) and Zhuang et al. (2016). It is easily seen that seeking an optimal $I$ within $\mathcal{I}$ is equivalent to seeking an optimal $\eta$ within the class

$$\tilde{\mathcal{I}} = \left\{ \eta : [0,1] \mapsto [0,1] \;\middle|\; 0 \leq \eta(x) \leq 1 \text{ for any } x \in [0, M] \right\}.$$

Denote by $J(I)$ the objective function of problem (3.3). If $I^*$ is an optimal indemnity function, then given any $I \in \mathcal{I}$, we have $\epsilon I^* + (1-\epsilon)I \in \mathcal{I}$ for any $\epsilon \in [0,1]$. The first and second order derivatives of $J(\epsilon I^* + (1-\epsilon)I)$ with respect to $\epsilon$ are

$$\begin{aligned}\frac{dJ(\epsilon I^* + (1-\epsilon)I)}{\epsilon} =& 2\mathbb{E}[(\epsilon I^*(X) + (1-\epsilon)I(X)) \cdot (I^*(X) - I(X)) \cdot \psi_2(X)] \\ &- 2\mathbb{E}[X \cdot (I^*(X) - I(X)) \cdot \psi_1(X)] + \lambda \cdot \mathbb{E}[(I^*(X) - I(X)) \cdot \psi_1(X)],\end{aligned}$$

$$\frac{d^2 J(\epsilon I^* + (1-\epsilon)I)}{d\epsilon^2} = 2\mathbb{E}[(I^*(X) - I(X))^2 \cdot \psi_2(X)] \geq 0.$$

Therefore, $J(\epsilon I^* + (1-\epsilon)I)$ is convex with respect to $\epsilon$. It reaches its minimum at $\epsilon = 1$, and thus

$$
\begin{aligned}
&\frac{dJ(\epsilon I^* + (1-\epsilon)I)}{d\epsilon}\Big|_{\epsilon=1} \leq 0 \\
\Longrightarrow\ & 2\mathbb{E}[I^*(X)\cdot\big(I^*(X)-I(X)\big)\cdot\psi_2(X)] - 2\mathbb{E}[X\cdot(I^*(X)-I(X))\cdot\psi_1(X)] \\
&\quad + \lambda\cdot\mathbb{E}[(I^*(X)-I(X))\cdot\psi_1(X)] \leq 0 \\
\Longrightarrow\ & 2\mathbb{E}[I^*(X)^2\psi_2(X)] - 2\mathbb{E}[X\cdot I^*(X)\cdot\psi_1(X)] + \lambda\mathbb{E}[I^*(X)\cdot\psi_1(X)] \\
&\quad \leq 2\mathbb{E}[I^*(X)\cdot I(X)\cdot\psi_2(X)] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)] + \lambda\mathbb{E}[I(X)\cdot\psi_1(X)].
\end{aligned}
$$

This implies

$$
I^* = \arg\min_{I\in\mathcal{I}}\ 2\mathbb{E}[I^*(X)\cdot I(X)\cdot\psi_2(X)] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)] + \lambda\mathbb{E}[I(X)\cdot\psi_1(X)].
$$

Note that

$$
\begin{aligned}
& 2\mathbb{E}[I^*(X)\cdot I(X)\cdot\psi_2(X)] - 2\mathbb{E}[X\cdot I(X)\cdot\psi_1(X)] + \lambda\mathbb{E}[I(X)\cdot\psi_1(X)] \\
=\ & \int_0^M \big(2I^*(x)\psi_2(x) - 2x\psi_1(x) + \lambda\psi_1(x)\big)\, I(x)dF(x) \\
=\ & \int_0^M \big(2I^*(x)\psi_2(x) - 2x\psi_1(x) + \lambda\psi_1(x)\big)\left(\int_0^x \eta(t)dt\right) dF(x) \\
=\ & \int_0^M \left\{\int_t^M \big(2I^*(x)\psi_2(x) - 2x\psi_1(x) + \lambda\psi_1(x)\big)\, dF(x)\right\}\eta(t)dt \\
=\ & \int_0^M \left\{\int_t^M 2\psi_2(x)\left(I^*(x) - \frac{\psi_1(x)}{\psi_2(x)}(x-\frac{\lambda}{2})\right) dF(x)\right\}\eta(t)dt,
\end{aligned} \tag{A.1}
$$

where the third equation holds due to the Fubini's theorem. Now let

$$
L(t; I^*, \lambda) = \int_t^M \psi_2(x)\left(I^*(x) - \frac{\psi_1(x)}{\psi_2(x)}(x-\frac{\lambda}{2})\right) dF(x),
$$

it is straightforward that (A.1) gets minimized if its integrand function $2L(t; I^*, \lambda)\eta(t)$ gets minimized for each $t\in[0,M]$. Since $I\in\mathcal{I} \Longleftrightarrow \eta\in\tilde{\mathcal{I}}$, we have

$$
\eta^*(t) = \begin{cases} 1, & \text{if } L(t; I^*, \lambda) < 0, \\ \xi(t), & \text{if } L(t; I^*, \lambda) = 0, \\ 0, & \text{if } L(t; I^*, \lambda) > 0, \end{cases}
$$

where $\xi(t)\in[0,1]$ such that $\eta^*\in\tilde{\mathcal{I}}$. This ends the proof.

### A.2 Proof of Theorem 3.1

To prove (1), we first show by contradiction that there does not exist a point $t^*\in S_{m,j_m}$ such that $L(t^*; I^*, \lambda) > 0$ and $L'(t^*; I^*, \lambda)\geq 0$. Note that

$$
L'(t; I^*, \lambda) = -\psi_2(t)(I^*(t) - \phi_\lambda(t))f(t),
$$

and thus $L'(t^*; I^*, \lambda) \geq 0$ is equivalent to $I^*(t^*) - \phi_\lambda(t^*) \leq 0$. Since $\phi'_\lambda(x) > 1$ over $S_{m,j_m}$ if $j_m = 1$ and $I^{*\prime}(x)$ is always bounded by 0 and 1, $I^{*\prime}(t) - \phi'_\lambda(t) < 0$. Therefore $I^*(t) - \phi_\lambda(t) \leq 0$ for all $t \in [t^*, M]$, which implies that $L'(t; I^*, \lambda) \geq 0$ over $[t^*, M]$. However, this also implies that

$$L(t^*; I^*, \lambda) = L(M; I^*, \lambda) - \int_{t^*}^{M} L'(x; I^*, \lambda) dx = \int_{t^*}^{M} -L'(x; I^*, \lambda) dx \leq 0,$$

which contradicts with $L(t^*; I^*, \lambda) > 0$. Therefore, such $t^*$ does not exist. This implies that $L(t; I^*, \lambda)$ cannot up-cross the t-axis on $S_{m,j_m}$. Furthermore, $L(t; I^*, \lambda) = 0$ cannot hold over any sub-intervals of $S_{m,j_m}$ as otherwise

$$L'(t; I^*, \lambda) = 0 \quad \Longrightarrow \quad I^*(x) = \phi_\lambda(x)$$

over these sub-intervals. However, this contradicts with $I^{*\prime}(x) \in [0, 1]$ since $\phi'_\lambda(x) > 1$. Now define $t_0 = \inf\left\{t \in S_{m,j_m} : L(t; I^*, \lambda) \leq 0\right\}$, we have $L(t; I^*, \lambda) > 0$ for $t \in [x_{m-1}, t_0)$ and $L(t; I^*, \lambda) \leq 0$ for $t \in [t_0, M]$. As per Lemma 3.1, we get $\eta^*(x) = \mathbb{1}_{[t_0, M]}(x)$. The optimal indemnity function over $S_{m,j_m}$ in this case is given by

$$I^*(x) = I^*(x_{m-1}) + \int_{x_{m-1}}^{x} \eta^*(t) dt = I^*(x_{m-1}) + (x - t_0)_+.$$

To prove (2), we show by contradiction that there does not exist points $t^*, t^{**} \in S_{m,j_m}$ such that

$$L(t^*; I^*, \lambda) > 0 \quad \text{and} \quad L'(t^*; I^*, \lambda) \geq 0,$$

$$L(t^{**}; I^*, \lambda) < 0 \quad \text{and} \quad L'(t^{**}; I^*, \lambda) \leq 0.$$

If such $t^*$ exists, then from $L'(t^*; I^*, \lambda) \geq 0$ we get $I^*(t^*) - \phi_\lambda(t^*) \leq 0$. since $L(t^*; I^*, \lambda) > 0$, $I^{*\prime}(t^*) = 0$ as per Lemma 3.1. As such, $I^{*\prime}(t^*) - \phi'_\lambda(t^*) \leq 0$. This implies that $I^*(t) - \phi_\lambda(t) \leq 0$ for any $t \in [t^*, M]$. Therefore $L'(t; I^*, \lambda) \geq 0$ for any $t \in [t^*, M]$. This leads to

$$L(t^*; I^*, \lambda) = \int_{t^*}^{M} -L'(x; I^*, \lambda) dx \leq 0,$$

which contradicts with $L(t^*; I^*, \lambda) > 0$.

Similarly, if such $t^{**}$ exists, then from $L(t^{**}; I^*, \lambda) \leq 0$ we get $I^*(t^{**}) - \phi_\lambda(t^{**}) \geq 0$. Since $L(t^{**}; I^*, \lambda) < 0$, $I^{*\prime}(t^{**}) = 1$ as per Lemma 3.1. As such, $I^{*\prime}(t^{**}) - \phi'_\lambda(t^{**}) \geq 0$. This implies that $I^*(t) - \phi_\lambda(t) \geq 0$ for any $t \in [t^{**}, M]$. Therefore $L'(t; I^*, \lambda) \leq 0$ for any $t \in [t^{**}, M]$. This leads to

$$L(t^{**}; I^*, \lambda) = \int_{t^{**}}^{M} -L'(x; I^*, \lambda) dx \geq 0,$$

which contradicts with $L(t^{**}; I^*, \lambda) < 0$.

Based on the above findings, $L(t; I^*, \lambda)$ cannot cross the t-axis on $S_{m,j_m}$. Note that when $L(t; I^*, \lambda) = 0$ over any sub-intervals of $S_{m,j_m}$, we have $I^*(x) = \phi_\lambda(x)$ over those intervals. Now define $t_1 = \inf\left\{t \in S_{m,j_m} : L(t; I^*, \lambda) = 0\right\}$, then we have the following situations:

(i). $L(t; I^*, \lambda) > 0$ over $[x_{m-1}, t_1)$ and $L(t; I^*, \lambda) = 0$ over $[t_1, M]$. This leads to $\eta^*(x) = \phi'_\lambda(x)\mathbb{1}_{[t_1,M]}(x)$.

(ii). $L(t; I^*, \lambda) < 0$ over $[x_{m-1}, t_1)$ and $L(t; I^*, \lambda) = 0$ over $[t_1, M]$. This leads to $\eta^*(x) = \mathbb{1}_{[x_{m-1},t_1)}(x) + \phi'_\lambda(x)\mathbb{1}_{[t_1,M]}(x)$.

Applying the basic formula $I^*(x) = I^*(x_{m-1}) + \int_{x_{m-1}}^{x} \eta^*(t)dt$ leads to the result in (2).

To prove (3), we can show similarly that does not exist a point $t^{**} \in S_{m,j_m}$ such that $L(t^{**}; I^*, \lambda) < 0$ and $L'(t^{**}; I^*, \lambda) \leq 0$. As such, $L(t; I^*, \lambda)$ cannot down-cross the t-axis on $S_{m,j_m}$. Furthermore, $L(t; I^*, \lambda) = 0$ cannot hold on any sub-intervals of $S_{m,j_m}$ as otherwise $I^*(x) = \phi_\lambda(x)$ on those intervals, which is a contradiction since $I^{*\prime}(x) \in [0, 1]$ but $\phi'_\lambda(x) < 0$ in this case. Now define $t_2 = \inf\left\{t \in S_{m,j_m} : L(t; I^*, \lambda) \geq 0\right\}$, we have $L(t; I^*, \lambda) < 0$ for $t \in [x_{m-1}, t_2)$ and $L(t; I^*, \lambda) \geq 0$ for $t \in [t_2, M]$. According to Lemma 3.1, we have $\eta^*(x) = \mathbb{1}_{[x_{m-1},t_2)}(x)$, for which the corresponding $I^*$ is given by (3).

To prove (4), we note that for any $t \in [x_{i-1}, x_i]$

$$L(t; I^*, \lambda) = \int_t^{x_i} -L'(x; I^*, \lambda)dx + L(x_i; I^*, \lambda),$$

where $L'(t; I^*, \lambda) = -\psi_2(t)(I^*(t) - \phi_\lambda(t))f(t)$. If $j_i = 1$, then for any $t \in [x_{i-1}, x_i]$, $I^{*\prime}(t) - \phi'_\lambda(t) < 0$. We next focus on the case where the root of $I^*(t) = \phi_\lambda(t)$ exists on $(x_{i-1}, x_i)$. Other cases could be studied in a similar way and are thus omitted.

Denote by $t_{r_1}$ the root of $I^*(t) = \phi_\lambda(t)$ on $(x_{i-1}, x_i)$, then $L'(t; I^*, \lambda) < 0$ for $t \in [x_{i-1}, t_{r_1})$ and $L'(t; I^*, \lambda) > 0$ for $t \in (t_{r_1}, x_i]$. That means, $L(t; I^*, \lambda)$ can cross the t-axis at most twice and on $(x_{i-1}, t_{r_1})$ and $(t_{r_1}, x_i)$ respectively. Let

$$t_2 = \inf\left\{t \in [x_{i-1}, x_i] : L(t; I^*, \lambda) \leq 0\right\} \quad t_3 = \inf\left\{t \in [t_2, x_i] : L(t; I^*, \lambda) \geq 0\right\},$$

then $L(t; I^*, \lambda) > 0$ over $[x_{i-1}, t_2)$, $L(t; I^*, \lambda) < 0$ over $(t_2, t_3)$ and $L(t; I^*, \lambda) > 0$ over $(t_3, x_i]$. According to Lemma 3.1, we have $\eta^*(x) = \mathbb{1}_{(t_2,t_3)}(x)$. Applying the equation $I^*(x) = I^*(x_{i-1}) + \int_{x_{i-1}}^{x} \eta^*(t)dt$ leads to the result in (4).

To prove (5), we show by contradiction that there cannot exist two or more than two sub-intervals of $[x_{i-1}, x_i]$ such that $L(t; I^*, \lambda) = 0$. If there exist two sub-intervals, e.g. $[a, b]$ and $[c, d]$ where $x_{i-1} \leq a < b < c \leq x_i$, such that $L(t; I^*, \lambda) = 0$ for $t \in [a, b] \cup [c, d]$, then there must exist a point $t^*$ or $t^{**}$ in $(b, c)$ as described in the proof of (2). However, if $t^*$ exists, then similar to the proof of (2) we get $L'(t; I^*, \lambda) \geq 0$ over $[t^*, x_i]$. As such $L(t; I^*, \lambda) > 0$ for $t \in [t^*, x_i]$, which contradicts with $L(t; I^*, \lambda) = 0$ over $[c, d] \subseteq [t^*, x_i]$. If $t^{**}$ exists, then similar to the proof of (2) we get $L'(t; I^*, \lambda) \leq 0$ over $[t^{**}, x_i]$. As such $L(t; I^*, \lambda) < 0$ for $t \in [t^{**}, x_i]$, which also contradicts with $L(t; I^*, \lambda) = 0$ over $[c, d] \subseteq [t^{**}, x_i]$. Therefore, there exists at most one sub-interval, e.g. $[t_4, t_5] \subseteq [x_{i-1}, x_i]$, on which $L(t; I^*, \lambda) = 0$. We have four situations based on the sign of $L(t; I^*, \lambda)$ on $[x_{i-1}, t_4)$ and $(t_5, x_i]$:

(i). $L(t;I^*,\lambda)>0$ over $[x_{i-1},t_4)$ and $L(t;I^*,\lambda)>0$ over $(t_5,x_i]$. This leads to $\eta^*(x)=\phi'_\lambda(x)\mathbb{1}_{[t_4,t_5]}(x)$.

(ii). $L(t;I^*,\lambda)<0$ over $[x_{i-1},t_4)$ and $L(t;I^*,\lambda)>0$ over $(t_5,x_i]$. This leads to $\eta^*(x)=\mathbb{1}_{[x_{i-1},t_4)}(x)+\phi'_\lambda(x)\mathbb{1}_{[t_4,t_5]}(x)$.

(iii). $L(t;I^*,\lambda)>0$ over $[x_{i-1},t_4)$ and $L(t;I^*,\lambda)<0$ over $(t_5,x_i]$. This leads to $\eta^*(x)=\phi'_\lambda(x)\mathbb{1}_{[t_4,t_5]}(x)+\mathbb{1}_{(t_5,x_i]}(x)$.

(iv). $L(t;I^*,\lambda)<0$ over $[x_{i-1},t_4)$ and $L(t;I^*,\lambda)<0$ over $(t_5,x_i]$. This leads to $\eta^*(x)=\mathbb{1}_{[x_{i-1},t_4)}(x)+\phi'_\lambda(x)\mathbb{1}_{[t_4,t_5]}(x)+\mathbb{1}_{(t_5,x_i]}(x)$.

Applying the formula $I^*(x)=I^*(x_{i-1})+\int_{x_{i-1}}^x\eta^*(t)dt$ leads to the result in (5).

To prove (6), we note that $I^{*\prime}(t)-\phi'_\lambda(t)>0$ in this case. We next focus on the case where the root of $I^*(t)=\phi_\lambda(t)$ exists on $(x_{i-1},x_i)$. Other cases could be studied in a similar way and are thus omitted.

Denote by $t_{r_2}$ the root of $I^*(t)=\phi_\lambda(t)$ on $(x_{i-1},x_i)$, then $L'(t;I^*,\lambda)>0$ for $t\in[x_{i-1},t_{r_2})$ and $L'(t;I^*,\lambda)<0$ for $t\in(t_{r_2},x_i]$. That means $L(t;I^*,\lambda)$ can cross the t-axis at most twice and on $(x_{i-1},t_{r_2})$ and $(t_{r_2},x_i)$ respectively. Similar to the proof of (4), let

$$t_6=\inf\left\{t\in[x_{i-1},x_i]:L(t;I^*,\lambda)\geq 0\right\},\quad t_7=\inf\left\{t\in[t_6,x_i]:L(t;I^*,\lambda)\leq 0\right\},$$

then $L(t;I^*,\lambda)<0$ over $[x_{i-1},t_6)$, $L(t;I^*,\lambda)>0$ over $(t_6,t_7)$ and $L(t;I^*,\lambda)<0$ over $(t_7,x_i]$. According to Lemma 3.1, we have $\eta^*(x)=\mathbb{1}_{[x_{i-1},t_6)}(x)+\mathbb{1}_{(t_7,x_i]}(x)$. Applying the basic formula $I^*(x)=I^*(x_{i-1})+\int_{x_{i-1}}^x\eta^*(t)dt$ leads to the result in (6). This ends the proof.

## A.3 Proof of Proposition 3.1

Under the conditions of this proposition, for any $t\in[0,\frac{\lambda}{2})$,

$$I^*(t)-\frac{\psi_1(t)}{\psi_2(t)}(t-\frac{\lambda}{2})>0.$$

As such, for $t\in[0,\frac{\lambda}{2})$,

$$L'(t;I^*,\lambda)=-\psi_2(t)\left(I^*(t)-\frac{\psi_1(t)}{\psi_2(t)}(t-\frac{\lambda}{2})\right)f_X(t)<0.$$

Let $t_1=\inf\left\{t\in[0,\frac{\lambda}{2}]:L(t;I^*,\lambda)\leq 0\right\}$, then as per the monotonicity of $L(t;I^*,\lambda)$ over $[0,\frac{\lambda}{2})$, we have $L(t;I^*,\lambda)>0$ for $t\in[0,t_1)$ and $L(t;I^*,\lambda)<0$ for $t\in(t_1,\frac{\lambda}{2})$. Applying Lemma 3.1 gives $I^{*\prime}(x)=\mathbb{1}_{(t_1,\frac{\lambda}{2})}(x)$. As such, for any $x\in[0,\frac{\lambda}{2}]$

$$I^*(x)=\int_0^x I^{*\prime}(t)dt=(x-t_1)_+.$$

This finishes the proof.

## References


K. J. Arrow. Optimal insurance and generalized deductibles. Rand Corporation. Technical report, R-1108-OEO, 1973.

H. Assa. On optimal reinsurance policy with distortion risk measures and premiums. *Insurance: Mathematics and Economics*, 61:70–75, 2015.

C. Bernard and M. Ludkovski. Impact of counterparty risk on the reinsurance market. *North American Actuarial Journal*, 16(1):87–111, 2012.

T. J. Boonen. Equilibrium recoveries in insurance markets with limited liability. *Journal of Mathematical Economics*, 85:38–45, 2019.

T. J. Boonen and M. Ghossoub. On the existence of a representative reinsurer under heterogeneous beliefs. *Insurance: Mathematics and Economics*, 88:209–225, 2019.

K. Borch. *An attempt to determine the optimum amount of stop loss reinsurance.* 1960.

J. Cai, C. Lemieux, and F. Liu. Optimal reinsurance with regulatory initial capital and default risk. *Insurance: Mathematics and Economics*, 57:13–24, 2014.

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

K. C. Cheung, W. F. Chong, and A. Lo. Budget-constrained optimal reinsurance design under coherent risk measures. *Scandinavian Actuarial Journal*, 2019(9):729–751, 2019.

Y. Chi. Optimal reinsurance under variance related premium principles. *Insurance: Mathematics and Economics*, 51(2):310–321, 2012.

Y. Chi and K. S. Tan. Optimal reinsurance under VaR and CVaR risk measures: a simplified approach. *ASTIN Bulletin: The Journal of the IAA*, 41(2):487–509, 2011.

Y. Chi and K. S. Tan. Optimal incentive-compatible insurance with background risk. *ASTIN Bulletin: The Journal of the IAA*, 51(2):661–688, 2021.

Y. Chi and W. Wei. Optimum insurance contracts with background risk and higher-order risk attitudes. *ASTIN Bulletin: The Journal of the IAA*, 48(3):1025–1047, 2018.

Y. Chi and W. Wei. Optimal insurance with background risk: An analysis of general dependence structures. *Finance and Stochastics*, 24(4):903–937, 2020.

Y. Chi and S. C. Zhuang. Optimal insurance with belief heterogeneity and incentive compatibility. *Insurance: Mathematics and Economics*, 92:104–114, 2020.

J. D. Cummins and O. Mahul. Optimal insurance with divergent beliefs about insurer total default risk. *Journal of Risk and Uncertainty*, 27(2):121–138, 2003.

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

N. A. Doherty and H. Schlesinger. The optimal deductible for an insurance policy when initial wealth is random. *Journal of Business*, 56(4):555–565, 1983.

D. Filipović, R. Kremslehner, and A. Muermann. Optimal investment and premium policies under risk shifting and solvency regulation. *Journal of Risk and Insurance*, 82(2):261–288, 2015.

G. Franke, H. Schlesinger, and R. C. Stapleton. Multiplicative background risk. *Management Science*, 52(1):146–153, 2006.

M. Ghossoub. Budget-constrained optimal insurance with belief heterogeneity. *Insurance: Mathematics and Economics*, 89:79–91, 2019a.

M. Ghossoub. Optimal insurance under rank-dependent expected utility. *Insurance: Mathematics and Economics*, 87:51–66, 2019b.

C. Gollier. Optimum insurance of approximate losses. *Journal of Risk and Insurance*, pages 369–380, 1996.

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

R. Ibragimov, D. Jaffee, and J. Walden. Pricing and capital allocation for multiline insurance firms. *Journal of Risk and Insurance*, 77(3):551–578, 2010.

C. Li and X. Li. On the optimal risk sharing in reinsurance with random recovery rate. *Risks*, 6 (4):114, 2018.

S. C. Zhuang, C. Weng, K. S. Tan, and H. Assa. Marginal indemnification function formulation for optimal reinsurance. *Insurance: Mathematics and Economics*, 67:65–76, 2016.