OPTIMAL INCENTIVE-COMPATIBLE INSURANCE WITH BACKGROUND RISK

BY

YICHUN CHI AND KEN SENG TAN

ABSTRACT

In this paper, the optimal insurance design is studied from the perspective of an insured, who faces an insurable risk and a background risk. For the reduction of *ex post* moral hazard, alternative insurance contracts are asked to satisfy the principle of indemnity and the incentive-compatible condition. As in the literature, it is assumed that the insurer calculates the insurance premium solely on the basis of the expected indemnity. When the insured has a general mean-variance preference, an explicit form of optimal insurance is derived explicitly. It is found that the stochastic dependence between the background risk and the insurable risk plays a critical role in the insured's risk transfer decision. In addition, the optimal insurance policy can often change significantly once the incentive-compatible constraint is removed.

KEYWORDS

Background risk, conditional expectation function, incentive compatibility, mean-variance preference, optimal insurance design.

JEL codes: D81, G22

1. INTRODUCTION

The optimal risk sharing between an insured and an insurer (or between an insurer and a reinsurer) is one of the most important topics in insurance economics. Important breakthroughs linking insurance to general economic theory and the optimal design of risk sharing were made in the 1960s by a number of researchers including Borch (1960), Arrow (1963), and Mossin (1968). A fundamental assumption made in these studies and many other subsequent studies is that the risk faced by an individual is insurable. This implies that an individual could use any desirable insurance contract to insure his/her risk. In practice, an individual is also exposed to other sources of risk, which can be

Astin Bulletin 51(2), 661-688. doi:10.1017/asb.2021.7 © 2021 by Astin Bulletin. All rights reserved.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

attributed to the market, catastrophes, human capital, etc. These risks are typically uninsurable and are collectively denoted as the background risk. Because of the background risk, the insurance market is incomplete in the sense that it is impossible to provide an insurance coverage for all the risks faced by an individual.

As shown in Mayers and Smith (1983) and Doherty and Schlesinger (1983), the presence of background risk has a non-trivial effect on the demand for insurance. Their results have subsequently sparked considerable research interests. These results can be categorised according to model assumptions. One categorisation is the relationship between the existing insurable risk and the background risk, which can be additive or non-additive (Graf and Schulenburg, 1986; Gollier, 1996; Mahul, 2000; Vercammen, 2001; Dana and Scarsini, 2007; Franke *et al.*, 2011; Huang *et al.*, 2013). Another possible categorisation is based on the generality of the assumed set of admissible insurance contracts. Many results have been obtained by confining the analysis to a specific type of insurance policies such as stop-loss insurance and quota-share insurance (Eeckhoudt and Kimball, 1992). However, assuming a quite general set of alternative contracts and the additive relationship between the insurable risk and the background risk, Gollier (1996) concluded that the optimal insurance policy may be a disappearing deductible; see also Mahul (2000), Vercammen (2001) and Dana and Scarsini (2007).

Although the aforementioned works have made advances in the study of optimal insurance with background risk, many of the assumed models could be criticised for being either too restrictive or too unrealistic. In one extreme case, simplified assumptions are often imposed to ensure the tractability of the resulting model. It is common to confine the analysis of the optimal solution to a specific type of insurance policies. The remaining task is to determine the optimal parameter of the assumed policy (such as the optimal retention). Although such assumption leads to mathematical tractability, the drawback is the lack of generality. In another extreme case, if appropriate constraints are not imposed on alternative insurance policies, the resulting optimal insurance strategies could be unrealistic and, hence, hinder their usefulness in the study of optimal risk sharing. One example is that some results have indicated the optimality of the disappearing deductible. In a disappearing deductible, the risk retained by the insured may become less for a larger insurable loss. It encourages *ex post* moral hazard in the sense that the insured has an adverse incentive to report a larger loss (see Huberman *et al.*, 1983).

The above two extremes caution that some compromising constraints need to be imposed on admissible insurance policies to ensure the appropriateness and the usefulness of the resulting model. As argued in Huberman *et al.* (1983), an effective way of alleviating *ex post* moral hazard is to impose the incentive-compatible constraint. It ensures that both the insured and the insurer will pay more for a larger realisation of the insurable loss. For this reason, it is of significant interest to investigate the optimal incentive-compatible insurance for an insured who is exposed to both the insurable risk and the background risk.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

In this paper, we revisit the optimal insurance with background risk under the typical assumption that the insurance premium is calculated solely on the basis of the expected indemnity. A critical feature that distinguishes our proposed model from the existing ones is the presence of the incentive-compatible constraint on admissible insurance policies, in addition to the principle of indemnity. As in Gollier (1996), we assume that the insurable risk and the background risk are additive. When the insured has a general mean-variance preference, we derive an explicit form of optimal insurance. Another interesting finding of our result is that the optimal insurance solution depends heavily on the conditional expectation function (CEF) of the background risk with respect to the insurable risk. This, in turn, provides additional simplification for some classes of CEFs with simple types.

We now summarise the key contributions of this study. First, our result confirms the importance of the incentive-compatible condition when designing the optimal insurance with background risk. Although this condition was already introduced by Huberman *et al*. (1983) more than 30 years ago as a way of excluding *ex post* moral hazard, it is often neglected by researchers. There are two possible explanations to account for this phenomenon. One explanation is the increased complexity of the model attributed to the incentive-compatible condition. Another explanation is that some might argue this condition does not have a significant impact on the design of an optimal insurance policy in the absence of background risk. Hence, this condition is perceived to be redundant in the optimal insurance model set-up, even when a background risk is considered. The latter explanation, however, is unwarranted, and it is not true even in the absence of background risk (see Chi and Tan, 2011). By comparing our derived explicit optimal insurance form with those of Gollier (1996), Huang *et al*. (2013), and Dana and Scarsini (2007), we find that the incentive-compatible condition has a profound impact on the optimal insurance, particularly when the background risk and the insurable risk are negatively dependent.

Second, we find that the stochastic dependence between the background risk and the insurable risk plays a very critical role in the insured's risk transfer decision when the incentive-compatible constraint is imposed. We show that the optimal ceded loss function is piecewise and that the number of fragments depends crucially on the monotonicity of the CEF. If the sign of the derivative of the CEF changes infrequently, then the optimal insurance has a relatively simple form. When the CEF is a non-decreasing function, then the background risk cannot be used to hedge the insurable risk, and the stop-loss insurance is shown to be optimal. If they are extremely negatively dependent such that one plus the derivative of the CEF is always non-positive, the insurable risk can be naturally hedged by the background risk and it is prudent to have no insurance.

Finally, under the mean-variance preference of the insured, we derive the optimal insurance form explicitly irrespective of the dependence structure between the background risk and the insurable risk. It is quite different from the analysis under the expected utility framework, which has to be restricted

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

to some dependence cases (Lu *et al.*, 2012, 2018; Chi and Wei, 2018, 2020). The derived optimal form of insurance is often quite new and is seldom quota share. Thus, it seems inappropriate to restrict the analysis to the optimal quota-share insurance, as discussed in Mayers and Smith (1983) and Eeckhoudt and Kimball (1992).

The rest of the paper is organised as follows. Section 2 introduces an optimal insurance model with background risk and an incentive-compatible constraint. Using a constructive approach, Section 3 derives explicitly an optimal insurance form. Section 4 provides more insights into our results by discussing in detail the optimal insurance for some special classes of CEFs. Section 5 concludes the paper, and all the proofs are collected in Appendices.

2. Model formulation

Suppose that over a fixed time horizon, an insured faces two sources of risk: X and Y, where X is a non-negative insurable loss and Y is a background risk (which may be negative). Background risk Y can be an uninsurable risk, the catastrophic risk, market risk, the risk of human capital or the aggregation of these risks. Both of them are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with finite variance, that is, $var[X] < \infty$ and $var[Y] < \infty$, where $var[Z]$ is the variance of a random variable Z. We assume that the insured is interested in managing his/her risk exposure via an insurance policy, which is written on the insurable risk X. Under such an insurance arrangement, the insured cedes an amount of risk $I(X)$ to an insurer and retains the residual insurable risk $X - I(X)$. Thus, $I(x)$ and $x - I(x)$ are called the insured's ceded loss function and retained loss function, respectively. The problem of optimal insurance therefore boils down to seeking an optimal insurance policy $I^*(x)$.

As pointed out in the introduction, it is important to impose the principle of indemnity and the incentive-compatible condition on $I(x)$. Mathematically, the principle of indemnity asks $0 \le I(x) \le x$, and the latter condition implies that both the ceded and retained loss functions are non-decreasing. This, in turn, is equivalent to $I(x)$ being non-decreasing and Lipschitz-continuous. Formally, we have

$$0 \le I(x_1) - I(x_2) \le x_1 - x_2 \quad \text{for any} \quad 0 \le x_2 \le x_1. \tag{2.1}$$

Following Huberman *et al.* (1983), we define the set of admissible ceded loss functions as

$$\mathfrak{C} \triangleq \{0 \le I(x) \le x : \text{ both } I(x) \text{ and } x - I(x) \text{ are non-decreasing functions}\}. \tag{2.2}$$

There is a cost to the insured for transferring $I(X)$ to an insurer. This cost corresponds to the insurance premium that the insured needs to pay to the insurer. As in Gollier (1996), we assume that the insurer is risk-neutral

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

and the insurance premium is calculated solely on the basis of the expected indemnity. More specifically, if $\pi(\cdot)$ denotes the insurance premium principle of the insurer, then $\pi(I(X)) = \mathcal{P}(\mathbb{E}[I(X)])$ for some differentiable function $\mathcal{P}(\cdot)$ satisfying $\mathcal{P}(0) = 0$ and $\mathcal{P}'(x) \geq 1$ for any $x \geq 0$. In the special case where $\mathcal{P}(x) = (1+\theta)x$ for some $\theta > 0$, we recover the standard expected value premium principle. It should be emphasised that this assumption of premium principle plays an important role in the subsequent analysis and our results may not hold for other premium principles.

We now consider the impact of insurance. In the presence of insurance, the insured's total risk exposure, $\mathcal{T}_I(X;Y)$, is given by the sum of the retained insurable risk, the insurance premium and the background risk, that is,

$$\mathcal{T}_I(X;Y) = X - I(X) + \pi(I(X)) + Y. \tag{2.3}$$

The above decomposition highlights the intricate roles among various key factors. On one hand, if the insured is conservative, much (or all) of the risk can be transferred to an insurer by a meticulous choice of ceded loss functions, but at the expense of high cost in the form of insurance premium. On the other hand, if the insured has a higher risk tolerance, he/she could reduce the cost of insurance but would expose him/her to higher potential loss. This demonstrates the classical trade-off between the risk retention and the insurance premium paid out. The overall risk profile of the insured is further complicated by the presence of background risk, which can have a non-trivial effect depending on its relative magnitude as well as on its correlation with the insurable risk. According to these characteristics of $\mathcal{T}_I(X;Y)$, it is of interest to develop an optimal insurance strategy for the insured. This can be achieved via an optimisation problem such as minimising an appropriate risk measure on $\mathcal{T}_I(X;Y)$ or maximising the expected utility of the insured's final wealth.

When analysing the optimal insurance design under the expected utility framework, besides the difficulty in eliciting the utility function from the insured, the solution is seldom derived explicitly except for some special types of dependence between background risk and the insurable risk (see Chi and Wei, 2020). Motivated by the popularity of the mean-variance optimisation criterion in finance and insurance, this work studies the optimal insurance design by also assuming that the insured has a general mean-variance preference. This assumption is consistent with Mayers and Smith (1983). The general formulation of our proposed optimal insurance model can be stated as

$$\min_{I \in \mathfrak{C}} \mathcal{M}\big(\mathbb{E}\big[\mathcal{T}_I(X;Y)\big],\ var\big[\mathcal{T}_I(X;Y)\big]\big), \tag{2.4}$$

where $\mathcal{M}(x,y)$ is a non-decreasing function over $\mathbb{R}_+^2$. We emphasise that the above framework is quite general. For example, if $\mathcal{M}(x,y) = x + \frac{\mathcal{A}}{2}y$ for some $\mathcal{A} > 0$, then the above optimisation criterion is equivalent to the maximisation of the insured's mean-variance utility, that is, $\max_{I \in \mathfrak{C}} \mathbb{E}[W - \mathcal{T}_I(X;Y)] - \frac{\mathcal{A}}{2} var[W - \mathcal{T}_I(X;Y)]$, where W is the insured's initial wealth.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

3. Optimal insurance design

This section derives the solution to Problem (2.4). It is instructive to consider the first two moments of $\mathcal{T}_I(X;Y)$ as these moments directly affect the objective function of our proposed optimal insurance model. For any admissible ceded loss function $I(x) \in \mathfrak{C}$ and $Q(x) \triangleq \mathcal{P}(x) - x$ for all non-negative x, the mean of $\mathcal{T}_I(X;Y)$ is given by

$$\mathbb{E}[\mathcal{T}_I(X;Y)] = \mathbb{E}[X] + \mathbb{E}[Y] + Q(\mathbb{E}[I(X)]). \tag{3.1}$$

For the variance of $\mathcal{T}_I(X;Y)$, it is useful to introduce an auxiliary notation that describes the stochastic dependence between the insurable risk and the background risk. More specifically, we will use the CEF of the background risk with respect to the insurable risk, defined as

$$\psi(x) \triangleq \mathbb{E}[Y|X=x], \quad x \in \mathcal{S}(X), \tag{3.2}$$

where $\mathcal{S}(Z)$ is the support of a random variable Z. Especially, $\psi(x)$ is a constant when Y and X are independent, and it is well known that $\psi(x) = F_Y^{-1}(F_X(x))$ if Y and X are comonotonic and the cumulative distribution function $F_X(X)$ is continuous, where $F_Y^{-1}(p) \triangleq \inf\{y \in \mathbb{R} : F_Y(y) \geq p\}$ for $p \in [0,1]$ and $\inf \emptyset = \infty$ by convention. Following Hong *et al.* (2011), we say that the background risk Y is strongly positively expectation dependent (SPED) with respect to the insurable risk X if $\psi(x)$ is non-decreasing in x. It reflects a positive dependence structure between X and Y and yields non-negative covariance. Similarly, Y is said to be strongly negatively expectation dependent (SNED) with respect to X if $\psi(x)$ is non-increasing.

For any random variables Z_1 and Z_2 with finite second moments, we have the well-known decomposition $var(Z_1) = \mathbb{E}[var(Z_1|Z_2)] + var(\mathbb{E}[Z_1|Z_2])$, where $var(Z_1|Z_2)$ is the variance of Z_1 conditional on Z_2. With this decomposition, the variance of the insured's total risk exposure becomes

$$var[\mathcal{T}_I(X;Y)] = \mathbb{E}[var(Y|X)] + var(X + \psi(X) - I(X)). \tag{3.3}$$

By defining $(x)_+ \triangleq \max\{x, 0\}$, we can conclude that there exists a ceded loss function

$$I_\kappa(x) = \min\{(x + \psi(x) - \kappa)_+,\ x\} \quad \text{for some} \quad \kappa \in \mathbb{R} \tag{3.4}$$

such that

$$\mathbb{E}[I_\kappa(X)] = \mathbb{E}[I(X)] \quad \text{and} \quad var(X + \psi(X) - I_\kappa(X)) \leq var(X + \psi(X) - I(X)). \tag{3.5}$$

The proof of this statement is given in Appendix A. The above result is key to our approach to solving Problem (2.4). If we begin with an initial insurance policy $I(x)$, the above result asserts that there exists another insurance policy $I_\kappa(x)$,

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

which is at least as efficient as the initial $I(x)$. If $x+\psi(x)$ is non-decreasing and Lipschitz-continuous (satisfying (2.1)), then $I_\kappa(x)$ is an admissible ceded loss function and the problem of seeking an optimal insurance policy boils down to finding an optimal parameter κ of $I_\kappa(x)$. Unfortunately, the function $x+\psi(x)$ typically does not satisfy (2.1) such that $I_\kappa(x)\notin\mathfrak{C}$, and, hence, the optimal insurance must admit other forms. To make matters worse, the monotonic constraint (2.1) on insurance contracts implies that traditional methods such as the optimal control theory, the calculus of variation, and the Lagrangian multiplier approach are no longer applicable to solving Problem (2.4). More recently, a constructive approach has been shown to be very powerful in solving some other optimal constrained insurance problems. See, for example, Cheung (2010) and Chi and Tan (2011). In this study, we similarly adopt the constructive approach and show that it is also effective in solving Problem (2.4), even with the presence of background risk.

The dependence between X and Y, which is realised via the CEF $\psi(x)$ under the mean-variance framework, plays a critical role in the optimal insurance design. To simplify the analysis, we impose the following technical assumption:

Assumption 3.1. *$\mathcal{S}(X)=[0,\infty]$ and $\psi(x)$ is a continuously differentiable function.*

This assumption facilitates us to split $\mathcal{S}(X)$ according to the value of $\psi'(x)$. It, together with a constructive approach, allows us to devise a new insurance strategy that is at least as efficient as the original one. Notably, similar results can also often be obtained if this assumption is not met. We now present a three-step procedure to derive an optimal insurance form.

Step 1: Partition $\mathcal{S}(X)$ into a set of disjoint intervals, according to whether $x+\psi(x)$ is non-decreasing and Lipschitz-continuous at a neighbourhood of x or not. Using Assumption 3.1, we have

$$[0,\infty)=\bigcup_{i=1}^{m} S_{i,j_i}, \tag{3.6}$$

where m is a positive integer and $j_i\in\{1,2,3\}$ is determined by

$$\psi'(x)\begin{cases}\geq 0, & x\in S_{i,1};\\ \in(-1,0), & x\in S_{i,2};\\ \leq -1, & x\in S_{i,3}.\end{cases} \tag{3.7}$$

By design, m represents the minimum number of the partition satisfying (3.6) and (3.7). Moreover, the function $x+\psi(x)$ is increasing and Lipschitz-continuous over $S_{i,2}$, whereas it is non-increasing over $S_{i,3}$. Finally, conditional on $X\in S_{i,1}$, the background risk Y is SPED with respect to the insurable risk X.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

To further elaborate the additional properties of the above partition, we find it useful to introduce the following notation:

$$x_{i-1} \triangleq \inf\{x : x \in S_{i,j_i}\}, \qquad i = 1, \ldots, m \quad \text{and} \quad x_m = \infty. \tag{3.8}$$

Without loss of generality, we assume $0 = x_0 \le x_1 \le \cdots \le x_{m-1} < \infty$. It follows from the definition of m that $j_{i+1} \neq j_i$, $|j_{i+1} - j_i| = 1$ and $x_i = \sup\{x : x \in S_{i,j_i}\}$ for $i = 1, \ldots, m$. In our context, we refer to $\{x_i\}_{i=1}^{m-1}$ as the change points of $\psi'(x)$. Trivially, there are $m-1$ change points for a partition with m disjoint intervals. The example below exemplifies the idea of the partition.

Example 3.1. Let the CEF $\psi(x)$ have the following form:

$$\psi(x) = \frac{\rho}{\sigma} ln(x + \epsilon), \ x \ge 0,$$

for some $\rho \in [-1, 1]$ and positive ϵ and σ satisfying $0 < \epsilon\sigma < 1$. Hence, we have $\psi'(x) = \frac{\rho}{\sigma(x+\epsilon)}$. The partition of the set $[0, \infty)$ depends critically on the value of parameter ρ, as shown in the following three cases:

(i) If $\rho \ge 0$, then $\psi'(x) \ge 0$ for all non-negative x. In this case, we have $m = 1$ and $[0, \infty) = S_{1,1}$.
(ii) If $-\epsilon\sigma < \rho < 0$, then $m = 1$ and $[0, \infty) = S_{1,2}$.
(iii) If $-1 \le \rho \le -\epsilon\sigma$, then $\psi'(x)$ exhibits a change point at $-\frac{\rho}{\sigma} - \epsilon$ as

$$[0, \infty) = S_{1,3} \cup S_{2,2} = \left[0, -\frac{\rho}{\sigma} - \epsilon\right] \cup \left(-\frac{\rho}{\sigma} - \epsilon, \infty\right).$$

Step 2: Capitalising on the partition from Step 1 and the current insurance policy $I(x)$, we now describe a procedure to construct a desirable insurance policy $\tilde{I}(x)$ that reduces the insured's total risk exposure. The resulting $\tilde{I}(x)$ is a concatenation of m piecewise insurance policies such that

$$\tilde{I}(x_i) = I(x_i), \ i = 0, \ldots, m-1, \tag{3.9}$$

and that $\tilde{I}(X)$ and $I(X)$ have the same expected value over each interval S_{i,j_i}, that is,

$$\mathbb{E}[\tilde{I}(X); X \in S_{i,j_i}] \triangleq \mathbb{E}[\tilde{I}(X)\mathcal{I}_{\{X \in S_{i,j_i}\}}] = \mathbb{E}[I(X); X \in S_{i,j_i}]. \tag{3.10}$$

In the above expression, $\mathcal{I}_A$ denotes the indicator function of an event A.

For $0 \le a \le b$, let us denote $\mathcal{L}_{(a,b]}(Z)$ as a layer $(a, b]$ of risk Z with the following representation:

$$\mathcal{L}_{(a,b]}(Z) \triangleq \min\{(Z-a)_+, b-a\} = (Z-a)_+ - (Z-b)_+.$$

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

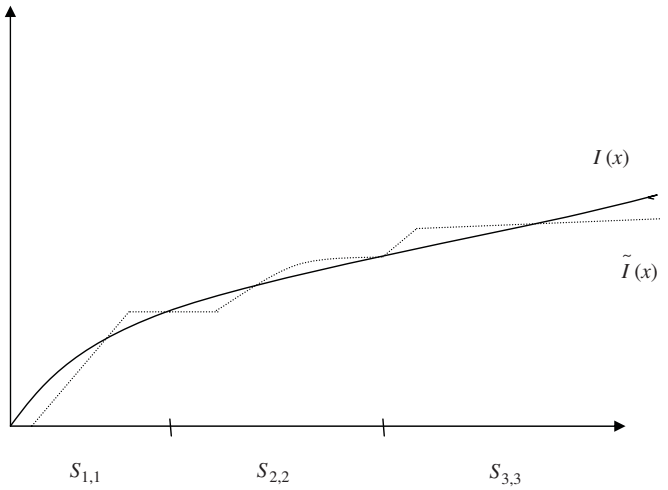


FIGURE 1: Illustration of the relationship between $\tilde{I}(x)$ (dotted line) and $I(x)$ (solid line).

Then, the desirable ceded loss function can be constructed as

$$\tilde{I}(x) \triangleq \begin{cases} I(x_{i-1}) + \mathcal{L}_{(y_{i,1}, y_{i,1}+(I(x_i)-I(x_{i-1}))]}(x), & x \in S_{i,1}; \\ \min\{\max\{x+\psi(x)-\lambda_i, I(x_i)+x-x_i, I(x_{i-1})\}, \\ \qquad I(x_{i-1})+x-x_{i-1}, I(x_i)\}, & x \in S_{i,2}; \\ I(x_{i-1})+x-x_{i-1}-\mathcal{L}_{(y_{i,3}, y_{i,3}+(x_i-x_{i-1})-(I(x_i)-I(x_{i-1}))]}(x), & x \in S_{i,3} \end{cases} \tag{3.11}$$

for $i = 1, \ldots, m-1$, and

$$\tilde{I}(x) \triangleq \begin{cases} I(x_{m-1}) + (x - y_{m,1})_+, & x \in S_{m,1}; \\ I(x_{m-1}) + \min\{(x+\psi(x)-\lambda_m)_+, x - x_{m-1}\}, & x \in S_{m,2}; \\ I(x_{m-1}) + \mathcal{L}_{(x_{m-1}, y_{m,3}]}(x), & x \in S_{m,3}, \end{cases} \tag{3.12}$$

where the parameters $y_{i,1} \in [x_{i-1},\ x_i - I(x_i) + I(x_{i-1})]$, $y_{i,3} \in [x_{i-1},\ x_{i-1} + (I(x_i) - I(x_{i-1}))]$ and $\lambda_i \in \mathbb{R}$ are determined by (3.10). Obviously, we have $\tilde{I} \in \mathfrak{C}$. Figure 1 provides a graphical comparison between $I(x)$ and $\tilde{I}(x)$ for a particular CEF with the partition $[0, \infty) = S_{1,1} \cup S_{2,2} \cup S_{3,3}$.

Step 3: In the final step, we compare the constructed ceded loss function $\tilde{I}(x)$ from Step 2 with $I_\kappa(x)$ in (3.4). When the CEF $\psi(x)$ plus the identity function is a non-decreasing and Lipschitz-continuous function, they are exactly the same. However, for other cases, $\tilde{I}(x)$ appears much more complicated than $I_\kappa(x)$, and they also have a slight difference even over $S_{i,2}$ for $1 \le i \le m-1$. In fact, noting that $\tilde{I}(x)$ has the same values as $I(x)$ at the end points of S_{i,j_i}, it follows from (2.1) that:

$$\max\{I(x_i)+x-x_i, I(x_{i-1})\} \le \tilde{I}(x) \le \min\{I(x_{i-1})+x-x_{i-1}, I(x_i)\} \text{ for all } x \in S_{i,2},$$

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

which leads to the slight deviation. With this constraint, $\tilde{I}(x)$ can be obtained using state-wise optimisation approach.

The monotonicity of the function $x+\psi(x)$ also dictates the optimal insurance form. Recall that, over $S_{i,1}$, $\psi(x)$ is non-decreasing such that $x-I(x)$ and $\psi(x)$ are comonotonic. Using Lemma 3 in Ohlin (1969), we can easily verify that, conditional on $X\in S_{i,1}$, $X-I(X)+\psi(X)$ is less dispersed than $X-\tilde{I}(X)+\psi(X)$. Equivalently, conditional on $X\in S_{i,1}$, $X-I(X)+\psi(X)$ is larger than $X-\tilde{I}(X)+\psi(X)$ in the convex order. It is denoted as $[X-I(X)+\psi(X)|X\in S_{i,1}]\geq_{cx}[X-\tilde{I}(X)+\psi(X)|X\in S_{i,1}]$, where $Z_1\geq_{cx}Z_2$ means $\mathbb{E}[g(Z_1)]\geq\mathbb{E}[g(Z_2)]$ for any convex function g provided that their expectations exist. Noting that variance preserves the convex order, we can conclude that it is better for the insured to cede the middle part of the risk over $S_{i,1}$. Similarly, for $S_{i,3}$, it can be found that, conditional on $X\in S_{i,3}$, $I(X)-[X+\psi(X)]$ is less dispersed than $\tilde{I}(X)-[X+\psi(X)]$, and, hence, the insured is encouraged to retain the middle risk. These analyses explain why we construct $\tilde{I}(x)$ from $I(x)$ in this way. By comparing the current insurance policy with the newly constructed one, we obtain the main result of the paper. This is stated in the following theorem and its proof is relegated to Appendix B.

Theorem 3.1. *Under Assumption 3.1, the admissible ceded loss function $I(x)$ is dominated by $\tilde{I}(x)$ in the sense of*

$$\mathcal{M}\left(\mathbb{E}\left[\mathcal{T}_{\tilde{I}}(X;Y)\right],\ var\left[\mathcal{T}_{\tilde{I}}(X;Y)\right]\right)\leq\mathcal{M}\left(\mathbb{E}\left[\mathcal{T}_{I}(X;Y)\right],\ var\left[\mathcal{T}_{I}(X;Y)\right]\right). \quad (3.13)$$

The above theorem formally establishes that any given admissible insurance policy is suboptimal to an insurance contract with the ceded loss function $\tilde{I}(x)$ stipulated in (3.11) and (3.12). This theorem also highlights a few other important findings. First, the optimal ceded loss function $\tilde{I}(x)$ heavily relies on the CEF $\psi(x)$. In other words, the stochastic dependence between the background risk and the insurable risk plays a very critical role in the insured's optimal risk transfer. Second, the incentive-compatible constraint often has a non-trivial impact on the optimal insurance design. More precisely, the above comparison between $\tilde{I}(x)$ and $I_\kappa(x)$ illustrates that our optimal insurance form is typically different from those obtained by Gollier (1996) and Huang *et al.* (2013) without the incentive-compatible constraint.

It is also worth noting that, from (3.11) and (3.12), all the newly constructed ceded loss functions have the same form involving a finite number of free parameters, namely, $\tilde{I}(x_i)$, $y_{i,1}$, $y_{i,3}$ and λ_i. Consequently, the optimal insurance policy can be completely specified once these optimal parameters are known. This finding also highlights the effectiveness of the constructive approach. With this approach, the study of the infinite-dimensional optimal insurance model (2.4) simplifies to deriving the optimal $\tilde{I}(x)$, which is equivalent to solving a minimisation problem with less than $2m-1$ decision variables. For a large m, it seems very challenging to solve this minimisation problem analytically and

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

may need to resort to some numerical procedures. However, the CEFs considered in the literature tend to have simple structures with a relatively small m. Thus, it will be illuminating to analyse the derivation and the implication of the optimal insurance policy for the CEFs whose derivative has less than one change point. This is discussed in detail in the next section.

4. Special cases

4.1. The case of no change point ($m = 1$)

In this subsection, we assume that the derivative of the CEF $\psi(x)$ has no change point, that is, $m = 1$. By Theorem 3.1, the analysis of the optimal insurance model (2.4) is reduced to studying a minimisation problem with one decision variable. As shown in Appendix C, the problem for this special case can be solved easily and the optimal insurance contracts for all three cases are collected in the following proposition.

Proposition 4.1.

(i) If the background risk Y is SPED with respect to the insurable risk X, then the stop-loss insurance, which is full insurance above a deductible, is optimal. In particular, full insurance is optimal under a fair premium.

(ii) If Y and $X + Y$ are SNED and SPED with respect to X, respectively, then, for some $\lambda \geq \psi(0)$, an optimal ceded loss function can be in the form of

$$I^*(x) = (x + \psi(x) - \lambda)_+, \ x \geq 0.$$

Furthermore, given a fair premium, the optimal ceded loss function is given by

$$I^*(x) = x + \psi(x) - \psi(0).$$

(iii) If $X + Y$ is SNED with respect to X, then no insurance is optimal.

Recall that the optimal insurance policies obtained in Dana and Scarsini (2007) do not take into consideration the elimination of *ex post* moral hazard. It is therefore instructive to compare the optimal insurance policies given by the above proposition with that from Dana and Scarsini (2007) to highlight the importance of the incentive-compatible constraint.

(1) If Y is stochastically increasing in X (denoted by $Y \uparrow_{st} X$), that is, $\mathbb{E}[u(Y)|X = x]$ is non-decreasing over $\mathcal{S}(X)$ for any non-decreasing function u such that the expectation exists, then Y is SPED with respect to X. For this case, Proposition 4.4 in Dana and Scarsini (2007) shows that the optimal insurance contract is a disappearing deductible, which could lead to *ex post* moral hazard. In contrast, when the incentive-compatible

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

constraint is imposed, the stop-loss insurance is optimal, as asserted in the above proposition. That is, the indemnity is restricted to partially compensate for the insurable loss. It should be emphasised that, although in this case we obtain the same optimal insurance form as that of Arrow (1963) in the absence of background risk Y, the optimal deductible is often different and depends explicitly on the joint distribution of Y and X. This result can be explained: since the background risk cannot be used to hedge the insurable risk X, then it remains optimal for the insured to cede all the tail risk to an insurer just like the case in the absence of background risk.

(2) When $X+Y$ is stochastically decreasing in X (denoted by $X+Y\downarrow_{st}X$), it is easy to show that $X+Y$ is SNED with respect to X. For this case, Propositions 6.2 and 6.4 in Dana and Scarsini (2007) confirm that neither no insurance nor full insurance is optimal and that the optimal ceded loss function is decreasing over some interval. However, part (iii) of Proposition 4.1 demonstrates that no insurance is optimal when the incentive-compatible constraint is imposed. This result, in fact, is intuitive. Because $X+Y$ is SNED with respect to X, the insurable risk becomes a natural hedge to the background risk and, as such, any insurance purchase may even be harmful.

(3) When $Y\downarrow_{st}X$ and $X+Y\uparrow_{st}X$, it follows from Proposition 5.1 in Dana and Scarsini (2007) that the optimal insurance solution would satisfy the incentive-compatible constraint. Thus, for this case, we have the same design of optimal insurance. For this case, the background risk provides a partial hedge to the insurable risk and the residual risk $X+\psi(X)$ has to be ceded by a stop-loss contract. Notably, the optimality of partial insurance above a deductible is also obtained in Vercammen (2001) with a non-separable background risk.

The above comparisons clearly signify the importance of the incentive-compatible constraint. The design of an optimal insurance policy can be remarkably different, depending on the incentive-compatible constraint. The presence of background risk can further affect the insured's risk transfer decision.

4.2. The case of one change point ($m=2$)

In this subsection, we extend the analysis by assuming that $\psi'(x)$ has one change point, that is, $m=2$. Thus, the set $[0,\infty)$ can only be partitioned in the following four ways:

$$S_{1,3}\cup S_{2,2},\quad S_{1,1}\cup S_{2,2},\quad S_{1,2}\cup S_{2,1},\quad S_{1,2}\cup S_{2,3}.$$

In view of the length of the paper, we focus on deriving the optimal insurance solution only for the first two cases. In contrast to the cases with no change point, the analysis of optimal insurance design becomes much complicated because the hedge of background risk to the insurable risk changes

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

over different parts of X. Fortunately, by Theorem 3.1, the optimal insurance model (2.4) boils down to a minimisation problem with three decision variables. Although this result has greatly simplified the difficulty of solving the original optimal insurance model (2.4), it should be emphasised that solving a three-dimensional optimisation problem can still be quite challenging, even when resorting to numerical procedures. Recognising the potential challenge, we attempt to further reduce the dimensionality of the minimisation problem from three to two. This is asserted in the following Propositions 4.2 and 4.3 (see Appendices D and E for the proofs).

Proposition 4.2. *For the CEF $\psi(x)$ with the partition $[0,\infty)=S_{1,3}\cup S_{2,2}$, a solution to the optimal insurance model (2.4) could be written by*

$$I^*(x)=\mathcal{L}_{(d,x_1]}(x)+(x+\psi(x)-c)_+\mathcal{I}_{\{x>x_1\}},$$

for some $0\le d\le x_1$ and $c\ge x_1+\psi(x_1)$, where x_1 is defined in (3.8). If c satisfies

$$\mathbb{E}\left[X+\psi(X)|X\le x_1\right]\ge\mathbb{E}\left[\min\{X+\psi(X),c\}|X>x_1\right],\tag{4.1}$$

then we have $d=x_1$.

Proposition 4.3. *If the set $[0,\infty)$ is partitioned into $S_{1,1}\cup S_{2,2}$ according to the CEF $\psi(x)$, then any given admissible ceded loss function is suboptimal to*

$$I^*(x)=\mathcal{L}_{(d,z]}(x)+\big(x+\psi(x)-(z+\psi(z))\big)_+,\tag{4.2}$$

for some $z\ge x_1$ and $d\in[0,x_1]\cup\{z\}$.

For both cases, we can see that the residual risk $X+\psi(X)$ after the hedge of background risk is ceded by a stop-loss insurance policy over $[x_1,\infty)$. This is consistent with Proposition 4.1(ii). However, the monotonicity of $\psi(x)$ over $[0,x_1)$ does affect the insurance strategy on the risk beside x_1. More precisely, this part of risk is often not ceded for the former case, especially when (4.1) is satisfied, but it is transferred to some degree in the latter case. These two propositions show that the optimal insurance contract can be in the form of a combination of layer insurance for middle risk and partial insurance for tail risk, which is quite different from previous studies except Chi and Wei (2020). More precisely, in the absence of background risk, layer insurance and partial insurance above a deductible are often shown to be optimal, especially under some risk measure; see, for example, Raviv (1979), Chi and Tan (2011), Chi (2012). The incorporation of background risk complicates the form of optimal insurance.

To gain more insights into the optimal insurance policy dictated by Proposition 4.2 and how the optimal insurance is affected by the degree of the insured's risk aversion and the stochastic dependence between the background risk and the insurable risk, we conclude the section by providing the following example:

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

Example 4.1. Setting $\mathcal{M}(x,y)=x+\frac{\mathcal{A}}{2}y$, we consider to maximise the insured's mean-variance utility with a risk aversion coefficient $\mathcal{A}>0$. We assume that the insurance premium is calculated by the expected value principle with a safety loading coefficient θ, and that the insured shorts some shares of insurance companies which underwrite the same type of risks. The short position provides a way of partially hedging insured's insurable loss, in addition to an insurance contract. For simplicity, we assume the investment risk Y is normally distributed and the insurable risk X is lognormally distributed. These are commonly adopted assumptions in the respective practice. See, for example, Markowitz (1952) for the normally distributed investment risk and Marlin (1984) for the lognormally distributed insurance risk. With these assumptions, we have

$$Y=\mu_1+\sigma_1 e_Y,$$

where $\mu_1\in\mathbb{R}$, $\sigma_1>0$ and e_Y follows the standard normal distribution, and

$$lnX=\mu_2+\sigma_2 e_X,$$

where $\mu_2\in\mathbb{R}$ and $\sigma_2>0$. Here, (e_X,e_Y) follows a two-dimensional standard normal distribution with a correlation coefficient $\rho\in[-1,0)$.

With the above assumptions, it is easy to see that the background risk is SNED with respect to the insurable loss X. Moreover, we have

$$\psi(x)=\mu_1+\sigma_1\mathbb{E}\left[e_Y|e_X=\frac{lnx-\mu_2}{\sigma_2}\right]=\frac{\rho\sigma_1}{\sigma_2}(lnx-\mu_2)+\mu_1,\ x>0,$$

which, in turn, implies $(0,\infty)=S_{1,3}\cup S_{2,2}=\left(0,-\frac{\rho\sigma_1}{\sigma_2}\right]\cup\left(-\frac{\rho\sigma_1}{\sigma_2},\infty\right)$.

From (3.1) and (3.3), we can see that the value of μ_1 has no effect on the optimal insurance design under the above assumption. Without loss of generality, we set $\mu_1=0$ to simplify the numerical calculation. By Proposition 4.2, we know that a solution to Problem (2.4) is

$$I_{d,c}(x)\triangleq\mathcal{L}_{(d,x_1]}(x)+(x+\psi(x)-c)_+\mathcal{I}_{\{x>x_1\}} \tag{4.3}$$

for some $0\le d\le x_1$ and $c\ge x_1+\psi(x_1)$. Therefore, the residual task is to derive the optimal d and c, which is equivalent to solving the following minimisation problem:

$$\min_{\substack{0\le d\le -\frac{\rho\sigma_1}{\sigma_2}\\ c\ge -\frac{\rho\sigma_1}{\sigma_2}+\psi(-\frac{\rho\sigma_1}{\sigma_2})}}\ \mathcal{B}(d,c),$$

where

$$\begin{aligned}\mathcal{B}(d,c)&\triangleq\mathbb{E}[\mathcal{T}_{I_{d,c}}(X;Y)]+\frac{\mathcal{A}}{2}var[\mathcal{T}_{I_{d,c}}(X;Y)]-\mathbb{E}[X]-\mathbb{E}[Y]-\frac{\mathcal{A}}{2}\mathbb{E}[var[Y|X]]\\&=\theta\left(\int_d^{-\frac{\rho\sigma_1}{\sigma_2}}\mathbb{P}(X>t)\mathrm{d}t+\int_c^{\infty}\mathbb{P}\left(X+\psi(X)>y,X>-\frac{\rho\sigma_1}{\sigma_2}\right)\mathrm{d}y\right)\end{aligned}$$

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

TABLE 1

EFFECTS OF THE INSURED'S RISK AVERSION AND THE CORRELATION COEFFICIENT ρ ON THE VALUE OF $(d^* \times 1000, \frac{c(d^*)}{10})$ FOR $\sigma_1 = \sigma_2 = 2$, $\theta = 0.1$ AND $\mu_2 = 14$.

	$\mathcal{A}=1$	$\mathcal{A}=2$	$\mathcal{A}=3$	$\mathcal{A}=4$
$\rho=-0.1$	(97.81, 115.3283)	(97.81, 91.3355)	(97.81, 79.8424)	(97.81, 72.6376)
$\rho=-0.2$	(179.49, 115.4029)	(179.49, 91.4123)	(179.49, 79.9204)	(179.49, 72.7166)
$\rho=-0.3$	(278.99, 115.4774)	(278.99, 91.489)	(278.99, 79.9984)	(294.61, 72.7955)
$\rho=-0.4$	(376.69, 115.552)	(376.69, 91.5658)	(376.69, 80.0765)	(393.11, 72.8744)
$\rho=-0.5$	(495.39, 115.6265)	(495.39, 91.6426)	(495.39, 80.1545)	(495.39, 72.9534)
$\rho=-0.6$	(507.22, 115.7011)	(507.22, 91.7193)	(507.22, 80.2325)	(507.22, 73.0323)
$\rho=-0.7$	(655.57, 115.7756)	(655.57, 91.7961)	(655.57, 80.3106)	(649.42, 73.1112)
$\rho=-0.8$	(753.27, 115.8502)	(753.27, 91.8728)	(753.27, 80.3886)	(769.69, 73.1901)
$\rho=-0.9$	(850.97, 115.9248)	(850.97, 91.9496)	(850.97, 80.4666)	(850.97, 73.2691)
$\rho=-1$	(993.7, 115.9993)	(993.7, 92.0263)	(966.02, 80.5447)	(966.02, 73.348)

$$+\frac{\mathcal{A}}{2}var\big(X+\psi(X)-(X+\psi(X)-c)_+\mathcal{I}_{\{X>-\frac{\rho\sigma_1}{\sigma_2}\}}\big)$$
$$+\mathcal{A}\mathbb{E}\left[(X+\psi(X))\left(-\frac{\rho\sigma_1}{\sigma_2}-d-\mathcal{L}_{(d,-\frac{\rho\sigma_1}{\sigma_2}]}(X)\right)\right]+\frac{\mathcal{A}}{2}var(\mathcal{L}_{(d,-\frac{\rho\sigma_1}{\sigma_2}]}(X))$$
$$-\mathcal{A}\int_d^{-\frac{\rho\sigma_1}{\sigma_2}}\mathbb{P}(X\le t)\mathrm{d}t\mathbb{E}\left[X+\psi(X)-(X+\psi(X)-c)_+\mathcal{I}_{\{X>-\frac{\rho\sigma_1}{\sigma_2}\}}\right].$$

Differentiating the objective function $\mathcal{B}(d,c)$ with respect to c yields

$$\frac{\partial\mathcal{B}(d,c)}{\partial c}=\mathcal{A}\mathbb{P}\left(X+\psi(X)>c, X>-\frac{\rho\sigma_1}{\sigma_2}\right)V(d,c),\ a.e., \tag{4.4}$$

where $V(d,c)\triangleq-\mathbb{E}[X+\psi(X)]-\frac{\theta}{\mathcal{A}}-\int_d^{-\frac{\rho\sigma_1}{\sigma_2}}\mathbb{P}(X\le t)\mathrm{d}t+c+\mathbb{E}\big[(X+\psi(X)-c)_+\mathcal{I}_{\{X>-\frac{\rho\sigma_1}{\sigma_2}\}}\big]$. It is easy to see that $V(d,c)$ is a non-decreasing continuous function of c with $\lim_{c\to\infty}V(d,c)=\infty$; then, it follows from (4.4) that, given $d\in[0,-\frac{\rho\sigma_1}{\sigma_2}]$, the minimal value of $\mathcal{B}(d,c)$ is attainable at

$$c(d)=\inf\left\{c\ge-\frac{\rho\sigma_1}{\sigma_2}+\psi\left(-\frac{\rho\sigma_1}{\sigma_2}\right):V(d,c)\ge0\right\}.$$

In what follows, we can derive the optimal d^* by solving the minimisation problem $\min_{0\le d\le-\frac{\rho\sigma_1}{\sigma_2}}\mathcal{B}(d,c(d))$. As a result, an optimal ceded loss function is given by

$$I^*(x)=\mathcal{L}_{(d^*,-\frac{\rho\sigma_1}{\sigma_2}]}(x)+(x+\psi(x)-c(d^*))_+\mathcal{I}_{\{x>-\frac{\rho\sigma_1}{\sigma_2}\}},\ x\ge0.$$

The numerical results for $(d^*, c(d^*), \mathcal{B}(d^*, c(d^*)))$ are given in Tables 1 and 2. The first table studies the effects of the insured's risk aversion and the stochastic

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

TABLE 2

CHANGES IN $(d^* \times 1000, \frac{c(d^*)}{10}, \frac{\mathcal{B}(d^*,c(d^*))}{10{,}000})$ WITH RESPECT TO THE SAFETY LOADING COEFFICIENT θ AND THE MEAN OF lnX BY ASSUMING $\sigma_1 = \sigma_2 = 2$, $\mathcal{A} = 2$ AND $\rho = -0.5$

	$\theta = 0.05$	$\theta = 0.1$	$\theta = 0.15$	$\theta = 0.2$
$\mu_2 = 8$	(499.99, 1.823, 1.1006)	(499.99, 2.292, 2.2009)	(499.99, 2.642, 3.301)	(499.99, 2.931, 4.4009)
$\mu_2 = 10$	(499.98, 5.634, 8.136)	(499.98, 7.173, 16.27)	(499.98, 8.297, 24.404)	(499.98, 9.216, 32.537)
$\mu_2 = 12$	(499.29, 19.612, 60.123)	(499.29, 24.866, 120.242)	(498.11, 28.648, 180.359)	(498.11, 31.713, 240.474)
$\mu_2 = 14$	(473.93, 72.953, 444.277)	(495.39, 91.643, 888.541)	(495.39, 104.935, 1332.8)	(495.39, 115.627, 1777)
$\mu_2 = 16$	(349.79, 283.066, 3282.8)	(349.79, 352.06, 6565.7)	(349.79, 400.647, 9848.5)	(349.79, 439.476, 13,131)

dependence between the background risk and the insurable risk on the optimal insurance design by setting $\sigma_1 = \sigma_2 = 2, \theta = 0.1$ and $\mu_2 = 14$. Numerical results illustrate that, when $\mathcal{A}$ is increasing, $c(d^*)$ is decreasing, whereas d^* is usually insensitive to this change. Thus, we can conclude that, as the insured becomes more risk-averse, he/she would cede more insurable risk above $-\frac{\rho\sigma_1}{\sigma_2}$. The numerical results also highlight the sensitivity of the correlation coefficient ρ to d^* but not to $c(d^*)$. In fact, as ρ decreases from -0.1 to -1, d^* increases from just under 0.1 to almost 1, irrespective of the risk aversion of the insured. This is consistent with our intuition in that as the shorting investment becomes more efficient in hedging the insurable loss, the insured will reduce the demand for insurance.

Table 2 shows the impacts of θ and μ_2 by assuming $\sigma_1 = \sigma_2 = 2, \mathcal{A} = 2$ and $\rho = -0.5$. It is easy to see that both $\mathcal{B}(d^*, c(d^*))$ and $c(d^*)$ are increasing in θ and μ_2. Furthermore, as insurance becomes more expensive, the insured would retain more risk at the expense of higher risk exposure. As μ_2 increases, d^* often decreases. Specially, $d^* \approx \bar{x}_1 = 0.5$ for $\mu_2 = 8$. This result matches with Proposition 4.2 when the condition (4.1) is satisfied.

5. Concluding remarks

In this paper, the design of an optimal insurance policy with background risk is analysed from the perspective of an insured, who has a general mean-variance preference. As in the literature, the insurance premium is assumed to be calculated based only on the expected indemnity. When admissible insurance policies are asked to satisfy the incentive-compatible condition and the principle of indemnity, an optimal insurance form is derived explicitly irrespective of the dependence structure between the background risk and the insurable risk. The derived optimal insurance policy can deviate from the existing ones in the literature. This highlights the importance of the background risk and the incentive-compatible constraint.

When the background risk is stochastically increasing in the insurable risk, Lemma 3.3 in Cai and Wei (2012) can be used to easily extend our results to the optimisation criteria that preserve the convex order. Remarkably, such extension seems very challenging for other dependence structures. Although the analysis in this paper is confined to the mean-variance preference of the insured, we believe that our results can provide constructive insights to finding optimal insurance contracts for other optimisation criteria. Especially, it is of interest to investigate the optimal insurance design by taking into consideration higher moments of the insured's risk exposure, as the insurable risk often has a heavy-tailed distribution. We leave this to future research. It is necessary to point out that, following this work, Lu *et al.* (2018), Chi and Wei (2018) and Chi and Wei (2020) have made contributions to this problem within the expected utility framework.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

ACKNOWLEDGEMENTS

We would like to thank the Editor-in-Chief and three anonymous reviewers for their valuable comments, which greatly improve the presentation of this paper. This work was supported by grants from the MOE (China) Project of the Key Research Institute of Humanities and Social Sciences at Universities (no. 16JJD790061), National Natural Science Foundation of China (no. 11971505), the Natural Sciences and Engineering Research Council of Canada (no. RGPIN-216949-2013) and the 111 Project of China (no. B17050).

REFERENCES

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

BORCH, K. (1960) An attempt to determine the optimum amount of stop loss reinsurance. In *Transactions of the 16th International Congress of Actuaries*, Vol. I, pp. 597–610.

CAI, J. and WEI, W. (2012) Optimal reinsurance with positively dependent risks. *Insurance: Mathematics and Economics*, **50**(1), 57–63.

CHEUNG, K.C. (2010) Optimal reinsurance revisited—A geometric approach. *ASTIN Bulletin*, **40**(1), 221–239.

CHI, Y. (2012) Reinsurance arrangements minimizing the risk-adjusted value of an insurer's liability. *ASTIN Bulletin*, **42**(2), 529–557.

CHI, Y. and TAN, K.S. (2011) Optimal reinsurance under VaR and CVaR risk measures: A simplified approach. *ASTIN Bulletin*, **41**(2), 487–509.

CHI, Y. and WEI, W. (2018). Optimum insurance contracts with background risk and higher-order risk attitudes. *ASTIN Bulletin*, **48**(3), 1025–1047.

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

DANA, R.A. and SCARSINI, M. (2007) Optimal risk sharing with background risk. *Journal of Economic Theory*, **133**, 152–176.

DHAENE, J., DENUIT, M., GOOVAERTS, M.J., KAAS, R. and VYNCKE, D. (2002) The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**(1), 3–33.

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

EECKHOUDT, L. and KIMBALL, M. (1992) Background risk, prudence, and the demand for insurance. In *Contributions to Insurance Economics*, (ed. G. Dionne), pp. 239–254. Springer.

FRANKE, G., SCHLESINGER, H. and STAPLETON, R.C. (2011) Risk taking with additive and multiplicative background risks. *Journal of Economic Theory*, **146**(4), 1547–1568.

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

GRAF, J.M. and SCHULENBURG, V.D. (1986) Optimal insurance purchasing in the presence of compulsory insurance and uninsurable risks. *The Geneva Papers on Risk and Insurance*, **38**, 5–16.

HONG, S.K., LEW, K.O., MACMINN, R. and BROCKETT, P. (2011) Mossin's theorem given random initial wealth. *The Journal of Risk and Insurance*, **78**(2), 309–324.

HUANG, H.H., SHIU, Y.M. and WANG, C.P. (2013) Optimal insurance contract with stochastic background wealth. *Scandinavian Actuarial Journal*, **2**, 119–139.

HUBERMAN, G., MAYERS, D. and SMITH JR., C.W. (1983) Optimal insurance policy indemnity schedules. *The Bell Journal of Ecoonomics*, **14**(2), 415–426.

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

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

LU, Z.Y., LIU, L.P., ZHANG, J.Y. and MENG, L.L. (2012) Optimal insurance under multiple sources of risk with positive dependence. *Insurance: Mathematics and Economics*, **51**(2), 462–471.
LU, Z.Y., MENG, S.W., LIU, L.P. and HAN, Z. (2018) Optimal insurance design under background risk with dependence. *Insurance: Mathematics and Economics*, **80**, 15–28.
MAHUL, O. (2000) Optimal insurance design with random initial wealth. *Economics Letters*, **69**(3), 353–358.
MARKOWITZ, H. (1952) Portfolio selection. *The Journal of Finance*, **7**(1), 77–91.
MARLIN, P. (1984) Fitting the log-normal distribution to loss data subject to multiple deductibles. *The Journal of Risk and Insurance,* **51**(4), 687–701.
MAYERS, D. and SMITH JR., C.W. (1983) The interdependence of individual portfolio decisions and the demand for insurance. *Journal of Political Economy,* **91**(2), 304–311.
MOSSIN, J. (1968) Aspects of rational insurance purchasing. *Journal of Political Economy*, **76**(4), 553–568.
OHLIN, J. (1969) On a class of measures of dispersion with application to optimal reinsurance. *ASTIN Bulletin*, **5**(2), 249–266.
RAVIV, A. (1979) The design of an optimal insurance policy. *American Economic Review*, **69**(1), 84–96.
VERCAMMEN, J. (2001) Optimal insurance with nonseparable background risk. *The Journal of Risk and Insurance*, **68**(3), 437–448.

YICHUN CHI (Corresponding author)
China Institute for Actuarial Science
Central University of Finance and Economics, Beijing 102206, China
E-Mail: yichun@cufe.edu.cn

KEN SENG TAN
Division of Banking & Finance, Nanyang Business School
Nanyang Technological University, Singapore
E-Mail: kenseng.tan@ntu.edu.sg

APPENDIX A. PROOF OF EQUATION (3.5)

We first consider the following optimisation problem:

$$\min_{0\le f(x)\le x} \mathbb{E}[(X+\psi(X)-f(X))^2]+2\kappa\mathbb{E}[f(X)], \tag{A.1}$$

for a parameter $\kappa\in\mathbb{R}$. This problem will be solved by the state-wise optimisation approach. More specifically, we introduce the following function:

$$L(y)\triangleq (x+\psi(x)-y)^2+2\kappa y,\ 0\le y\le x.$$

Obviously, for each non-negative x, $L(y)$ is convex in y and attains its minimum value at $y=I_\kappa(x)$ where $I_\kappa(x)$ is defined in (3.4). Thus, $I_\kappa(x)$ is a solution to Problem (A.1).

For any admissible insurance contract $I\in\mathfrak{C}$, noting that $\mathbb{E}[I_\kappa(X)]$ is a non-increasing and continuous function of κ with $0=\lim_{\kappa\to\infty}\mathbb{E}[I_\kappa(X)]\le\lim_{\kappa\to-\infty}\mathbb{E}[I_\kappa(X)]=\mathbb{E}[X]$, there

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

must exist a κ^* such that $\mathbb{E}[I_{\kappa^*}(X)] = \mathbb{E}[I(X)]$. Further, recalling that $I_\kappa(x)$ solves Problem (A.1), we have

$$\begin{aligned} &var[X + \psi(X) - I(X)] - var[X + \psi(X) - I_{\kappa^*}(X)] \\ &= \mathbb{E}[(X + \psi(X) - I(X))^2] + 2\kappa^* \mathbb{E}[I(X)] - \left(\mathbb{E}[(X + \psi(X) - I_{\kappa^*}(X))^2] + 2\kappa^* \mathbb{E}[I_{\kappa^*}(X)]\right) \geq 0. \end{aligned}$$

APPENDIX B. PROOF OF THEOREM 3.1

For any admissible ceded loss function $I(x) \in \mathfrak{C}$, it follows from (3.1) and (3.3) that

$$\mathbb{E}[\mathcal{T}_I(X;Y)] = \mathbb{E}[X] + \mathbb{E}[Y] + Q\left(\sum_{i=1}^{m} \mathbb{E}[I(X);\, X \in S_{i,j_i}]\right), \tag{B.1}$$

$$\begin{aligned} var[\mathcal{T}_I(X;Y)] = \mathbb{E}[var[Y|X]] &+ \sum_{i=1}^{m} var[X + \psi(X) - I(X); X \in S_{i,j_i}] \\ &- 2 \sum_{1 \leq i < k \leq m} \mathbb{E}[X + \psi(X) - I(X);\, X \in S_{i,j_i}] \mathbb{E}[X + \psi(X) - I(X);\, X \in S_{k,j_k}], \end{aligned}$$

where $\psi(x)$ and S_{i,j_i} are defined in (3.2) and (3.6), respectively.

Recall that $\mathcal{M}(x, y)$ is a non-decreasing function over $\mathbb{R}_+^2$. It follows from (3.10) and the above two equations that (3.13) holds if we can show that

$$\mathbb{E}\left[\left(X + \psi(X) - \tilde{I}(X)\right)^2;\, X \in S_{i,j_i}\right] \leq \mathbb{E}\left[(X + \psi(X) - I(X))^2;\, X \in S_{i,j_i}\right], \tag{B.2}$$

for $i = 1, \cdots, m$. To establish (B.2), we divide the proof into the following three cases: $j_i = 1, 2, 3$.

(a) We can see from the definition of S_{i,j_i} that $\psi(x)$ is non-decreasing over $S_{i,1}$. As noted in (2.1), $x - I(x)$ is non-decreasing and Lipschitz-continuous. Furthermore, it is easy to show that the function $x - I(x) + \psi(x)$ up-crosses $x - \tilde{I}(x) + \psi(x)$ over $S_{i,1}$, that is

$$\begin{cases} x - I(x) + \psi(x) \leq x - \tilde{I}(x) + \psi(x), & x_{i-1} \leq x \leq z_0; \\ x - I(x) + \psi(x) \geq x - \tilde{I}(x) + \psi(x), & z_0 \leq x \leq x_i, \end{cases}$$

for some $z_0 \in [x_{i-1}, x_i]$. Consequently, Lemma 3 in Ohlin (1969), together with (3.10), implies

$$\left[X - \tilde{I}(X) + \psi(X) \middle| X \in S_{i,1}\right] \leq_{cx} \left[X - I(X) + \psi(X) \middle| X \in S_{i,1}\right],$$

which, in turn, leads to (B.2) as $g(x) = x^2$ is a convex function over $\mathbb{R}$.

(b) Note that $x + \psi(x)$ is non-increasing over $S_{i,3}$. By comparing $I(x)$ with $\tilde{I}(x)$, we can easily find that both $I(x) - (x + \psi(x))$ and $\tilde{I}(x) - (x + \psi(x))$ are non-decreasing functions and that the former up-crosses the latter over $S_{i,3}$. Using (3.10) and Lemma 3 in Ohlin (1969) again, we have

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

$$\left[\tilde{I}(X)-(X+\psi(X))\middle|X\in S_{i,3}\right]\leq_{cx}\left[I(X)-(X+\psi(X))\middle|X\in S_{i,3}\right].$$

This, in turn, implies (B.2)

(c) The proof of (B.2) for $S_{i,2}$ is further separated into two cases: $1\leq i\leq m-1$ and $i=m$.

(i) For $1\leq i\leq m-1$, (B.2) can be easily obtained if we can show that $\tilde{I}(x)$ is a solution to the following optimisation problem:

$$\min_{f\in\mathfrak{C}_I^i}\mathbb{E}\left[(X+\psi(X)-f(X))^2\,;\,X\in S_{i,2}\right],$$

where $\mathfrak{C}_I^i\triangleq\left\{f\in\mathfrak{C}:f(x_k)=I(x_k),k=i-1,i;\,\mathbb{E}[f(X);X\in S_{i,2}]=\mathbb{E}[I(X);\,X\in S_{i,2}]\right\}$. Similar to the proof in Appendix A, it will be proved using the state-wise optimisation approach. More specifically, it is trivial that $I(x)\in\mathfrak{C}_I^i$. Moreover, it follows from (3.9) and (3.10) that $\tilde{I}(x)$ also belongs to the above set. According to (2.1), we must have

$$\begin{aligned}L^i(x)&\triangleq\max\left\{I(x_{i-1}),I(x_i)+x-x_i\right\}\leq f(x)\leq U^i(x)\\&\triangleq\min\left\{I(x_i),I(x_{i-1})+x-x_{i-1}\right\},\ x\in S_{i,2},\end{aligned}$$

for any $f\in\mathfrak{C}_I^i$. Furthermore, noting that $(x+\psi(x)-y)^2+2\lambda_i y$, as a function of y, is convex in y and attains its minimum at $y=x+\psi(x)-\lambda_i$, we thus have

$$\tilde{I}(x)=\arg\min_{L^i(x)\leq y\leq U^i(x)}(x+\psi(x)-y)^2+2\lambda_i y,\ \forall x\in S_{i,2},$$

which leads to

$$\begin{aligned}&\mathbb{E}[(X+\psi(X)-\tilde{I}(X))^2+2\lambda_i\tilde{I}(X);X\in S_{i,2}]\\&\quad\leq\mathbb{E}[(X+\psi(X)-f(X))^2+2\lambda_i f(X);\,X\in S_{i,2}],\end{aligned}$$

for any $f\in\mathfrak{C}_I^i$. Thus, the result follows from the fact that $I(x)\in\mathfrak{C}_I^i$.

(ii) For $i=m$, we similarly define

$$\mathfrak{C}_I^m\triangleq\left\{f\in\mathfrak{C}:f(x_{m-1})=I(x_{m-1}),\ \mathbb{E}[f(X);X\in S_{m,2}]=\mathbb{E}[I(X);\,X\in S_{m,2}]\right\},$$

and then $I(x),\tilde{I}(x)\in\mathfrak{C}_I^m$. Using the arguments similar to that for the case $1\leq i\leq m-1$, we have

$$\tilde{I}(x)=\arg\min_{I(x_{m-1})\leq y\leq I(x_{m-1})+x-x_{m-1}}(x+\psi(x)-y)^2+2(\lambda_m-I(x_{m-1}))y,\quad\forall x\in S_{m,2}.$$

Therefore, $\tilde{I}(x)$ is a solution to the following minimisation problem:

$$\min_{f\in\mathfrak{C}_I^m}\mathbb{E}\left[(X+\psi(X)-f(X))^2\,;\,X\in S_{m,2}\right].$$

Collecting all the above arguments, we obtain (3.13) and the proof is thus complete.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

APPENDIX C. PROOF OF PROPOSITION 4.1

(i) If the background risk Y is SPED with respect to X, then $\psi(x)$ is non-decreasing and, hence, $m=1$. For this case, it follows from Theorem 3.1 that any given insurance policy is suboptimal to a stop-loss insurance contract with the ceded loss function:

$$\tilde{I}_d(x)=(x-d)_+,$$

for some $d\geq 0$.

Furthermore, given a fair premium, that is, $\mathcal{P}(x)=x$ for any $x\geq 0$, it follows from (3.1) and (3.3) that

$$\mathbb{E}[\mathcal{T}_{\tilde{I}_d}(X;Y)]=\mathbb{E}[X]+\mathbb{E}[Y],$$

and

$$var(\mathcal{T}_{\tilde{I}_d}(X;Y))=\mathbb{E}\left[var[Y|X]\right]+var\big(\min\{X,d\}-\mathbb{E}\left[\min\{X,d\}\right]+\psi(X)\big).$$

From Lemma A.2 in Chi (2012), we know that $\min\{X,d\}-\mathbb{E}[\min\{X,d\}]$ is increasing in d in the sense of convex order. Recalling that $\psi(x)$ is non-decreasing for this case, we can get from Corollary 1 in Dhaene *et al.* (2002) that $\min\{X,d\}-\mathbb{E}[\min\{X,d\}]+\psi(X)$ is increasing in d in the convex order. It is well known that variance preserves the convex order; then, full insurance is optimal.

(ii) If $X+Y$ and Y are SPED and SNED with respect to X, respectively, we have $\psi'(x)\in[-1,0]$. For this case, Theorem 3.1 implies that any admissible ceded loss function is dominated by the one in the form of

$$\hat{I}_\lambda(x)=\min\{(x+\psi(x)-\lambda)_+,x\},\quad x\geq 0,$$

for some $-\infty\leq\lambda\leq\infty$. Furthermore, using (3.1) and (3.3), we have

$$\mathbb{E}[\mathcal{T}_{\hat{I}_\lambda}(X;Y)]=\mathbb{E}[X]+\mathbb{E}[Y]+\mathcal{Q}(\mathbb{E}[\hat{I}_\lambda(X)]),$$

which is non-increasing in λ, and

$$var(\mathcal{T}_{\hat{I}_\lambda}(X;Y))=\mathbb{E}[var[Y|X]]+var\big(\varphi_\lambda(X)-\mathbb{E}[\varphi_\lambda(X)]\big),\tag{C.1}$$

where $\varphi_\lambda(x)\triangleq\max\{\psi(x),\min\{\lambda,x+\psi(x)\}\}$ for $x\geq 0$.

For any $\lambda\leq\psi(0)$, noting that $x+\psi(x)$ is non-decreasing over $\mathbb{R}_+$ in this case, we have $\varphi_\lambda(x)=\max\{\psi(x),\lambda\}$. Then,

$$\mathbb{E}[\varphi_\lambda(X)]-\varphi_\lambda(X)=\min\{\psi(0)-\psi(X),\psi(0)-\lambda\}-\mathbb{E}\left[\min\{\psi(0)-\psi(X),\psi(0)-\lambda\}\right],$$

which is decreasing in λ over $(-\infty,\psi(0)]$ in the sense of convex order according to Lemma A.2 in Chi (2012). Recall that the objective function $\mathcal{M}(x,y)$ is a non-decreasing function over $\mathbb{R}_+^2$. Thus, the optimal $\hat{I}_\lambda(x)$ must be attainable at $\lambda\geq\psi(0)$. Furthermore, for $\lambda\geq\psi(0)$ and noting that $\psi'(x)\in[-1,0]$, we have $(x+\psi(x)-\lambda)_+\leq x$, which, in turn, implies $\hat{I}_\lambda(x)=(x+\psi(x)-\lambda)_+$.

Finally, given a fair premium, we have $\mathbb{E}[\mathcal{T}_{\hat{I}_\lambda}(X;Y)]=\mathbb{E}[X]+\mathbb{E}[Y]$. Moreover, for any $\lambda\geq\psi(0)$ and noting for this case that $\psi(x)$ is non-increasing, we have $\varphi_\lambda(x)=\min\{\lambda,x+\psi(x)\}$ for $x\geq 0$. With the use of Lemma A.2 in Chi (2012) and given the fact that variance preserves the convex order, it follows from (C.1) that $var(\mathcal{T}_{\hat{I}_\lambda}(X;Y))$ is non-decreasing in λ. Hence, $\hat{I}_{\psi(0)}(x)$ is an optimal solution under a fair premium.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

(iii) If $X+Y$ is SNED with respect to X, then $\psi'(x) \le -1$. For this case, we have $m=1$, and it follows from Theorem 3.1 that any given admissible ceded loss function $I \in \mathfrak{C}$ is suboptimal to

$$f_K(x) = \min\{x, K\}, \quad x \ge 0,$$

for some $K \ge 0$. Furthermore, we can see from (3.1) that

$$\mathbb{E}[\mathcal{T}_{f_K}(X;Y)] = \mathbb{E}[X] + \mathbb{E}[Y] + Q\left(\mathbb{E}[\min\{X, K\}]\right),$$

which is non-decreasing in K. On the other hand, (3.3) implies

$$var[\mathcal{T}_{f_K}(X;Y)] = \mathbb{E}[var[Y|X]] + var\big(f_K(X) - \mathbb{E}[f_K(X)] - (X+\psi(X))\big).$$

Recalling that $\psi'(x) \le -1$ for this case, we find that $-(x+\psi(x))$ is a non-decreasing function over $\mathbb{R}_+$. Using the proof similar to that for case (i), we find that $f_K(X) - \mathbb{E}[f_K(X)] - (X+\psi(X))$ is increasing in K in the sense of convex order. Since variance preserves the convex order, then $f_K(x)$ is suboptimal to $f_0(x)=0$. As a consequence, no insurance is optimal and this completes the proof.

APPENDIX D. PROOF OF PROPOSITION 4.2

When the partition is $[0,\infty) = S_{1,3} \cup S_{2,2}$, it follows from Theorem 3.1 that any admissible ceded loss function is suboptimal to

$$I(x) = \mathcal{L}_{(0,a]}(x) + \mathcal{L}_{(b,x_1]}(x) + \min\big\{(x+\psi(x)-c)_+, (x-x_1)_+\big\},$$

for some $0 \le a \le b \le x_1$ and $c \in \mathbb{R}$, where x_1 is given in (3.8).

For any $c \le x_1 + \psi(x_1)$, noting that $x+\psi(x)$ is non-increasing over $S_{1,3}$ and increasing over $S_{2,2}$, we have $x+\psi(x)-c \ge 0$ for any $x \ge 0$. This, in turn, leads to

$$L_c(x) \triangleq x+\psi(x)-I(x)-H = \begin{cases} \mathcal{L}_{(a,b]}(x) + \psi(x) - H, & x \le x_1; \\ \max\{\psi(x)-\psi(x_1), c-x_1-\psi(x_1)\}, & \text{otherwise}, \end{cases}$$

where $H \triangleq \psi(x_1) - a + b$. From the above equation, we can see that $L_c(x)$ is a non-increasing function over $\mathbb{R}_+$. Moreover, for any $c_1 < c_2 \le x_1 + \psi(x_1)$, it is easy to show that $\mathbb{E}[L_{c_1}(X)] - L_{c_1}(x)$ up-crosses $\mathbb{E}[L_{c_2}(X)] - L_{c_2}(x)$. Consequently, it follows from Lemma 3 in Ohlin (1969) that

$$\mathbb{E}[L_{c_2}(X)] - L_{c_2}(X) \le_{cx} \mathbb{E}[L_{c_1}(X)] - L_{c_1}(X).$$

Hence, $var[L_c(X)]$ is non-increasing in c and so is $var\big[\mathcal{T}_I(X;Y)\big]$ according to (3.3). On the other hand, we can see from (3.1) that $\mathbb{E}[\mathcal{T}_I(X;Y)]$ is non-increasing in c. Thus, for any $0 \le a \le b \le x_1$, the optimal c must be attainable over $[x_1+\psi(x_1), \infty]$.

Next, for any $c \ge x_1 + \psi(x_1)$ and noting that $\psi'(x) \in (-1, 0)$ for any $x \in S_{2,2}$, we have $(x+\psi(x)-c)_+ \le x - x_1$ for any $x \ge x_1$. Thus, $I(x)$ can be rewritten as $I(x) = K_1(x) + K_2(x)$, where

$$K_1(x) \triangleq \mathcal{L}_{(0,a]}(x) + \mathcal{L}_{(b,x_1]}(x) \quad \text{and} \quad K_2(x) \triangleq (x+\psi(x)-c)_+ \mathcal{I}_{\{x>x_1\}}, \forall x \ge 0,$$

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

and then

$$var(X+\psi(X)-I(X)) = 2cov\,(X+\psi(X)-K_2(X), K_1(x_1)-K_1(X)) + var(X+\psi(X)-K_2(X)) + var(K_1(X)), \tag{D.1}$$

where $cov(Z_1, Z_2)$ is the covariance between random variables Z_1 and Z_2.

Trivially, there must exist $d \in [b-a, b]$ such that

$$\mathbb{E}\left[\tilde{K}_1(X)-\tilde{K}_1(x_1)\right] = \mathbb{E}\left[K_1(X)-K_1(x_1)\right], \tag{D.2}$$

where $\tilde{K}_1(x) \triangleq \mathcal{L}_{(d,x_1]}(x)$ for $x \geq 0$. Moreover, we can see that $K_1(x)-K_1(x_1)$ up-crosses $\tilde{K}_1(x)-\tilde{K}_1(x_1)$. Therefore, Lemma 3 in Ohlin (1969) implies

$$var\big(\tilde{K}_1(X)\big) = var\big(\tilde{K}_1(X)-\tilde{K}_1(x_1)\big) \leq var\big(K_1(X)-K_1(x_1)\big) = var(K_1(X)).$$

Furthermore, noting that $\psi(x)+x$ is non-increasing over $[0, x_1]$ and that $(K_1(x_1)-K_1(x))-\left(\tilde{K}_1(x_1)-\tilde{K}_1(x)\right)$ is non-negative for any $x \in [0, d-(b-a)]$ and non-positive for any $x \geq d-(b-a)$, we have

$$\begin{aligned}
&cov\Big(X+\psi(X)-K_2(X), K_1(x_1)-K_1(X)\Big) - cov\Big(X+\psi(X)-K_2(X),\ \tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\\
&= \mathbb{E}\Big[\Big(X+\psi(X)-K_2(X)\Big)\Big(K_1(x_1)-K_1(X)-\Big(\tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\Big); 0 \leq X \leq x_1\Big]\\
&= \mathbb{E}\Big[(X+\psi(X))\Big(K_1(x_1)-K_1(X)-\Big(\tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\Big); 0 \leq X \leq d-(b-a)\Big]\\
&\quad + \mathbb{E}\Big[(X+\psi(X))\Big(K_1(x_1)-K_1(X)-\Big(\tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\Big); d-(b-a) < X \leq x_1\Big]\\
&\geq (d-b+a+\psi(d-b+a))\mathbb{E}\Big[\Big(K_1(x_1)-K_1(X)-\Big(\tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\Big); 0 \leq X \leq x_1\Big]\\
&= (d-b+a+\psi(d-b+a))\mathbb{E}\Big[\Big(K_1(x_1)-K_1(X)-\Big(\tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\Big)\Big] = 0,
\end{aligned}$$

where the first and last equalities are derived from (D.2).

Collecting the above arguments, we can see from (D.1) that

$$var(X+\psi(X)-I(X)) \geq var(X+\psi(X)-I_{d,c}(X)),$$

where $I_{d,c}(x)$ is defined in (4.3). Furthermore, noting that $\tilde{K}_1(x_1) = x_1 - d \leq K_1(x_1)$, we have $I_{d,c}(x) \leq I(x)$ for any $x \geq 0$, which, together with (3.1), implies $\mathbb{E}[\mathcal{T}_{I_{d,c}}(X;Y)] \leq \mathbb{E}[\mathcal{T}_I(X;Y)]$. Recall that $\mathcal{M}(x, y)$ is a non-decreasing function. Thus, we can conclude from (3.3) that $I(x)$ is suboptimal to $I_{d,c}(x)$.

Finally, we show that $I_{d,c}(x)$ is also suboptimal to $I_{x_1,c}(x)$ when $c \geq x_1 + \psi(x_1)$ satisfies the condition (4.1). More specifically, it is easy to see that $I_{d,c}(x)$ is non-increasing in d and so is $\mathbb{E}[\mathcal{T}_{I_{d,c}}(X;Y)]$ according to (3.1). On the other hand, we have

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

$$
\begin{aligned}
&cov\Big(X+\psi(X)-K_2(X),\tilde{K}_1(x_1)-\tilde{K}_1(X)\Big)\\
&=\mathbb{P}(X\le x_1)\mathbb{E}\left[(X+\psi(X))\left(x_1-d-\mathcal{L}_{(d,x_1]}(X)\right)|X\le x_1\right]\\
&\quad-\mathbb{E}\left[X+\psi(X)-(X+\psi(X)-c)_+\mathcal{I}_{\{X>x_1\}}\right]\mathbb{E}\left[x_1-d-\mathcal{L}_{(d,x_1]}(X)\right]\\
&\ge\mathbb{P}(X\le x_1)\mathbb{E}[X+\psi(X)|X\le x_1]\mathbb{E}\left[x_1-d-\mathcal{L}_{(d,x_1]}(X)|X\le x_1\right]\\
&\quad-\mathbb{E}\left[X+\psi(X)-(X+\psi(X)-c)_+\mathcal{I}_{\{X>x_1\}}\right]\mathbb{E}[x_1-d-\mathcal{L}_{(d,x_1]}(X)]\\
&=\mathbb{P}(X>x_1)\mathbb{E}[x_1-d-\mathcal{L}_{(d,x_1]}(X)]\\
&\quad\times\big(\mathbb{E}[X+\psi(X)|X\le x_1]-\mathbb{E}[\min\{X+\psi(X),c\}|X>x_1]\big)\ge 0,
\end{aligned}
$$

where the first inequality is derived from the fact that both $x+\psi(x)$ and $\tilde{K}_1(x_1)-\tilde{K}_1(x)$ are non-increasing over $[0,x_1]$, and the last inequality follows from Assumption (4.1). Consequently, we can see from (D.1) that

$$
var\big(X+\psi(X)-I_{d,c}(X)\big)\ge var\big(X+\psi(X)-K_2(X)\big)=var\big(X+\psi(X)-I_{x_1,c}(X)\big).
$$

The final result can be obtained from (3.3), and the proof is thus complete.

APPENDIX E. PROOF OF PROPOSITION 4.3

Since it is assumed that $[0,\infty)=S_{1,1}\cup S_{2,2}$, then Theorem 3.1 implies that any admissible ceded loss function is suboptimal to

$$
I(x)=\mathcal{L}_{(a,a+c]}(x)+\min\{(x+\psi(x)-\lambda)_+,(x-x_1)_+\},\tag{E.1}
$$

for some $0\le a\le a+c\le x_1$ and $\lambda\in\mathbb{R}$, where x_1 is the change point of $\psi'(x)$ as defined in (3.8).

If $\lambda\ge x_1+\psi(x_1)$, noting that $x+\psi(x)$ is non-decreasing over $\mathbb{R}_+$ and that $\psi(x)$ is decreasing over $[x_1,\infty)$, we have $(x+\psi(x)-\lambda)_+\le(x-x_1)_+$ for any $x\ge 0$. This, in turn, leads to

$$
I(x)=\mathcal{L}_{(a,a+c]}(x)+(x+\psi(x)-\lambda)_+.
$$

For this case, we have $x+\psi(x)-I(x)=-\mathcal{L}_{(a,a+c]}(x)+\min\{x+\psi(x),\lambda\}$, which is a non-decreasing function as $\psi'(x)\ge 0$ for $0\le x\le x_1$. Furthermore, there must exist $d\in[a,x_1-c]$ such that $\mathbb{E}\left[\mathcal{L}_{(d,x_1]}(X)\right]=\mathbb{E}\left[\mathcal{L}_{(a,a+c]}(X)\right]$, and then we can define

$$
f(x)\triangleq\mathcal{L}_{(d,x_1]}(x)+(x+\psi(x)-\lambda)_+,\quad x\ge 0.
$$

It is easy to show that $x+\psi(x)-I(x)$ up-crosses $x+\psi(x)-f(x)$, and then using Lemma 3 in Ohlin (1969), we can obtain $X+\psi(X)-f(X)\le_{cx}X+\psi(X)-I(X)$. Noting that variance preserves the convex order, we get from (3.1) and (3.3) that $I(x)$ is suboptimal to $f(x)$.

Next, building upon $f(x)$, we can define

$$
h_t(x)\triangleq\mathcal{L}_{(t,x_1]}(x)+(x+\psi(x)-x_1-\psi(x_1))_+,\quad d\le t\le x_1,\tag{E.2}
$$

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

which is a special case of (4.2) and is non-increasing in t. It is easy to see that $h_d(x) \geq f(x)$ for any $x \geq 0$. Furthermore, if $\mathbb{E}[h_{x_1}(X)] \leq \mathbb{E}[f(X)]$, then there must exist $t \in [d, x_1]$ such that $\mathbb{E}[h_t(X)] = \mathbb{E}[f(X)]$. Moreover, it is easy to see that $x + \psi(x) - f(x)$ up-crosses the function $x + \psi(x) - h_t(x)$. Note that these two functions are non-decreasing over $\mathbb{R}_+$. Then, it follows from Lemma 3 in Ohlin (1969) that $X + \psi(X) - h_t(X) \leq_{cx} X + \psi(X) - f(X)$, which implies that $f(x)$ is suboptimal to $h_t(x)$. Otherwise, if $\mathbb{E}[h_{x_1}(X)] > \mathbb{E}[f(X)]$, then there must exist $z > x_1$ such that $\mathbb{E}[h(X)] = \mathbb{E}[f(X)]$, where $h(x) \triangleq (x + \psi(x) - z - \psi(z))_+$, $x \geq 0$. Using similar arguments, we know that $x + \psi(x) - f(x)$ up-crosses the function $x + \psi(x) - h(x)$, and, hence, Lemma 3 in Ohlin (1969) leads to

$$X + \psi(X) - h(X) \leq_{cx} X + \psi(X) - f(X).$$

Thus, we conclude that $f(x)$ is suboptimal to $h(x)$, which is also a special case of (4.2).

Next, if $\lambda < x_1 + \psi(x_1)$, we demonstrate that $I(x)$ is suboptimal to a ceded loss function defined in (4.2). More specifically, if $a + c$ is equal to x_1 or 0, the result easily follows from (E.1). Otherwise, if $0 < a + c < x_1$, we let

$$\gamma_0 \triangleq \inf\{x \geq x_1 : \psi(x) \leq \lambda - x_1\}.$$

It is well defined as $\psi(x)$ is decreasing and continuous over $[x_1, \infty)$. Now we have

$$I(x) = \begin{cases} \mathcal{L}_{(a,a+c]}(x) + (x - x_1)_+, & 0 \leq x \leq \gamma_0; \\ c + x + \psi(x) - \lambda, & \text{otherwise.} \end{cases}$$

Note that $x + \psi(x) - I(x)$ is non-decreasing over $[0, x_1]$, decreasing over $[x_1, \gamma_0]$ and becomes a constant afterwards. For any $l \in [x_1, \gamma_0]$, we define

$$x(l) \triangleq \inf\left\{x \in [0, x_1] : x + \psi(x) - I(x) \geq l + \psi(l) - I(l) = x_1 + \psi(l) - c\right\},$$

which is non-increasing in l, and let $\lambda(l) \triangleq 2x_1 - c - x(l) + I(x(l)) + \psi(l)$. Because $\psi(l) \geq \lambda - x_1$, we have $\lambda(l) \geq x_1 - c - x(l) + I(x(l)) + \lambda \geq x_1 - c - x_1 + I(x_1) + \lambda = \lambda$. On the other hand, we can obtain from the definition of $x(l)$ that $\lambda(l) \leq x_1 + \psi(x(l)) \leq x_1 + \psi(x_1)$. Moreover, we have $\lambda(l) = 2x_1 - c + \psi(l)$ if $x(l) = 0$ and $\lambda(l) = x_1 + \psi(x(l))$ otherwise.

Building upon $x(l)$ and $\lambda(l)$, we construct a ceded loss function:

$$g_l(x) \triangleq \mathcal{L}_{(x(l)-I(x(l)),x_1]}(x) + \min\left\{(x + \psi(x) - \lambda(l))_+, (x - x_1)_+\right\},$$

then $g_l(x_1) = x_1 - x(l) + I(x(l)) \geq I(x_1)$ and $g_l(l) = I(l) = c + l - x_1$. Noting that both $x(l)$ and $\lambda(l)$ are non-increasing in l, we find that $g_l(x)$ is non-decreasing in l. In particular, we have $x(x_1) = x_1$ such that $g_{x_1}(x) \leq I(x)$ for all $x \geq 0$.

If there exists $l_0 \in [x_1, \gamma_0]$ such that $\mathbb{E}[g_{l_0}(X)] = \mathbb{E}[I(X)]$, we can show that $I(x)$ is suboptimal to $g_{l_0}(x)$, which is a special case of (4.2). More specifically, it is easy to see that $g_{l_0}(x) - I(x)$ is non-negative over $[x(l_0), l_0]$ and is non-positive otherwise, and is equal to zero at $x = l_0$ and $x = x(l_0)$. Note that both $x + \psi(x) - I(x)$ and $x + \psi(x) - g_{l_0}(x)$ are non-decreasing over $[0, x_1]$ and are non-increasing for $x \geq x_1$. Thus, we have

$$\begin{aligned} &var\big(X + \psi(X) - I(X)\big) - var\big(X + \psi(X) - g_{l_0}(X)\big) \\ &\quad = \mathbb{E}\left[\big(2(X + \psi(X)) - I(X) - g_{l_0}(X)\big)\big(g_{l_0}(X) - I(X)\big)\right] \end{aligned}$$

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

$$
\begin{aligned}
&= \mathbb{E}\left[(2(X+\psi(X)) - I(X) - g_{l_0}(X))(g_{l_0}(X) - I(X)); X \le x_1\right] \\
&\quad + \mathbb{E}\left[(2(X+\psi(X)) - I(X) - g_{l_0}(X))(g_{l_0}(X) - I(X)); X > x_1\right] \\
&\ge \left(2(x(l_0) + \psi(x(l_0))) - I(x(l_0)) - g_{l_0}(x(l_0))\right) \mathbb{E}[g_{l_0}(X) - I(X); X \le x_1] \\
&\quad + \left(2(l_0 + \psi(l_0)) - I(l_0) - g_{l_0}(l_0)\right) \mathbb{E}[g_{l_0}(X) - I(X); X > x_1] \\
&\ge 2\,(l_0 + \psi(l_0) - I(l_0))\,\mathbb{E}[g_{l_0}(X) - I(X)] = 0,
\end{aligned}
$$

where the last inequality is derived by

$$
x(l_0) + \psi(x(l_0)) - I(x(l_0)) - (l_0 + \psi(l_0)) + I(l_0) \begin{cases} = 0, & x(l_0) > 0; \\ \ge 0, & x(l_0) = 0. \end{cases}
$$

The above results follow from the definition of $x(l)$ and the fact that $g_{l_0}(x) - I(x) \ge 0, \forall 0 \le x \le x_1$ when $x(l_0) = 0$.

Otherwise, if $\mathbb{E}[g_l(X)] < \mathbb{E}[I(X)]$ for any $l \in [x_1, \gamma_0]$, then we must have $a + c < x(l) \le x_1$ for each $x_1 \le l \le \gamma_0$. More specifically, if $x(\tilde{l}) \le a + c$ for some $\tilde{l} \in [x_1, \gamma_0]$, we have $g_l(y) \ge I(y)$, $\forall 0 \le y \le x_1$ for any $l \in [\tilde{l}, \gamma_0]$. Furthermore, if $\gamma_0 < \infty$, it is easy to see that $g_{\gamma_0}(y) \ge I(y)$ for any $y \ge 0$, which, in turn, implies $\mathbb{E}[g_{\gamma_0}(X)] \ge \mathbb{E}[I(X)]$; otherwise, if $\gamma_0 = \infty$ and recalling that $g_l(y)$ is non-decreasing in l, we must have $\lim_{l \uparrow \infty} \mathbb{E}[g_l(X)] > \mathbb{E}[I(X)]$. This contradicts the assumption that $\mathbb{E}[g_l(X)] < \mathbb{E}[I(X)]$ for all $l \in [x_1, \gamma_0]$.

For this case, the result $a + c < x(l) \le x_1$, together with the definition of $x(l)$, suggests that $x(l) + \psi(x(l)) = x_1 + \psi(l)$ and that

$$
\begin{aligned}
&c + x_1 - x(l) + l + \psi(l) - (x_1 + \psi(x_1)) \\
&\quad = c + l - x_1 + \psi(x(l)) - \psi(x_1) \le c + l - x_1 = I(l).
\end{aligned}
$$

Now, we can introduce the following ceded loss function:

$$
\tilde{g}_l(x) \triangleq \mathcal{L}_{(a,a+c]}(x) + \mathcal{L}_{(x(l),x_1]}(x) + \min\left\{(x + \psi(x) - (x_1 + \psi(x(l))))_+, (x - x_1)_+\right\},\ x \ge 0,
$$

such that $\tilde{g}_l(x) = I(x)$ for $x = x(l)$ or l. It is easy to see that $\tilde{g}_l(x)$ is non-decreasing in l and $\tilde{g}_{x_1}(x) \le I(x) \le \tilde{g}_{\gamma_0}(x)$ for any $x \ge 0$. Thus, there must exist $l \in (x_1, \gamma_0)$ such that $\mathbb{E}[\tilde{g}_l(X)] = \mathbb{E}[I(X)]$. Using the arguments similar to the above, we obtain that $I(x)$ is suboptimal to $\tilde{g}_l(x)$ under a general mean-variance preference assumed in this paper.

Furthermore, using the proof similar to that of Theorem 3.1 (see Appendix B), we know that $\tilde{g}_l(x)$ is suboptimal to the following ceded loss function:

$$
I_1(x) = \mathcal{L}_{(a_1,a_1+c_1]}(x) + \min\left\{(x + \psi(x) - \lambda_1)_+, (x - x_1)_+\right\},
$$

where $a_1 > a$, $c_1 = c + x_1 - x(l) > c$ and $\lambda_1 = x_1 + \psi(x(l)) \in [\lambda, x_1 + \psi(x_1)]$. It is easy to find that $I_1(x)$ has the same form as that of $I(x)$ with $\lambda \le x_1 + \psi(x_1)$ in (E.1). Repeating the steps of the above proof, we can define $\gamma_k, x_k(l), g_l^k(x), \tilde{g}_l^k(x)$ and $I_{k+1}(x)$ for $k = 1, 2, \ldots$. If for some n we can find $l \in [x_1, \gamma_n]$ such that $\mathbb{E}[g_l^n(X)] = \mathbb{E}[I_n(X)]$, then similar arguments would lead to the conclusion that $g_l^n(x)$ dominates $I_n(x)$, which has been shown to dominate $I(x)$ in the previous steps. Hence, the result can be obtained by noting that $g_l^n(x)$ is a special case of (4.2) and the construction stops. Otherwise, we can construct a sequence $\{I_n(x)\}_{n \ge 1}$, in

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press

which $I_n(x)$ is suboptimal to $I_{n+1}(x)$, both $\{a_n\}_{n\geq 1}$ and $\{a_n + c_n\}_{n\geq 1}$ are increasing sequences and λ_n is non-decreasing in n. Consequently, $I(x)$ is suboptimal to

$$I^{\infty}(x) \triangleq \lim_{n\to\infty} I_n(x) = \mathcal{L}_{(\lim_{n\to\infty} a_n, \lim_{n\to\infty} a_n + c_n]}(x) + \min\{(x + \psi(x) - \lim_{n\to\infty} \lambda_n)_+, (x - x_1)_+\}.$$

From the construction procedure of $\{I_n(x)\}_{n\geq 1}$, we must have $\lim_{n\to\infty} a_n + c_n = x_1$ or $\lim_{n\to\infty} \lambda_n = x_1 + \psi(x_1)$. For the former case, it is easy to see that $I^{\infty}(x)$ has the same form as the ceded loss function in (4.2). For the latter case, we have shown that $I^{\infty}(x)$ is suboptimal to $h_t(x)$ for some $t \in [0, x_1]$, which is defined in (E.2) and is also a special case of (4.2). The proof is finally completed.

https://doi.org/10.1017/asb.2021.7 Published online by Cambridge University Press