A unifying approach to risk-measure-based optimal reinsurance problems with practical constraints

Ambrose Lo

Department of Statistics and Actuarial Science, The University of Iowa
241 Schaeffer Hall, Iowa City, IA 52242, USA

June 3, 2016

Abstract

The design of optimal reinsurance treaties in the presence of multifarious practical constraints is a substantive but underdeveloped topic in modern risk management. To examine the influence of these constraints on the contract design systematically, this article formulates a generic constrained reinsurance problem where the objective and constraint functions take the form of Lebesgue integrals whose integrands involve the unit-valued derivative of the ceded loss function to be chosen. Such a formulation provides a unifying framework to tackle a wide body of existing and novel distortion-risk-measure-based optimal reinsurance problems with constraints that reflect diverse practical considerations. Prominent examples include insurers' budgetary, regulatory and reinsurers' participation constraints. An elementary and intuitive solution scheme based on an extension of the cost-benefit technique in Cheung and Lo (2015) [Cheung, K.C., Lo, A. (2015). Characterizations of optimal reinsurance treaties: a cost-benefit approach. *Scandinavian Actuarial Journal* (in press). DOI: 10.1080/03461238.2015.1054303.] is proposed and illuminated by analytically identifying the optimal risk-sharing schemes in several concrete optimal reinsurance models of practical interest. Particular emphasis is placed on the economic implications of the above constraints in terms of stimulating or curtailing the demand for reinsurance, and how these constraints serve to reconcile the possibly conflicting objectives of different parties.

Keywords: Budget constraint; Regulatory constraint; Participation constraint; Risk constraint; VaR; TVaR; Distortion; 1-Lipschitz

1 Introduction

Recent decades have witnessed an unprecedented surge in the frequency and severity of catastrophes, which have highlighted the importance of appropriate risk diversification strategies and protective regulatory measures in place. Among the various risk transfer methodologies available in the market, reinsurance remains an extensively used strategy due to the practical role it plays in alleviating the volatility of the risk exposure and in supporting the strategic business planning objectives of an insurer. The fundamental mechanism of reinsurance in a one-period setting is as follows. Let us fix a non-negative unbounded random variable X with a known distribution modeling the nonhedgeable loss faced by an insurer in a certain reference period. As a result of entering into a reinsurance arrangement, the reinsurer pays $f(x)$ to the insurer when x is the realization of X, with the insurer retaining a loss of $x - f(x)$. The function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is known as the *ceded loss function* and defines a reinsurance treaty. In return, the insurer is charged the reinsurance premium $\mu(f(X))$, which is a function of the ceded loss $f(X)$ and is payable to the reinsurer. Therefore, the total risk exposure of the insurer corresponding to a ceded loss function f is changed from the ground-up loss X to the sum of the retained loss and the reinsurance premium: ("T" stands for "total")

$$T_f(X) := X - f(X) + \mu(f(X)). \tag{1.1}$$

A full description of any optimal reinsurance model necessitates the specification of four components: (1) The optimization criterion; (2) The class of feasible ceded loss functions; (3) The reinsurance premium principle; (4) Optimization constraints. With respect to (1), the study of optimal reinsurance has undergone an impressive metamorphosis. Early works such as Borch (1960) and Arrow (1963) predominantly centered on the variance-minimization and expected utility (EU) maximization of the terminal wealth of the insurer, and established the optimality of stop-loss reinsurance. These celebrated results were revisited more recently in Gollier and Schlesinger (1996), Young (1999) and Kaluszka and Okolewski (2008). Nevertheless, it is well documented that some of the basic tenets of the EU theory are systematically violated in practice (see Section 13.1 of Eeckhoudt et al. (2005) for some anomalies). In view of the deficiencies of the EU paradigm to explain human behavior and the prevalence of risk measures in the banking and insurance industries, over the past decade there was a proliferation of research work on risk-measure-based reinsurance models. Initiated by Cai and Tan (2007) in the setting of Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR), investigated more extensively by Cai et al. (2008) and Chi and Tan (2011) among others, and later extended by Cui et al. (2013) to the general framework of Yaari's dual theory of choice (see Yaari (1987) for more information), these optimal reinsurance models explicitly employ risk measures in their objective functions as alternative vehicles to quantify risk. Since then, (distortion-)risk-measure-based optimal reinsurance problems progressively replace their EU counterparts and dominate the optimal reinsurance research arena (see, for example, Zheng and Cui (2014) and Cheung and Lo (2015) for further development).

In spite of the vast literature on optimal reinsurance, the majority of prior studies can be justifiably criticized for being in a vacuum detached from practicalities and externalities. The failure of classical models to take practical considerations into account deprives themselves of economic relevance, whereas the ability to design a robust reinsurance policy that readily applies to reality is what practicing actuaries are urgently in need of. In this connection, some recent studies have started to devote more attention to practical constraints concerned with (1) the financial limitations of the insurer; (2) the specific solvency requirements from regulatory authorities; (3) the risk-bearing capability of the reinsurer. With respect to (1), a budget constraint was incorporated by Cui et al. (2013) in the context of a distortion-risk-measure-based optimal reinsurance model. The presence of such a budget involves the deployment of limited resources to achieve minimality of risk. The same constrained problem was studied more transparently and systematically in Cheung and Lo (2015) from a cost-benefit perspective. Regulatory constraints, designed to promote the stability of the insurance market, were considered in Bernard and Tian (2009), where a VaR-regulatory constraint was translated into the minimization of the insurer's insolvency probability. The standard EU maximization problem was extended in Bernard and Tian (2010) with the inclusion of a VaR-regulatory constraint. As far as (3) is concerned, Cummins and Mahul (2004) were among the first to revisit Arrow's classical optimal reinsurance model by imposing a deterministic upper bound on each allowable ceded loss function. Subsequently, in a series of published papers initiated in Zhou and Wu (2008) and followed up in Zhou and Wu (2009) and Zhou et al. (2010)[1], risk constraints in the form of expected tail loss or VaR were imposed on the reinsurer's net risk exposure. Such reinsurer's risk constraints take into account not only the financial goals of the insurer but also the level of risk tolerance of the reinsurer. In the framework of risk-measure-based optimal reinsurance models, reinsurer's risk constraints were first incorporated in Cheung et al. (2012), where the TVaR of the insurer's total loss was minimized under a VaR-constraint on the reinsurer's net loss.

Given the multifarious types of constraints that prevail in practice, the question arises naturally as to whether a unifying framework can be developed to accommodate a wide spectrum of constrained optimal reinsurance models. Such an approach dispenses with model-specific methods and provides a general recipe for solving a large class of constrained optimal reinsurance problems efficiently. Moreover, the quantitative analysis of external optimization constraints in the realm of general distortion-risk-measure-based reinsurance models remains scarce in the literature but is of interest to both practicing and academic actuaries as risk measures continue to gain popularity. Driven by these needs, the principal objective of this article is to mathematically examine the economic implications of multifaceted constraints for optimal reinsurance via the introduction of a unifying approach. Utilizing Yaari's dual theory of choice as a decision-making vehicle, we embed constraints motivated from practical considerations in a general distortion-risk-measure-theoretic framework and investigate their financial effects on the reinsurance market. The constraints studied in this article include insurers' premium budget constraints, regulatory constraints and reinsurers' risk constraints. To place all of them under the same umbrella, a generic constrained optimization problem with general objective and constraint functions is formulated and solved intuitively via a variation of the cost-benefit approach recently introduced in Cheung and Lo (2015). This problem, which is central to the unifying approach pursued in this paper, exploits the commonality among all the seemingly disparate models that their objective and constraint functions can be exhibited as integrals whose integrands involve the $[0, 1]$-valued derivative of the ceded loss functions to be chosen. Via the prescriptions of appropriate objective and constraint functions into this general model, the explicit solutions to our desired constrained reinsurance problems can be obtained with minimal effort. Our solutions indicate that an insurer's premium budget generally reduces reinsurance coverage, while a regulatory constraint and reinsurer's risk constraint can potentially stimulate an insurer's

[1]To be precise, these papers studied optimal insurance, which is technically identical to optimal reinsurance. In optimal insurance problems, the insured and insurer play the same roles as the insurer and reinsurer in optimal reinsurance problems.

demand for reinsurance.

It is imperative to emphasize that, although the study of the effects of regulatory constraints on optimal reinsurance is not an entirely new topic, the investigation in this article differs from the literature in two conceptually and practically substantive aspects. First and foremost, note that regulatory requirements only constrain the reinsurance strategies that an insurer can select. As such, they do *not* define risk management and strategic priorities. Thus the adoption of the regulatory requirement as the optimization criterion, as in Bernard and Tian (2009), does not necessarily align with the strategic goals of an insurance company. It appears to be more natural and meaningful, as this article entails, for an insurer to optimize a risk functional which best dovetails with its risk profile subject to prescribed regulatory constraints. Second, the issue of moral hazard is not appropriately handled in many existing studies, resulting in counter-intuitive optimal solutions that are not marketable in practice. Throughout this paper, we restrict our analysis to non-decreasing and 1-Lipschitz ceded loss functions, so that both the insurer and reinsurer incur a higher loss with a heavier ground-up loss, ruling out ex post moral hazard issues arising from the manipulation of losses (see Cheung et al. (2014) for how non-decreasing and 1-Lipschitz indemnity schedules appeal to both the insurer and reinsurer in an EU setting).

This article proceeds as follows. Section 2 lays the mathematical groundwork of distortion-risk-measure-based optimal reinsurance models, and introduces three constrained models of practical interest. In Section 3, a generic constrained optimal reinsurance problem is formulated. The solution scheme, along with the underlying heuristic considerations, is presented, followed by its successive specializations in Section 4 to the three specific reinsurance models. The economic consequences of the constraints on the demand for reinsurance are highlighted. Finally, Section 5 concludes the paper.

2 Model formulation

2.1 Basic setting

Following Cheung and Lo (2015), we start by describing a distortion-risk-measure-based optimal reinsurance model with regard to the first three components sketched in the introductory section: (1) The optimization criterion; (2) The class of feasible ceded loss functions; (3) The reinsurance premium principle. The fourth component, namely optimization constraints, will be specified in the next subsection. In the current model, risk is quantified by *distortion risk measures*, which in turn are defined by means of distortion functions. By definition, a *distortion function* $g\colon [0,1] \to [0,1]$ is a non-decreasing function such that $g(0) = 0$ and $g(1) = 1$. Corresponding to a distortion function g, the *distortion risk measure* of a non-negative random variable Y is defined by

$$\rho_g(Y) := \int_0^\infty g(S_Y(t))\,\mathrm{d}t, \tag{2.1}$$

where $S_Y(t) := \mathbb{P}(Y > t)$ is the survival function of Y. Throughout this article, all random variables are tacitly assumed to be sufficiently integrable in the sense that their distortion risk measures are well-defined and finite.

In addition to having far-reaching implications for decision-making under uncertainty, a significant advantage of using distortion risk measures is their ability to incorporate a wide class of risk measures. In the later part of this article, we analyze specifically Value-at-Risk and Tail Value-at-Risk, which are two prominent examples of distortion risk measures. Their definitions are recalled below. In the remainder of this article, we denote the indicator function of a given set A by 1_A, i.e. $1_A(x) = 1$ if $x \in A$, and $1_A(x) = 0$ otherwise, and the empty set by $\emptyset$.

Definition 2.1. *(Definitions of VaR and TVaR)* Let Y be a random variable with distribution function F_Y and $\alpha \in (0, 1)$ be fixed.

1. The *Value-at-Risk* (VaR) of Y at the probability level of α is defined by

$$\mathrm{VaR}_\alpha(Y) = F_Y^{-1}(\alpha) := \inf\{y \in \mathbb{R} \mid F_Y(y) \geq \alpha\}$$

with the convention $\inf \emptyset = \infty$. It is known that (see Equation (4) of Dhaene et al. (2006))

$$F_Y^{-1}(\alpha) \leq y \quad \Leftrightarrow \quad \alpha \leq F_Y(y) \quad \text{for all } \alpha \in (0, 1). \tag{2.2}$$

The distortion function that gives rise to VaR is $g(x) = 1_{\{x>1-\alpha\}}$ (see Equation (44) of Dhaene et al. (2006)).

2. The *Tail Value-at-Risk* (TVaR) of Y at the probability level of α is defined by

$$\mathrm{TVaR}_\alpha(Y) := \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_p(Y)\,\mathrm{d}p.$$

As Equation (45) of Dhaene et al. (2006) indicates, the distortion function corresponding to TVaR is $g(x) = \frac{x}{1-\alpha} \wedge 1$.

To mitigate its risk exposure to the ground-up loss X, the insurer can purchase a reinsurance scheme f from a reinsurer so that its resulting total risk exposure is changed from X to $T_f(X)$ as defined in Equation (1.1). In this article, the feasible class of reinsurance policies is restricted to the set $\mathcal{F}$ of non-decreasing and 1-Lipschitz functions dominated by the identity function, i.e.

$$\mathcal{F} = \left\{ f\colon \mathbb{R}^+ \to \mathbb{R}^+ \,\middle|\, \begin{array}{l} 0 \leq f(x) \leq x \text{ for all } x \geq 0, \\ 0 \leq f(x_1) - f(x_2) \leq x_1 - x_2 \text{ for } 0 \leq x_2 \leq x_1 \end{array} \right\}.$$

For any reinsurance treaty selected from the set $\mathcal{F}$, the ceded loss cannot exceed the ground-up loss, and both the insurer and reinsurer will suffer as the ground-up loss becomes more severe, thereby having no incentive to misrepresent claims. Besides ruling out the issue of moral hazard, the 1-Lipschitzity condition results in a useful technical by-product: every $f \in \mathcal{F}$ is absolutely continuous with a derivative f' which exists almost everywhere and is bounded between 0 and 1.

Premium-wise, we assume that the reinsurance premium is calibrated, for a given ceded loss function $f \in \mathcal{F}$, by the formula

$$\mu_r\left(f(X)\right) := \int_0^\infty r\left(S_{f(X)}(t)\right)\mathrm{d}t, \tag{2.3}$$

where $r\colon [0,1] \to \mathbb{R}^+$ is a non-decreasing function satisfying $r(0) = 0$.

The following technical lemma, which can be found in Lemma 2.1 of Cheung and Lo (2015), makes the optimal selection of f more transparent and will be intensively used in the sequel.

Lemma 2.2. (Integral representations of objective and premium functions) *For any distortion function g and ceded loss function f in $\mathcal{F}$, we have*

$$\rho_g\left(T_f(X)\right) = \rho_g(X) + \int_0^\infty \left[r(S_X(t)) - g(S_X(t))\right] f'(t)\,\mathrm{d}t$$

and

$$\mu_r\left(f(X)\right) = \int_0^\infty r(S_X(t)) f'(t)\,\mathrm{d}t.$$

2.2 Three motivating models

Armed with the mathematical underpinnings in Subsection 2.1, in this subsection we set up three concrete constrained optimal reinsurance models which form the main motivation of this article. Their economic interpretations are elucidated, with particular attention devoted to the practical significance of the additional constraints. These models appear disparate in terms of the specification of their constraint functions, as well as their underlying considerations. In lieu of solving each of these three problems separately by *ad hoc* means, their commonalities will be explored and exploited in Sections 3 and 4, leading naturally to a unifying solution scheme that applies to all three models.

- *Model 1: Budget-constrained risk minimization*

$$\begin{cases} \inf\limits_{f\in\mathcal{F}} & \rho_{g_I}\left(T_f(X)\right) \\ \text{s.t.} & \mu_r\left(f(X)\right) \le \pi, \end{cases} \tag{2.4}$$

  where g_I is the distortion risk measure adopted by the insurer, and π is an exogenously assigned strictly positive quantity which can be interpreted as the budget allocated to reinsurance activities. In this model, the presence of the budget constraint, which is often in place in reality, poses a strategic challenge to the insurer, which must decide how to make the most of the limited budget to achieve the greatest reduction in the total retained risk. This problem has its genesis in Cui et al. (2013), where some convoluted partition arguments were presented under extra technical assumptions on g_I and r. Subsequently, it was solved in complete generality and more transparently in Cheung and Lo (2015) using a cost-benefit approach.

- *Model 2: Risk minimization subject to a regulatory constraint*

$$\begin{cases} \inf\limits_{f\in\mathcal{F}} & \rho_{g_I}(T_f(X)) \\ \text{s.t.} & \rho_{g_r}(T_f(X)) \le \pi, \end{cases} \tag{2.5}$$

where g_r is another distortion risk measure designated by regulatory authorities. In this model, the insurer is minimizing its total retained risk quantified by its favored distortion risk measure g_I, while complying with the regulatory constraint $\rho_{g_r}(T_f(X)) \leq \pi$. In general, the two distortion risk measures g_I and g_r are different, with the latter usually being more conservative than the former in terms of risk measurement, i.e. $g_I \leq g_r$. We are interested in the possible effects the regulatory constraint exerts on the form of the optimal reinsurance treaty and, in particular, whether the insurer under regulatory scrutiny is encouraged to implement a more prudent reinsurance strategy as regulators would very much hope. As far as the author is aware, this article is the first of its kind to examine the economic ramifications of regulations in the context of a (distortion-)risk-measure-based optimal reinsurance model without moral hazard.

While the unifying solution scheme in this article allows us to solve Problem (2.5) for general g_I and g_r, the implications of the regulatory constraint can be demonstrated more concretely when specific distortion risk measures are prescribed. To this end, we suppose specifically that the insurer is a VaR-adopter which is TVaR-regulated:

$$\begin{cases} \inf\limits_{f \in \mathcal{F}} & \mathrm{VaR}_\alpha(T_f(X)) \\ \text{s.t.} & \mathrm{TVaR}_\beta(T_f(X)) \leq \pi, \end{cases} \tag{2.6}$$

where α and β are given probability levels in $(0, 1)$ chosen by the insurer and regulator respectively. The two risk measures are prescribed due to the agreement of their characteristics with the interests of the concerned parties. It is well-known that the unconstrained version of Problem (2.6) is solved by a limited stop-loss reinsurance treaty which has no protection on extreme losses (see, for example, Chi and Tan (2011), Cui et al. (2013), Cheung and Lo (2015)). It would be intriguing to explore whether the enforcement of the TVaR-regulatory constraint would stimulate the demand for reinsurance and compel the insurer to protect itself against even extreme losses.

- *Model 3: Risk minimization in the presence of a reinsurer's risk constraint*

$$\begin{cases} \inf\limits_{f \in \mathcal{F}} & \rho_{g_I}\left(T_f(X)\right) \\ \text{s.t.} & \rho_{g_R}\left(f(X) - \mu_r\left(f(X)\right)\right) \leq \pi, \end{cases} \tag{2.7}$$

where g_R is the distortion function selected by the reinsurer to quantify its net risk exposure $f(X) - \mu_r\left(f(X)\right)$. This model recognizes the two-party nature of reinsurance, which suggests that a reinsurance treaty designed solely from the perspective of one party while completely neglecting the interests of the other may fail to be mutually acceptable and practically realistic. The economic significance of the model lies in the selection of a reinsurance arrangement which is optimal to the insurer and simultaneously acceptable to the reinsurer, whose exposure is below the level of π.

Parallel to Problem (2.6), we consider the following specific problem as a useful special case of Problem (2.7):

$$\begin{cases} \inf\limits_{f \in \mathcal{F}} & \mathrm{TVaR}_\alpha(T_f(X)) \\ \text{s.t.} & \mathrm{VaR}_\beta\left(f(X) - \mu_r\left(f(X)\right)\right) \leq \pi, \end{cases} \tag{2.8}$$

where α and β are fixed probability levels in $(0,1)$ chosen respectively by the insurer and reinsurer. Our aim is to examine quantitatively whether the imposition of the reinsurer's risk constraint creates additional incentives or disincentives for the insurer to employ reinsurance to manage its loss.

In the traditional EU framework, the study of the insurer's EU-maximization problem subject to a reinsurer's risk constraint was pioneered in Cummins and Mahul (2004) and later extended in Zhou and Wu (2008, 2009) and Zhou et al. (2010). It was shown that the optimal reinsurance treaty changes from a stop-loss insurance to a limited stop-loss insurance with a reduced demand for reinsurance. With Problems (2.7) and (2.8) considered as a distortion-risk-measure version of the problem studied in these papers, it will be of both theoretical and practical interest to investigate whether and how the reinsurer's risk constraint alters the optimal reinsurance treaty.

Among the three constrained optimal reinsurance problems, (2.4), (2.6) and (2.8), it is worth restating that Problem (2.6) is entirely novel and that Problem (2.8) was solved only implicitly in Cheung et al. (2012) under the expectation premium principle without profit loading, and without providing an explicit construction of the optimal reinsurance policy. The linchpin of this article is an innovative and versatile solution scheme that can be readily specialized not only to retrieve the solutions of Problem (2.4) as determined in Cheung and Lo (2015), but also to tackle Problems (2.6) and (2.8) in full and expeditiously.

3 A generic constrained optimal reinsurance model

To set forth a unifying solution to all of the three constrained optimal reinsurance problems introduced in Subsection 2.2, it is instructive to observe, by virtue of Lemma 2.2, that each of them can be represented in the general form of

$$\begin{cases} \inf_{f\in\mathcal{F}} & \int_0^\infty G(S_X(t))f'(t)\,\mathrm{d}t \\ \text{s.t.} & \int_0^\infty H(S_X(t))f'(t)\,\mathrm{d}t \le \pi, \end{cases} \tag{3.1}$$

where G and H are some functions defined on $[0,1]$, and π is a fixed real quantity. In this technical section, we are prompted to solve the generic constrained optimization problem (3.1), which is the abstraction of the three concrete problems in Subsection 2.2. Apart from being interesting in its own right, such a problem formulation is fruitful in the sense that the solutions to the three desired optimal reinsurance problems can be obtained readily from the general solutions of Problem (3.1) by prescribing the appropriate functions G and H.

3.1 Heuristic considerations

As a precursor to the mathematical and rigorous solutions of Problem (3.1), in this subsection the heuristics that demystifies the derivations of the optimal solutions is outlined. The considerations involve a non-trivial extension of the cost-benefit arguments recently developed in Cheung and Lo (2015). The formal solutions are obtained by formalizing and

	Set A	Set B	Set C	Set D
Definition	$G(S_X(\cdot)) \leq 0$ $H(S_X(\cdot)) \leq 0$	$G(S_X(\cdot)) \leq 0$ $H(S_X(\cdot)) > 0$	$G(S_X(\cdot)) > 0$ $H(S_X(\cdot)) \leq 0$	$G(S_X(\cdot)) > 0$ $H(S_X(\cdot)) > 0$
Sign of $R_X(\cdot)$	Positive	Negative	Negative	Positive
Optimal Strategy	Always cede all losses in this "good set"	If possible, first cede losses on which R_X takes the most negative values	If needed, first cede losses on which R_X takes the least negative values	Never cede any loss in this "bad set"

Table 1: Heuristic considerations underlying the solution of Problem (3.1).

transcribing these heuristic considerations into precise mathematical statements. As we shall see, the *signs* and *magnitudes* of the functions $G(S_X(\cdot))$ and $H(S_X(\cdot))$ are useful indications of where reinsurance coverage is most "effective" and "efficient", in a sense to be described.

To begin with, we partition the non-negative real line $\mathbb{R}^+$ into four sets in accordance with the signs of $G(S_X(\cdot))$ and $H(S_X(\cdot))$, as depicted in Table 1. The contributions of these four sets to the objective and constraint functions of Problem (3.1) are expounded as follows:

Set A. $G(S_X(\cdot)) \leq 0$ *and* $H(S_X(\cdot)) \leq 0$*:* Ceding each unit of excess loss in this set simultaneously reduces the objective function, in line with the goal of the minimization problem, as well as the constraint function, providing additional room for reinsurance coverage on losses in other sets contributing to further reduction in the objective function. In an attempt to minimize the objective function in the presence of the integral constraint on $H(S_X(\cdot))$, it is always advisable to purchase full protection on losses in this "good set", or mathematically, to set $f'(t)$ to its maximum value of 1. In other words, optimally $f'(t) = 1$ whenever $G(S_X(t)) \leq 0$ and $H(S_X(t)) \leq 0$.

Set B. $G(S_X(\cdot)) \leq 0$ *and* $H(S_X(\cdot)) > 0$*:* Purchasing reinsurance coverage on this set decreases the objective function, as in Set A, at the expense of a rise in the value of the constraint function. We distinguish two cases:

- If ceding all the excess losses in Sets A and B fulfills the integral constraint, then the optimal reinsurance treaty is designed as full reinsurance coverage on the union of these two sets, or the set where the function $G(S_X(\cdot))$ is negative. No protection should be sought on any other sets of losses, for it only serves to raise the value of the objective function. In this case, the integral constraint is not binding.
- If full reinsurance coverage on Sets A and B together violates the integral constraint, then optimality is achieved by arranging $f'(t)$ to be 1 on Set A, as always, and, if coverage on Set B is allowed, the most "efficient" parts of Set B resulting in the greatest reduction in $G(S_X(\cdot))$ with the least positive $H(S_X(\cdot))$. This sense of efficiency can be measured by the ratio of the integrands of the objective and

constraint functions given by (with "R" signifying "ratio")

$$R_X(t) = \frac{G(S_X(t))}{H(S_X(t))},$$

whose *most negative* values (i.e. negative with the largest magnitude) on Set B emanating from the most negative values of $G(S_X(\cdot))$ in conjunction with the least positive values of $H(S_X(\cdot))$ are desired. Using the least level sets of R_X as the guide, we purchase full coverage and set $f'(t)$ to be one on

$$\begin{aligned} B_1^* &:= \{G(S_X(\cdot)) \leq 0, H(S_X(\cdot)) > 0, R_X(\cdot) \leq c_1^*\} \\ &= \{G(S_X(\cdot)) \leq 0, R_X(\cdot) \leq c_1^*\} \\ &= \{H(S_X(\cdot)) > 0, R_X(\cdot) \leq c_1^*\}, \end{aligned}$$

which is a subset of Set B, and c_1^* is selected so that the constraint function equals the upper bound π.

Set C. $G(S_X(\cdot)) > 0$ *and* $H(S_X(\cdot)) \leq 0$*:* Ceding the excess losses in this set undesirably contributes to an increase in the objective function, a departure from the aim of the constrained minimization problem, but a decline in the constraint function, a step towards making a reinsurance treaty feasible. In the event that the integral constraint is refuted by full reinsurance coverage on Set A because of an overly negative π, one is then forced to buy reinsurance on Set C as well to satisfy the integral constraint. To minimize the resulting rise in the objective function, we are prompted to cede on the subset of Set C corresponding to the smallest positive $G(S_X(\cdot))$ along with the most negative $H(S_X(\cdot))$, which together give rise to the *least negative* values (i.e. negative with the smallest magnitude) of $R_X(\cdot)$. Denote by c_2^* the cutoff level such that buying a full coverage on

$$\begin{aligned} C_2^* &:= \{G(S_X(\cdot)) > 0, H(S_X(\cdot)) \leq 0, c_2^* \leq R_X(\cdot) < 0\} \\ &= \{G(S_X(\cdot)) > 0, c_2^* \leq R_X(\cdot) < 0\} \\ &= \{H(S_X(\cdot)) \leq 0, c_2^* \leq R_X(\cdot) < 0\}, \end{aligned}$$

which is a subset of Set C, together with Set A, exactly binds the integral constraint.

To complete the design of the optimal reinsurance scheme, let us compare what the functional values of $R_X(\cdot)$ on Sets B and C indicate. By definition, the magnitude of $R_X(\cdot)$ measures the decrease in $G(S_X(\cdot))$ per unit increase in $H(S_X(\cdot))$ on Set B, and the increase in $G(S_X(\cdot))$ per unit decrease in $H(S_X(\cdot))$ on Set C.

- If c_2^* is greater than the infimum of $R_X(\cdot)$ on Set B, then each unit increase in $H(S_X(\cdot))$ on the most "efficient" part of Set B generates a reduction in $G(S_X(\cdot))$ which outweighs the increase in $G(S_X(\cdot))$ associated with a further unit decrease in $H(S_X(\cdot))$ on Set C, leading to an overall decrease in the objective function. In this case, reinsurance coverage on the subset of Set B with the most negative values of $R_X(\cdot)$ is effective, and optimality will be attained when the cutoff levels on Sets B and C are equal to a common value, say c^*, reaching somehow an "equilibrium". The optimal reinsurance policy is defined by full coverage on Sets A, B^* and C^*, where

$$B^* = \{H(S_X(\cdot)) > 0, R_X(\cdot) < c^*\} \quad \text{and} \quad C^* = \{H(S_X(\cdot)) \leq 0, c^* \leq R_X(\cdot) < 0\}$$

are subsets of Sets B and C respectively.

- If c_2^* is less than the infimum of $R_X(\cdot)$ on Set B, then ceding any losses in Set B will only backfire, and the optimal ceded loss function is constructed by full coverage on Sets A and C_2^*.

Set D. $G(S_X(\cdot)) > 0$ *and* $H(S_X(\cdot)) > 0$*:* Reinsurance coverage in this set thanklessly increases the objective function as well as the constraint function. This is a "bad set" which should be completely discarded in the design of the optimal reinsurance treaty.

Now that the distinct roles played by the four sets in devising the optimal ceded loss function are unraveled, Problem (3.1) can be heuristically solved by the following procedure:

Case 1. If ceding all the excess losses in Sets A and B fulfills the integral constraint, then the optimal solution is full coverage on $A \cup B$, or the set $\{G(S_X(\cdot)) \leq 0\}$.

Case 2. If hedging against the excess losses in Set A fulfills the constraint, but not in Sets A and B together, then the optimal solution is full coverage on Set A and the losses in Set B with the most negative values of R_X.

Case 3. If ceding all the excess losses in Sets A and C (i.e. $\{H(S_X(\cdot)) \leq 0\}$) satisfies the constraint, but not in Set A alone, then the optimal solution is full coverage on Set A and losses in Set C with the least negative values of R_X, possibly together with losses in Set B that correspond to the most negative values of R_X.

Case 4. If protecting itself against all the excess losses in Sets A and C violates the constraint, then the insurer can never satisfy the integral constraint, and Problem (3.1) has no solution.

3.2 Formal solutions

Drawing upon the heuristics in Subsection 3.1, we are now in a position to formulate the rigorous solutions to Problem (3.1) by distinguishing the range of values of the upper bound π and various partial integrals of the function $H(S_X(\cdot))$, and defining appropriate subsets of the four sets, $A, B, C,$ and D, in Table 1. The technical proof is relegated to the Appendix.

Theorem 3.1. (Solutions of Problem (3.1)) *Consider Problem* (3.1)*, and define* $R_X : \mathbb{R}^+ \to \bar{\mathbb{R}}$ *by*

$$R_X(t) = \begin{cases} \dfrac{G(S_X(t))}{H(S_X(t))}, & \textit{if } H(S_X(t)) \neq 0, \\ -\infty, & \textit{if } H(S_X(t)) = 0 \textit{ and } G(S_X(t)) < 0, \\ +\infty, & \textit{if } H(S_X(t)) = 0 \textit{ and } G(S_X(t)) > 0, \\ \textit{undefined}, & \textit{if } H(S_X(t)) = G(S_X(t)) = 0. \end{cases}$$

Case 1. *Suppose that* $\int_{A\cup B} H(S_X(t))\,\mathrm{d}t \le \pi$. *Then an optimal solution of Problem* (3.1) *is*

$$f^*(x) = \int_0^x 1_{A\cup B}(t)\,\mathrm{d}t = \int_0^x 1_{\{G(S_X(t))\le 0\}}\,\mathrm{d}t.$$

Case 2. *Suppose that* $\int_A H(S_X(t))\,\mathrm{d}t \le \pi < \int_{A\cup B} H(S_X(t))\,\mathrm{d}t$. *Define*

$$c_1^* = \sup\left\{c \le 0 \,\middle|\, \int_{\{H(S_X(\cdot))>0, R_X(\cdot)\le c\}} H(S_X(t))\,\mathrm{d}t \le \pi - \int_A H(S_X(t))\,\mathrm{d}t\right\}$$

and

$$B_1^* = \{H(S_X(\cdot)) > 0, R_X(\cdot) \le c_1^*\}.$$

Then an[2] *optimal solution of Problem* (3.1) *is*

$$f^*(x) = \int_0^x 1_{A\cup B_1^*}(t)\,\mathrm{d}t.$$

Case 3. *Suppose that* $\int_{A\cup C} H(S_X(t))\,\mathrm{d}t \le \pi < \int_A H(S_X(t))\,\mathrm{d}t$. *Define*

$$c_2^* = \sup\left\{c \le 0 \,\middle|\, \int_{\{H(S_X(\cdot))\le 0, R_X(\cdot)\ge c\}} H(S_X(t))\,\mathrm{d}t \le \pi\right\}.$$

Then an optimal solution of Problem (3.1) *is*

$$f^*(x) = \int_0^x 1_{A\cup B^*\cup C^*}(t)\,\mathrm{d}t,$$

where

$$B^* = \{H(S_X(\cdot)) > 0, R_X(\cdot) < c^*\} \text{ and } C^* = \{H(S_X(\cdot)) \le 0, c^* \le R_X(\cdot) < 0\},$$

and $c^* \in \left[\min\left(\inf_{t\in B} R_X(t), c_2^*\right), c_2^*\right]$ *satisfies*

$$\int_{A\cup B^*\cup C^*} H(S_X(t))\,\mathrm{d}t = \pi.$$

Case 4. *Suppose that* $\int_{A\cup C} H(S_X(t))\,\mathrm{d}t > \pi$. *Then Problem* (3.1) *admits no solution.*

[2] To minimize technicalities and simplify the presentation of Cases 2 and 3 of Theorem 3.1, we implicitly assume that the selection of c_1^* and c^* binding the integral constraint is always possible and unique. A sufficient condition for this is that $R_X(\cdot)$ is not locally constant, i.e. there does not exist any non-empty interval $[a, b]$ on which R_X takes the same value. See the statements of Theorems 3.1 (c), (d) and 3.3 (b) of Cheung and Lo (2015) about how the general case can be treated and how the whole set of optimal solutions can be determined.

4 Solutions of the three constrained optimal reinsurance problems

In this section, we prescribe explicit forms of the functions G and H in Problem (3.1), identify the four sets introduced in Table 1, and specialize Theorem 3.1 to solve the three concrete constrained optimal reinsurance problems presented in Subsection 2.2. Each of these models admits distinctive economic interpretations and possesses subtly unique features. In addition to explicitly deriving the optimal reinsurance treaties, we lay our emphasis on the interpretations of the economic implications of the additional constraints for the demand for reinsurance and the characteristics of the chosen risk measures.

For convenience, we write, for any real x and y, $x \wedge y = \min(x, y)$, $x \vee y = \max(x, y)$ and $x_+ = x \vee 0$.

4.1 Model 1: Budget-constrained risk minimization

Consider the budget-constrained risk minimization problem (2.4), which was first studied in Cui et al. (2013) under some technical assumptions on the functions g_I and r using rather ad-hoc partitioning methods. A considerably simpler approach that exploits the cost-benefit structure of optimal reinsurance problems was given in Cheung and Lo (2015). As a confirmation of the consistency of our novel unifying solution scheme, we demonstrate that the use of Theorem 3.1 readily retrieves the known results of Cheung and Lo (2015), a streamlined version of which is presented below.

Corollary 4.1. (Solutions of budget-constrained risk minimization problem (2.4)) *Define* $H_X \colon \mathbb{R}^+ \to \bar{\mathbb{R}}^+$ *by*

$$H_X(t) := \frac{g_I(S_X(t))}{r(S_X(t))} \tag{4.1}$$

and $c^* = \inf\left\{c \geq 1 \,\middle|\, \int_{\{H_X \geq c\}} r(S_X(t))\,\mathrm{d}t \leq \pi\right\}$. *Then an optimal solution of Problem* (2.4) *is*

$$f^*(x) = \int_0^x 1_{\{H_X(t) \geq c^*\}}\,\mathrm{d}t.$$

Proof. In the setting of Problem (2.4), the budget π is a priori assumed to be non-negative, and the function $H = r$ is a non-negative function, so Sets A and C defined in Table 1 are empty, and only Cases 1 and 2 of Theorem 3.1 prevail. Since

$$G(S_X(t)) \leq 0 \quad \Leftrightarrow \quad r(S_X(t)) \leq g_I(S_X(t)) \quad \Leftrightarrow \quad H_X(t) \geq 1$$

and

$$R_X(t) = \frac{r(S_X(t)) - g_I(S_X(t))}{r(S_X(t))} = 1 - H_X(t),$$

we have $\{R_X(\cdot) \leq c_1^*\} = \{H_X(\cdot) \geq 1 - c_1^*\}$, and an optimal solution of Problem (2.4) is

$$f^*(x) = \int_0^x 1_{\{H_X(t) \geq 1, H_X(t) \geq 1 - c_1^*\}}\,\mathrm{d}t = \int_0^x 1_{\{H_X(t) \geq 1 - c_1^*\}}\,\mathrm{d}t,$$

where

$$\begin{aligned} c_1^* &= \sup\left\{c \leq 0 \,\middle|\, \int_{\{G(S_X(\cdot))\leq 0, R_X(\cdot)\leq c\}} r(S_X(t))\,\mathrm{d}t \leq \pi\right\} \\ &= 1 - \inf\left\{c \geq 1 \,\middle|\, \int_{\{H_X(\cdot)\geq c\}} r(S_X(t))\,\mathrm{d}t \leq \pi\right\}. \end{aligned}$$

The result follows by relabeling $1 - c_1^*$ as $c^* = \inf\left\{c \geq 1 \,\middle|\, \int_{\{H_X \geq c\}} r(S_X(t))\,\mathrm{d}t \leq \pi\right\}$. □

The ratio H_X defined in Equation (4.1) admits a very intuitive cost-benefit interpretation. Observe that each unit of excess loss the insurer cedes leads to a marginal reduction of $g_I(S_X(\cdot))$ in the retained risk (*benefit*), but a marginal increase of $r(S_X(\cdot))$ in the reinsurance premium (*cost*). Corollary 4.1 suggests that the insurer, with a view to minimizing its retained risk, should purchase full coverage on the excess layers of loss on which the benefit-to-cost ratio H_X takes the greatest values, until the whole budget is consumed. These cost-benefit considerations, however, may not easily carry over to the other two constrained optimal reinsurance problems.

4.2 Model 2: Risk minimization in the presence of a regulatory constraint

The utility of Theorem 3.1 fully manifests itself when it is applied to solve the insurer's VaR minimization problem subject to a TVaR-regulatory constraint as given in Problem (2.6), in which the integrand of the constraint function $H(S_X(t)) = r(S_X(t)) - [S_X(t)/(1-\beta)] \wedge 1$ can take positive and negative values. This technical difficulty prohibits the direct application of the cost-benefit argument in Cheung and Lo (2015) and distinguishes Problem (2.6) from Problem (2.4). In this subsection, we strive to derive the explicit solutions of Problem (2.6) when the reinsurance premium is calibrated by the well-known expectation premium principle with a profit loading of θ, i.e., $r(x) = (1+\theta)x$, where $\theta \geq 0$. Adopted in this and the next subsection, this premium principle is mathematically tractable and extensively used in the literature, thereby facilitating the comparison of our novel results with those in existing studies. The explicit solutions we obtain demonstrate quantitatively the effect of the TVaR-regulatory constraint on the form of the optimal reinsurance strategy and how the resulting change in the reinsurance policy aligns with the objective of the regulator.

We start by using Lemma 2.2 and the translation invariance of VaR to recast Problem (2.6) equivalently but more simply as

$$\begin{cases} \inf\limits_{f\in\mathcal{F}} & \displaystyle\int_0^\infty [(1+\theta)S_X(t) - 1_{\{S_X(t)>1-\alpha\}}]\,\mathrm{d}f(t) \\ \text{s.t.} & \displaystyle\int_0^\infty \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right]\mathrm{d}f(t) \leq \pi', \end{cases} \tag{4.2}$$

where $\pi' = \pi - \mathrm{TVaR}_\beta(X)$. This representation also exhibits the appropriate functions one should set in Problem (3.1):

$$G(x) = (1+\theta)x - 1_{\{x>1-\alpha\}} \quad \text{and} \quad H(x) = (1+\theta)x - \frac{x}{1-\beta} \wedge 1.$$

To solve Problem (2.6), or equivalently, Problem (4.2), it remains to identify the sets of losses on which $G(S_X(\cdot))$ and $H(S_X(\cdot))$ take a definite sign, examine the structure of the ratio R_X, and apply Theorem 3.1.

Proposition 4.2. (Solutions of TVaR-constrained VaR minimization problem (2.6)) *Consider Problem* (2.6), *and assume that* $\theta/(\theta+1) \leq \alpha \wedge \beta$.

(a) If

$$\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{F_X^{-1}(\alpha)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t \leq \pi',$$

then an optimal solution of Problem (2.6) *is of a limited stop-loss form:*

$$f^*(x) = \left[x - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+ - \left[x - F_X^{-1}(\alpha)\right]_+ = \left[x \wedge F_X^{-1}(\alpha) - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+.$$

(b) Suppose that

$$\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{\infty} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t \leq \pi' < \int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{F_X^{-1}(\alpha)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t.$$

Then an optimal solution of Problem (2.6) *takes the form of two insurance layers:*

$$\begin{aligned} f^*(x) &= \left[x - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+ - \left[x - F_X^{-1}(\alpha)\right]_+ + (x-d^*)_+ \\ &= \left[x \wedge F_X^{-1}(\alpha) - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+ + (x-d^*)_+, \end{aligned}$$

where $d^* > F_X^{-1}(\alpha)$ *and satisfies*

$$\int_{\left[F_X^{-1}\left(\frac{\theta}{\theta+1}\right), F_X^{-1}(\alpha)\right] \cup [d^*,\infty)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t = \pi'.$$

(c) If

$$\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{\infty} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t > \pi',$$

then Problem (2.6) *has no solution.*

Proof. Since Cases (a) and (c) are trivial, we only prove Case (b), which is of the highest practical relevance. To utilize Theorem 3.1, we are prompted to identify the appropriate case in the theorem to which Case (b) of the current proposition corresponds. To this end, we observe from Equivalence (2.2) and the hypothesis $\theta/(\theta+1) \leq \alpha \wedge \beta$ that

$$G(S_X(t)) \leq 0 \Leftrightarrow (1+\theta)S_X(t) \leq 1_{\{S_X(t)>1-\alpha\}} \Leftrightarrow F_X^{-1}\left(\frac{\theta}{\theta+1}\right) \leq t < F_X^{-1}(\alpha)$$

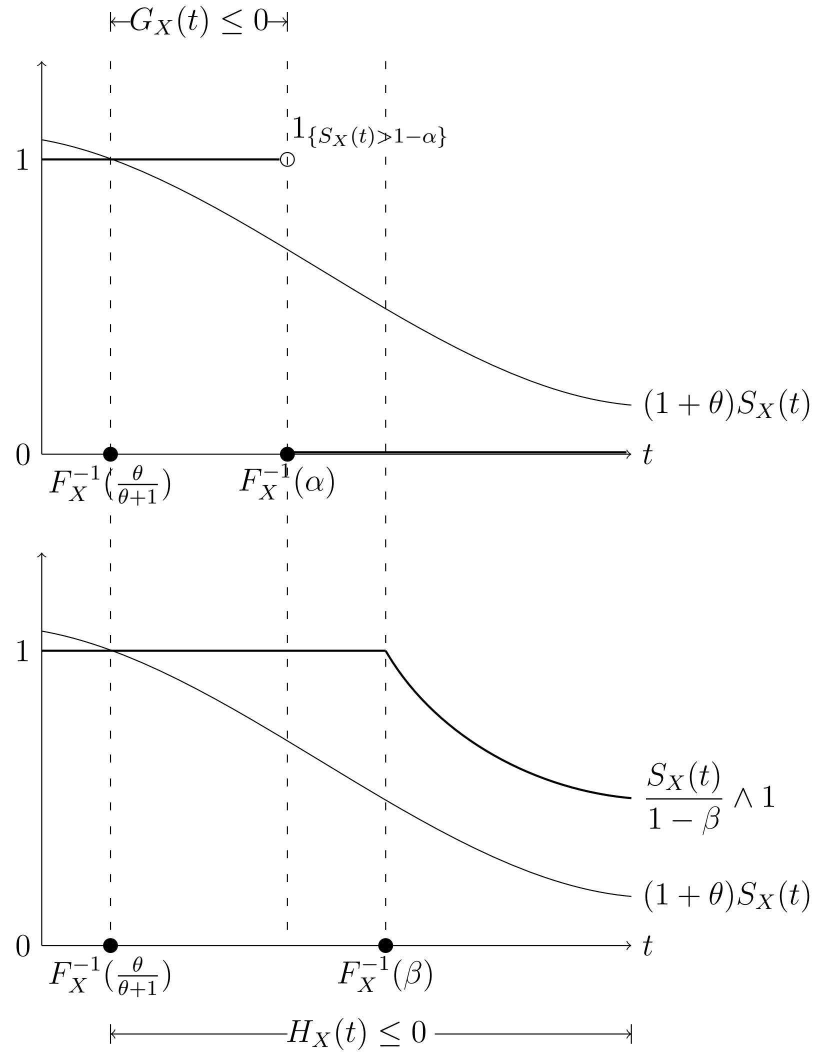

Figure 1: Illustrations of various functions arising in Problem (2.6) when $\alpha \le \beta$.

and that

$$H(S_X(t)) \le 0 \quad \Leftrightarrow \quad (1+\theta)S_X(t) \le \frac{S_X(t)}{1-\beta} \wedge 1 \quad \Leftrightarrow \quad t \ge F_X^{-1}\left(\frac{\theta}{\theta+1}\right).$$

Then the four sets introduced in Table 1 are given by (see Figure 1):

$$\begin{aligned}
A &= \{G(S_X(\cdot)) \le 0, H(S_X(\cdot)) \le 0\} = [F_X^{-1}(\theta/(\theta+1)), F_X^{-1}(\alpha)),\\
B &= \{G(S_X(\cdot)) \le 0, H(S_X(\cdot)) > 0\} = \emptyset,\\
C &= \{G(S_X(\cdot)) > 0, H(S_X(\cdot)) \le 0\} = [F_X^{-1}(\alpha), \infty),\\
D &= \{G(S_X(\cdot)) > 0, H(S_X(\cdot)) > 0\} = [0, F_X^{-1}(\theta/(\theta+1))).
\end{aligned}$$

In particular, notice that $A = A \cup B$, ruling out Case 2 in Theorem 3.1. Since the condition

$$\int_{A\cup C} H(S_X(t))\,\mathrm{d}t \le \pi' < \int_A H(S_X(t))\,\mathrm{d}t$$

is now equivalent to

$$\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{\infty} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t \leq \pi' < \int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{F_X^{-1}(\alpha)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t,$$

the hypothesis of Case (b) of the current proposition coincides with that of Case 3 of Theorem 3.1. Moreover, on Set C,

$$R_X(t) = \begin{cases} \dfrac{(1+\theta)S_X(t)}{(1+\theta)S_X(t)-1} = 1 + \dfrac{1}{(1+\theta)S_X(t)-1}, & \text{if } F_X^{-1}(\alpha) \leq t < F_X^{-1}(\beta), \\ \dfrac{(1+\theta)S_X(t)}{(1+\theta)S_X(t) - S_X(t)/(1-\beta)} = \dfrac{(1+\theta)(1-\beta)}{(1+\theta)(1-\beta)-1}, & \text{if } F_X^{-1}(\beta) \leq t, \end{cases}$$

for $\alpha \leq \beta$, and

$$R_X(t) = \frac{(1+\theta)S_X(t)}{(1+\theta)S_X(t) - S_X(t)/(1-\beta)} = \frac{(1+\theta)(1-\beta)}{(1+\theta)(1-\beta)-1}$$

for $\beta < \alpha$. Regardless of the relative values of α and β, $R_X(t)$ is non-decreasing in t for all $t \geq F_X^{-1}(\alpha)$ (i.e. when t lies in Set C), so the losses corresponding to the least negative values of R_X are precisely the losses in the right tail. In accordance with Case 3 of Theorem 3.1, an optimal solution of Problem (2.6) is (note that B^*, as a subset of Set B, is empty)

$$f^*(x) = \int_0^x 1_{\left[F_X^{-1}\left(\frac{\theta}{\theta+1}\right), F_X^{-1}(\alpha)\right) \cup [d^*,\infty)}(t)\,\mathrm{d}t = \left[x \wedge F_X^{-1}(\alpha) - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+ + (x-d^*)_+,$$

where $d^* > F_X^{-1}(\alpha)$ and satisfies

$$\int_{\left[F_X^{-1}\left(\frac{\theta}{\theta+1}\right), F_X^{-1}(\alpha)\right] \cup [d^*,\infty)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t = \pi'. \qquad \square$$

We remark that the assumption $\theta/(\theta+1) \leq \alpha \wedge \beta$, which is imposed to simplify the presentation of the optimal solutions (although our methodology based on Theorem 3.1 also applies to the complementary case when $\theta/(\theta+1) > \alpha \wedge \beta$), is very mild indeed and, for all intents and purposes, satisfied in practice, because the profit loading θ charged by the reinsurer usually takes a small positive value whereas the probability levels α and β that define the VaR and TVaR risk measures tend to approach one. Furthermore, the regulator is likely to be more conservative than the insurer with regard to the measurement of risk, so the case when $\alpha \leq \beta$ is of much higher practical importance.

Proposition 4.2 is a clear manifestation of the flaws of VaR when it is adopted to aid decision-making, and how its critical deficiencies can be remedied by regulatory constraints based on appropriate tail risk measures. When the optimal reinsurance treaty is designed with the sole goal of minimizing the VaR of the insurer's total retained risk, the optimal solution, as a by-product of Case (a) of Proposition 4.2, takes a limited stop-loss form (see also Corollary 3.1 of Cui et al. (2013) and Example 3.5 of Cheung and Lo (2015)), consistent with the empirical findings in Froot (2001). This implies that the insurer cedes only moderate-sized losses, but leaves the worst losses uninsured and exposed. This potentially catastrophic phenomenon, highly undesirable from the regulatory perspective, can be explained by the fundamental character of VaR, which only utilizes information about the lower tail of the loss distribution and completely ignores the severity of extreme losses. In compliance with a stringent regulatory constraint as in Case (b), however, the insurer will be compelled to reinsure not only medium-sized losses to reduce the VaR of its total risk exposure, but also extreme large losses to maintain the TVaR of its exposure below the regulatory level. The resulting total retained risk of the insurer becomes bounded from above. The regulatory constraint effectively motivates the insurer to implement a more prudent risk management policy, where the right tail of the loss distribution is fully hedged, thereby stabilizing the financial well-being of the insurer and in turn safeguarding the interests of its policyholders.

We end this subsection with a numerical example illustrating the application of Proposition 4.2 to a specific loss distribution and, more importantly, the impact of a stringent regulatory constraint on the optimal reinsurance arrangement.

Example 4.3. *(Exponential ground-up loss with a stringent regulatory constraint)* In Problem (2.6), suppose the ground-up loss X is exponentially distributed with a mean of 1,000, $\alpha = 0.9$, $\beta = 0.95$ (i.e. the regulator is more conservative than the insurer), $\theta = 0.1$, and $\pi = 2,000$. Since $\pi' = \pi - \text{TVaR}_\beta(X) = 2,000 - (1,000 - 1,000 \ln 0.05) = -1,995.73$, and

$$\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{F_X^{-1}(\alpha)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t = \int_{F_X^{-1}\left(\frac{1}{11}\right)}^{F_X^{-1}(0.9)} (1.1\mathrm{e}^{-t/1,000} - 1)\, \mathrm{d}t = -1,317.27 > \pi',$$

but

$$\begin{aligned}
&\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{\infty} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t \\
&= \int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{F_X^{-1}(\beta)} [(1+\theta)S_X(t) - 1]\, \mathrm{d}t + \int_{F_X^{-1}(\beta)}^{\infty} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta}\right] \mathrm{d}t = -2,900.42 < \pi',
\end{aligned}$$

we are in the setting of Case (b) of Proposition 4.2. The optimal deductible d^* is determined via solving the equation

$$\begin{aligned}
&\int_{[F_X^{-1}\left(\frac{\theta}{\theta+1}\right), F_X^{-1}(\alpha)] \cup [d^*, \infty)} \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\beta} \wedge 1\right] \mathrm{d}t \\
&\qquad = \int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^{F_X^{-1}(\alpha)} [(1+\theta)S_X(t) - 1]\, \mathrm{d}t + \left[(1+\theta) - \frac{1}{1-\beta}\right] \int_{d^*}^{\infty} S_X(t)\, \mathrm{d}t = \pi',
\end{aligned}$$

resulting in $d^* = 3,327.10$. Therefore, the optimal solution of Problem (2.6) is given by

$$f^*(x) = \left[x \wedge F_X^{-1}(0.9) - F_X^{-1}\left(\frac{1}{11}\right)\right]_+ + (x - d^*)_+ = (x \wedge 2,302.59 - 95.31)_+ + (x - 3,327.10)_+.$$

As a comparison, the optimal solution of the unconstrained version of Problem (2.6) is $f^{*,\text{unconstrained}}(x) = (x \wedge 2,302.59 - 95.31)_+$. □

4.3 Model 3: Risk minimization subject to a reinsurer's risk constraint

We now turn to the insurer's TVaR minimization problem (2.8) under a reinsurer's VaR-risk constraint and the expectation premium principle, paying special attention to whether the additional participation constraint motivates or demotivates the insurer to exploit reinsurance, and whether the (dis)incentives stem from the choice of particular distortion risk measures. In view of Lemma 2.2 and the translation invariance of TVaR, Problem (2.8) is equivalent to

$$\begin{cases} \inf\limits_{f\in\mathcal{F}} & \displaystyle\int_0^\infty \left[(1+\theta)S_X(t) - \frac{S_X(t)}{1-\alpha}\wedge 1\right] \mathrm{d}f(t) \\ \text{s.t.} & \displaystyle\int_0^\infty \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}f(t) \le \pi, \end{cases}$$

which suggests setting

$$G(x) = (1+\theta)x - \frac{x}{1-\alpha}\wedge 1 \quad \text{and} \quad H(x) = 1_{\{x>1-\beta\}} - (1+\theta)x$$

in Problem (3.1) to yield the following explicit solutions of Problem (2.8).

Proposition 4.4. (Solutions of TVaR minimization problem (2.8) subject to reinsurer's VaR-risk constraint) *Consider Problem* (2.8)*, and assume that* $\theta/(\theta+1) \le \alpha\wedge\beta$.

(a) If

$$\int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^\infty \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t \le \pi,$$

then an optimal solution of Problem (2.8) *is of a stop-loss form:*

$$f^*(x) = \left[x - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+.$$

(b) Suppose that

$$\int_{F_X^{-1}(\beta)}^\infty \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t \le \pi < \int_{F_X^{-1}\left(\frac{\theta}{\theta+1}\right)}^\infty \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t.$$

Then an optimal solution of Problem (2.8) *is a double insurance layer:*

$$f^*(x) = \left[x\wedge u^* - F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right]_+ + [x - F_X^{-1}(\beta)]_+,$$

where $u^* \in [F_X^{-1}(\theta/(\theta+1)), F_X^{-1}(\beta))$ *and satisfies*

$$\int_{\left[F_X^{-1}\left(\frac{\theta}{\theta+1}\right),u^*\right]\cup[F_X^{-1}(\beta),\infty)} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t = \pi.$$

(c) Suppose that

$$\int_{F_X^{-1}(\beta)}^{\infty} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t > \pi$$

but

$$\int_{\left[0,F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right)\cup[F_X^{-1}(\beta),\infty)} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t \le \pi.$$

Then an optimal solution of Problem (2.8) *is a double insurance layer:*

$$f^*(x) = x \wedge u^* + [x - F_X^{-1}(\beta)]_+,$$

where $u^* \in (0, F_X^{-1}(\theta/(\theta+1))]$ *and satisfies*

$$\int_{[0,u^*]\cup[F_X^{-1}(\beta),\infty)} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t = \pi.$$

(d) If

$$\int_{\left[0,F_X^{-1}\left(\frac{\theta}{\theta+1}\right)\right)\cup[F_X^{-1}(\beta),\infty)} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t > \pi,$$

then Problem (2.8) *has no solution.*

Proof. For brevity, we only prove Case (c). The four sets defined in Table 1 are given by (see Figure 2)

$$\begin{aligned}
A &= \{G(S_X(\cdot)) \le 0, H(S_X(\cdot)) \le 0\} = [F_X^{-1}(\beta), \infty),\\
B &= \{G(S_X(\cdot)) \le 0, H(S_X(\cdot)) > 0\} = [F_X^{-1}(\theta/(\theta+1)), F_X^{-1}(\beta)),\\
C &= \{G(S_X(\cdot)) > 0, H(S_X(\cdot)) \le 0\} = [0, F_X^{-1}(\theta/(\theta+1))),\\
D &= \{G(S_X(\cdot)) > 0, H(S_X(\cdot)) > 0\} = \emptyset.
\end{aligned}$$

The hypothesis of Case (c) in the current proposition therefore corresponds to that of Case 3 in Theorem 3.1. On Set C, the ratio R_X is constant at

$$R_X(t) = \frac{(1+\theta)S_X(t) - 1}{1 - (1+\theta)S_X(t)} = -1,$$

while on Set B,

$$R_X(t) = \begin{cases} \dfrac{(1+\theta)S_X(t) - 1}{1 - (1+\theta)S_X(t)} = -1, & \text{if } F_X^{-1}\left(\dfrac{\theta}{\theta+1}\right) \le t < F_X^{-1}(\beta),\\[2ex] \dfrac{(1+\theta)S_X(t) - S_X(t)/(1-\alpha)}{1 - (1+\theta)S_X(t)}, & \text{if } F_X^{-1}(\alpha) \le t < F_X^{-1}(\beta) \text{ and } \alpha \le \beta, \end{cases}$$

which is non-decreasing in t. Since $c_2^* = c^* = -1 = \inf_{t\in B} R_X(t)$, it follows from Case 3 of Theorem 3.1 that an optimal ceded loss function is

$$f^*(x) = \int_0^x 1_{[0,u^*]\cup[F_X^{-1}(\beta),\infty)}(t)\,\mathrm{d}t = x \wedge u^* + [x - F_X^{-1}(\beta)]_+,$$

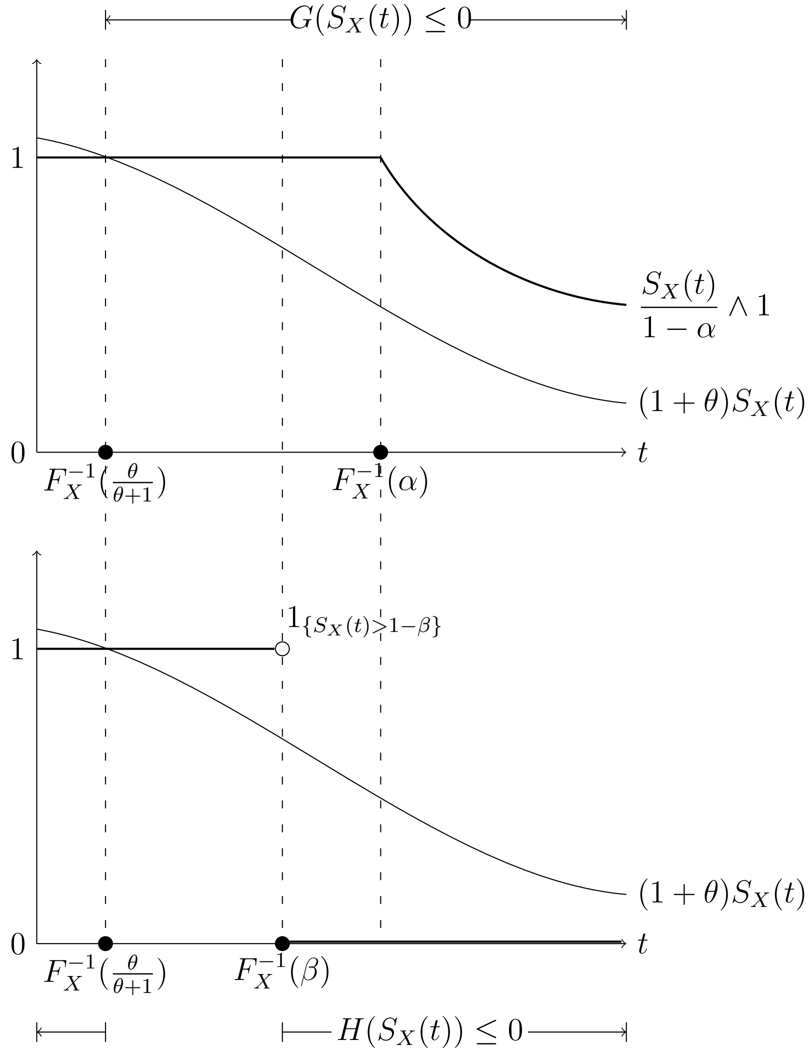

Figure 2: Illustrations of various functions arising in Problem (2.8) when $\alpha > \beta$.

where $u^* \in (0, F_X^{-1}(\theta/(\theta+1))]$ and satisfies

$$\int_{[0,u^*]\cup[F_X^{-1}(\beta),\infty)} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t = \pi. \qquad \square$$

The significance of Proposition 4.4 (b) and (c) consists in the unanticipated compulsion on the insurer to protect itself against *small* losses in the presence of a stringent reinsurer's risk constraint. When adopting TVaR as the risk metric, the insurer finds it beneficial to fully hedge the right tail of the loss distribution to minimize its total retained risk. This is a well-documented phenomenon (see, for example, Chi and Tan (2011), Cui et al. (2013)) consistent with TVaR being an arithmetic average of the severity of tail losses. Fully hedging the right tail, however, may expose the reinsurer to an unacceptably high level of risk and deter it from offering the reinsurance contract in the first place. To stabilize the reinsurer's retained risk and make the resulting reinsurance treaty mutually acceptable, the insurer is also prompted to cede small-sized losses, resulting in a slight rise in the insurer's retained risk but a decline in the reinsurer's risk exposure to an acceptable level. This explains the peculiarity of the insurer ceding small losses in conjunction with extreme large losses.

Example 4.5. *(Exponential ground-up loss with a stringent reinsurer's risk constraint)* In Problem (2.8), suppose that the ground-up loss X is exponentially distributed with a mean of 1,000, $\alpha = 0.95$, $\beta = 0.9$, $\theta = 0.1$, and $\pi = -112$. Since

$$\int_{F_X^{-1}(\beta)}^{\infty} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t = -1.1\int_{F_X^{-1}(0.9)}^{\infty} \mathrm{e}^{-t/1{,}000}\,\mathrm{d}t = -110 > \pi,$$

but

$$\begin{aligned}\int_{[0,F_X^{-1}(\theta/(\theta+1)))\cup[F_X^{-1}(\beta),\infty)} &\left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t \\ &= F_X^{-1}(1/11) - 1.1\int_{[0,F_X^{-1}(1/11))\cup[F_X^{-1}(0.9),\infty)} \mathrm{e}^{-t/1{,}000}\mathrm{d}t = -114.69 \leq \pi,\end{aligned}$$

we are in the setting of Case (c) of Proposition 4.4. Since $u^* = 22.85$ satisfies

$$\int_{[0,u^*]\cup[F_X^{-1}(\beta),\infty)} \left[1_{\{S_X(t)>1-\beta\}} - (1+\theta)S_X(t)\right] \mathrm{d}t = \pi,$$

an optimal solution of Problem (2.8) is $f^*(x) = x \wedge 22.85 + (x + 1,000\ln 0.1)_+$.

5 Concluding remarks

The main contributions of this article are twofold. Technically, it introduces an elementary solution scheme which lends itself to a universal framework for tackling a wide body of constrained optimal reinsurance problems. This heuristic solution scheme overcomes the technical difficulties emanating from the presence of optimization constraints and avoids the use of sophisticated mathematical techniques such as Lagrangian duality methods. Practically and more remarkably, this article is among the first to undertake a comprehensive analysis of the economic ramifications of various optimization constraints on the demand for reinsurance in the realm of general distortion-risk-measure-based reinsurance models. Upon specializing our analysis to VaR and TVaR, it is shown concretely that the insurer is compelled to hedge against large losses in the face of a stringent regulatory constraint (Problem (2.6)), and against small losses when confronted with a severe reinsurer's risk constraint (Problem (2.8)). These unanticipated features of the resulting optimal reinsurance strategies can be attributed to the necessity of balancing the conflicting interests of various parties.

Acknowledgments

The author would like to thank the anonymous reviewers for their careful reading and constructive comments. The author also gratefully acknowledges the support from Old Gold Summer Fellowship 2015–16 provided by the College of Liberal Arts and Sciences, The University of Iowa, and a Centers of Actuarial Excellence (CAE) Research Grant (2013–2016) from the Society of Actuaries.

References

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

Bernard, C., Tian, W., 2009. Optimal reinsurance arrangements under tail risk measures. The Journal of Risk and Insurance 76, 709–725.

Bernard, C., Tian, W., 2010. Insurance market effects of risk management metrics. The Geneva Risk and Insurance Review 35, 47–80.

Borch, K., 1960. An attempt to determine the optimum amount of stop loss reinsurance. Transactions of the 16th International Congress of Actuaries 1, 597–610.

Cai, J., Tan, K.S., 2007. Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measure. ASTIN Bulletin 37 (1), 93–112.

Cai, J., Tan, K.S., Weng, C., Zhang, Y., 2008. Optimal reinsurance under VaR and CTE risk measures. Insurance: Mathematics and Economics 43, 185–196.

Cheung, K.C., Dhaene, J., Lo, A., Tang, Q., 2014. Reducing risk by merging countermonotonic risks. Insurance: Mathematics and Economics 54, 58–65.

Cheung, K.C., Liu, F., Yam, S.C.P., 2012. Average Value-at-Risk minimizing reinsurance under Wang's premium principle with constraints. ASTIN Bulletin 42, 575–600.

Cheung, K.C., Lo, A., 2015. Characterizations of optimal reinsurance treaties: a cost-benefit approach. Scandinavian Actuarial Journal (in press). DOI:10.1080/03461238.2015.1054303.

Chi, Y., Tan, K.S., 2011. Optimal reinsurance under VaR and CVaR risk measures: a simplified approach. ASTIN Bulletin 41, 487–509.

Cui, W., Yang, J., Wu, L., 2013. Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles. Insurance: Mathematics and Economics 53, 74–85.

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

Dhaene, J., Vanduffel, S., Goovaerts, M.J., Kaas, R., Tang, Q., Vyncke, D., 2006. Risk measures and comonotonicity: A review. Stochastic Models 22, 573–606.

Eeckhoudt, L., Gollier, C., Schlesinger, H., 2005. Economic and Financial Decisions under Risk. Princeton University Press, New Jersey.

Froot, K.A., 2001. The market for catastrophe risk: a clinical examination. Journal of Financial Economics 60, 529–571.

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

Kaluszka, M., Okolewski, A., 2008. An extension of Arrow's result on optimal reinsurance contract. The Journal of Risk and Insurance 75, 275–288.

Yaari, M.E., 1987. The dual theory of choice under risk. Econometrica 55, 95–115.

Young, V., 1999. Optimal insurance under Wang's premium principle. Insurance: Mathematics and Economics 25, 109–122.

Zheng, Y., Cui, W., 2014. Optimal reinsurance with premium constraint under distortion risk measures. Insurance: Mathematics and Economics 59, 109–120.

Zhou, C., Wu, C., 2008. Optimal insurance under the insurer's risk constraint. Insurance: Mathematics and Economics 42, 992–999.

Zhou, C., Wu, C., 2009. Optimal insurance under the insurer's VaR constraint. The Geneva Risk and Insurance Review 34, 140–154.

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

A Appendix: Proof of Theorem 3.1

Proof. Since Cases 1 and 4 are trivial, and the proofs of Cases 2 and 3 closely resemble each other, to avoid unnecessary repetition we only prove Case 3, which is the most interesting and complex case. To start, we observe that f^*, by definition, resides in $\mathcal{F}$ with

$$\int_0^\infty H(S_X(t))(f^*)'(t)\,\mathrm{d}t = \int_{A\cup B^*\cup C^*} H(S_X(t))\,\mathrm{d}t = \pi \tag{A.1}$$

due to the choice of c^*. For any fixed $f \in \mathcal{F}$ satisfying $\int_0^\infty H(S_X(t))f'(t)\,\mathrm{d}t \leq \pi$, the optimality of f^* entails $\int_0^\infty G(S_X(t))[(f^*)'(t) - f'(t)]\,\mathrm{d}t \leq 0$, or equivalently,

$$\int_{A\cup B^*\cup C^*} G(S_X(t))[1 - f'(t)]\,\mathrm{d}t \leq \int_{\{H(S_X(\cdot))\leq 0, R_X(\cdot)<c^*\}\cup\{H(S_X(\cdot))>0, R_X(\cdot)\geq c^*\}} G(S_X(t))f'(t)\,\mathrm{d}t. \tag{A.2}$$

To prove this, notice that

$$A\cup C^* = \{H(S_X(\cdot)) \leq 0, R_X(\cdot) \geq c^*\} = \{H(S_X(\cdot)) \leq 0, G(S_X(\cdot)) \leq c^* H(S_X(\cdot))\} \tag{A.3}$$

and

$$B^* = \{H(S_X(\cdot)) > 0, R_X(\cdot) < c^*\} = \{H(S_X(\cdot)) > 0, G(S_X(\cdot)) < c^* H(S_X(\cdot))\}. \tag{A.4}$$

Therefore, keeping in mind that $1 - f'(t) \geq 0$ for any $t \geq 0$,

$$\begin{aligned}\int_{A\cup B^*\cup C^*} G(S_X(t))[1 - f'(t)]\,\mathrm{d}t &\leq c^* \int_{A\cup B^*\cup C^*} H(S_X(t))[1 - f'(t)]\,\mathrm{d}t \\ &= c^*\left(\pi - \int_{A\cup B^*\cup C^*} H(S_X(t))f'(t)\,\mathrm{d}t\right),\end{aligned} \tag{A.5}$$

where the last equality follows from Equation (A.1). Since

$$\begin{aligned}\int_0^\infty H(S_X(t))f'(t)\,\mathrm{d}t &= \int_{A\cup B^*\cup C^*} H(S_X(t))f'(t)\,\mathrm{d}t \\ &\quad + \int_{\{H(S_X(\cdot))\leq 0, R_X(\cdot)<c^*\}\cup\{H(S_X(\cdot))>0, R_X(\cdot)\geq c^*\}} H(S_X(t))f'(t)\,\mathrm{d}t \\ &\leq \pi,\end{aligned} \tag{A.6}$$

and $c^* < 0$, combining Inequalities (A.5) and (A.6) further yields

$$\int_{A\cup B^*\cup C^*} G(S_X(t))[1-f'(t)]\,\mathrm{d}t \leq c^*\left(\int_{\{H(S_X(\cdot))\leq 0, R_X(\cdot)<c^*\}\cup\{H(S_X(\cdot))>0, R_X(\cdot)\geq c^*\}} H(S_X(t))f'(t)\,\mathrm{d}t\right),$$

which, in view of (A.3) and (A.4) again, is bounded above by

$$\int_{\{H(S_X(\cdot))\leq 0, R_X(\cdot)<c^*\}\cup\{H(S_X(\cdot))>0, R_X(\cdot)\geq c^*\}} G(S_X(t))f'(t)\,\mathrm{d}t.$$

This proves Inequality (A.2) and establishes the optimality of f^*. □