A Continuous-Time Optimal Insurance Design with Costly Monitoring

Hisashi Nakamura*
Hitotsubashi University
Email: hisashi.nakamura@r.hit-u.ac.jp

November 30, 2011

Abstract

I provide a theoretical framework to study optimal insurance properties for players' general utility forms in a continuous-time environment where an insurer can observe neither the efforts nor the outcome of an insured firm. The insured may then cause two problems: the intentional loss and the exaggerated claim. I show theoretically that, using costly monitoring effectively, the two problems can be differentiated in an optimal insurance contract. Furthermore, if the insured is not downward-risk averse highly and if the participation constraint is not too tight, then the monitoring can mitigate the two problems. This model is tractable for numerical work.

Keywords: Insurance, Costly monitoring, Moral hazard, Ex-post informational asymmetry.
JEL Classification: D82, D86, G22, G32.

*Graduate School of Commerce and Management, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Phone: +81-42-580-8826. Fax: +81-42-580-8747. Email: hisashi.nakamura@r.hit-u.ac.jp. I am working with Koichiro Takaoka for improving mathematical descriptions. I would like to thank Takau Yoneyama for his constant encouragement and advice throughout the preparation of this work. I am also thankful to Ricard Gil, Jean Kwon, Soichiro Moridaira, Mahito Okura, Shuji Tanaka, and all seminar participants at Working Group on Globalization of Insurance Research (sponsored by The Dai-ichi Life Insurance Company, Limited and Dai-ichi Life Research Institute Inc.), Asia-Pacific Risk and Insurance Association 15th Annual Conference (Tokyo) and the Japanese Association of Risk, Insurance and Pensions Annual Conference (Tokyo). All errors are my own.

1 Introduction

As it is well known, insurers are exposed to information problems in most corporate insurance practices (e.g., MacMinn and Garven [12]). Specifically, the insurer can observe directly neither the efforts nor the outcome of an insured firm without a cost. The insured may then cause purposely the loss and/or the exaggerated claim. Call the former the problem of moral hazard and the latter the problem of ex-post informational asymmetry.

The ex-post informational asymmetry problem distorts the insured's effort incentives. In standard moral hazard models, it is assumed that the insurer can observe the insured's outcome ex post (e.g., Rogerson [15]). The insurer then faces only the incentive problem of inducing the insured to make desired efforts. In practice, by contrast, because of the ex-post informational asymmetry, the insurer needs to provide additionally the insured with the incentive to induce him to tell the truth ex post. In general, it is difficult to write appropriate insurance contracts to distinguish the problem of moral hazard and the problem of ex-post informational asymmetry. To identify and reduce the two information problems, the insurer routinely investigates a claim via a costly monitoring technology after it is filed (Harrington and Niehaus [9]).

The purpose of this paper is to provide a theoretical framework to study optimal insurance properties with monitoring under moral hazard and ex-post informational asymmetry by exploring a continuous-time optimal contracting model with costly monitoring. In this paper, I consider an environment in which an insurer can observe neither the efforts nor the outcome of an insured firm but can monitor the outcome only when incurring a monitoring cost. The insurer writes an optimal contract to maximize her own expected utility, inducing the insured to tell the truth while trying to reduce the expected monitoring cost. I examine optimal insurance properties, in particular, dynamic equilibrium interaction between the problem of moral hazard and the problem of ex-post informational asymmetry.

In relationships to previous literatures, the paper of Cvitanić and Zhang [5] is close to this paper, in that it studies optimal contracting in an environment in which there exist both hidden actions and hidden information in a continuous-time principal-agent model. Due to mathematical tractability, the continuous-time framework is useful for characterizing optimal contract properties in complex, dynamic environments.

Still, this present paper departs from Cvitanić and Zhang [5] mainly in two respects. First, regarding the problem of hidden information, this paper looks at ex-post informational asymmetry in the sense that the insurer cannot observe the insured's outcome ex post, whereas their paper looks at ex-ante adverse selection in the sense that the principal cannot observe the agent's ex ante production ability. Second, Cvitanić and Zhang [5] assume only costless reporting as a communication method, as usual in contract theory.[1] Their paper shows that, in general, it is very hard to distinguish the moral hard problem and the ex-ante adverse selection problem from a dynamic viewpoint as well as from a static viewpoint, because it is often difficult to compute dynamically the Lagrangian multipliers associated with the two information problems.[2] On the other hand, as it is mentioned above, in insurance practices, insurance companies routinely verify the reports of insured firms via costly monitoring after claims are filed. Thus monitoring is another crucial communication method in insurance contracts. In this paper, I incorporate costly monitoring in the model in order to study a crucial role to distinguish moral hazard and ex-post informational asymmetry in optimal insurance contracts.

Main results are as follows. If optimal efforts are attained, the insurer can write the optimal insurance contract that differentiates the problem of moral hazard and the problem of ex-post informational asymmetry, by using the costly monitoring effectively. In particular, because of the monitoring technology, the ex-post informational asymmetry problem is reduced, although the insured can still enjoy an information advantage only while in good shape. If the insured is not downward-risk averse highly and if the participation constraint of the insured is not too tight with respect to the monitoring cost, then a higher level of the monitoring technology (i.e., a lower monitoring cost) can mitigate the problem of moral hazard. This model is very tractable for numerical work. For example, consider the case that the insured has a log utility and the insurer is risk-neutral. When the monitoring cost is an immediate level, the monitoring action is undertaken only for poor cash flows, while the contract is deductible for very low outcomes. I.e.,

[1] In much of the literature on contract theory, communication games with costless reporting have been studied a lot in finite-horizon (typically, two or three period) discrete-time models. There are a few exceptions in a literature on insurance fraud (e.g., Dionne, Giuliano and Picard [6], Picard [14]).

[2] In a static model, in both cases, more comprehensive coverage is associated with high risk (Chiappori [4]). Accordingly, when observing the agent's poor outcome in a static way, it is difficult to identify whether it would be due to their ability or to their laziness. To overcome this difficulty, much empirical insurance literature has differentiated moral hazard and ex-ante adverse selection by making use of some different dynamic properties of the incentive structures between the two information problems in insurance contracts for various *exogenous* cases (typically, a reform of regulatory framework), without using dynamic optimal (i.e., *endogenous*) insurance properties (e.g., Abbring et al. [1]).

the optimal contract takes the form of an insurance contract with deductibles. Furthermore, The insurance premium (the compensation trigger, respectively) is slightly increasing (decreasing) in the monitoring cost. For very low monitoring costs, the monitoring action is undertaken for all feasible cash flows, and the allocations are state-dependent. On the other hand, for very high monitoring costs, the monitoring action is necessarily avoided in equilibrium. The optimal contract is then of a state-independent debt type for all feasible cash flows. The insured is then given no compensation for low cash flows.

This paper is organized as follows. Next section defines an environment. Section 3 studies optimal insurance properties. Section 4 obtains numerical results. Final section concludes.

2 Environment

I consider an optimal contracting problem between two players: an insurer (i.e., insurance company) and an insured (i.e., firm) on a time interval $[0, T]$ for a finite time $T > 0$. There exists a single consumption good. $U_i : \mathbf{R} \mapsto \mathbf{R}$ is player i's utility function of his or her own consumption $\gamma_i \in \mathbf{R}$ at time T $(i = 1, 2)$, where $i = 1$ denotes the insured and $i = 2$ denotes the insurer. The utility function U_i $(i = 1, 2)$ is three times continuously differentiable. $U_i' > 0$ for each $i = 1, 2$ and $U_1'' < 0$ and $U_2'' \leq 0$. Let a real-valued u denote instantaneous expected return of the insured's production. As it will be shown shortly below, the insured can control it – call it the insured's effort. Also, $G_T := \int_0^T g(u_t)\,\mathrm{d}t := \frac{1}{2}\int_0^T (u_t)^2\,\mathrm{d}t$ denotes the insured's utility cost for controlling his effort u. The insured is exogenously given a reservation utility at time 0, denoted by a constant $r \in \mathbf{R}$. For convenience, I will use female pronouns for the insurer, and male ones for the insured.

Fix a probability space $(\Omega, \mathcal{F}, P)$. Let B be a one-dimensional standard Brownian motion on the probability space, and $\mathbb{F}^B = \{\mathcal{F}_t\}_{0 \leq t \leq T}$ be the filtration generated by B up to time $T > 0$. The insured produces the process of cash flows (income) X, which is characterized by the following stochastic differential equation: for a finite constant $v > 0$, $X_t = x + vB_t$ where v stands for the riskiness of the cash-flow process. The insured's effort processes u are $\mathbb{F}^B$-adapted. For any $\mathbb{F}^B$-adapted real-valued processes u, let

$$B_t^u := B_t - \int_0^t u_s\,\mathrm{d}s, \quad M_t^u := \exp\left(\int_0^t u_s\,\mathrm{d}B_s - \frac{1}{2}\int_0^t u_s^2\,\mathrm{d}s\right), \quad \frac{\mathrm{d}P^u}{\mathrm{d}P} := M_T^u.$$

Assume that u satisfies the conditions required by the Girsanov Theorem (e.g., the Novikov condition) under P. Then M_t^u is a martingale and P^u is a probability measure. Moreover, B^u is a P^u-Brownian motion and

$$\mathrm{d}X_t = v\,\mathrm{d}B_t = u_t v\,\mathrm{d}t + v\,\mathrm{d}B_t^u, \quad X_0 = x.$$

The insured controls the effort u. In other words, a higher (lower) effort leads to a higher (lower) expected return of the cash flows.

Assume that v is observable by the insurer, but X, u are not.[3] The insured makes a report of the trajectory of X at time T without a cost. The report may be a lie. Moreover, a monitoring technology is available to the insurer at time T if she incurs a utility cost K_M. The technology is deterministic in the sense that, when demanded, it occurs with probability one, and delivers the true information of the time path of the cash flows to the insurer with perfect accuracy.

The insured enters into a contract with the insurer for insuring against his income risk (typically, liability risk). X_T is shared between the two players at time T according to terms of the contract. The insurer offers a menu of contract payoffs C_T to the insured. At time T, the insurer makes a report of time path of X, denoted by $\tilde{X}$. The report may be a lie. The insured's allocation C_T takes the form of $C_T(X, \tilde{X})$ as a functional of $X, \tilde{X}$. With small abuse of language, call C_T a contract.

3 Optimal insurance design with costly monitoring

3.1 Insured's problem

Define the admissible sets for the insured's controls of the efforts and the reports First, with regard to the efforts u,

Definition 3.1 $\mathcal{A}_0$ *is the admissible set for the efforts u that satisfy:*

(i) *u is $\mathbb{F}^B$-adapted,*

(ii) $P\left(\int_0^T (u_t)^2\,\mathrm{d}t < \infty\right) = 1,$

(iii) *u satisfies the conditions required by the Girsanov Theorem,*

(iv) $E\left[|M_T^u|^4\right] < \infty.$

[3] Because of this assumption, I will be able to use several convenient properties of the Brownian motion (e.g., see Musiela and Rutkowski [13]).

Second, with regard to the reports,

Definition 3.2 *$\mathcal{A}_1$ is the admissible set for $\mathbb{F}^B$-adapted real-valued processes $\tilde{u}$ such that, for any $u \in \mathcal{A}_0$, $\tilde{u} - u \in A_0$ and*

$$\begin{aligned} \mathrm{d}\tilde{X}_t &= -(\tilde{u}_t - u_t)v\,\mathrm{d}t + v\,\mathrm{d}B_t \\ &= -(\tilde{u}_t - u_t)v\,\mathrm{d}t + u_t v\,\mathrm{d}t + v\,\mathrm{d}B_t^u \\ &= v\,\mathrm{d}B_t^{\tilde{u}-u} = u_t v\,\mathrm{d}t + v\,\mathrm{d}B_t^{\tilde{u}}, \quad \tilde{X}_0 = X_0 = x \end{aligned}$$

The term $-(\tilde{u} - u)$ stands for a twisted part of the instantaneous expected return in the reported cash-flow process. In other words, if u is known, the insured can twist the reported measure from the true measure P^u into $P^{\tilde{u}}$. This is why there exists dynamic interaction between the problems of moral hazard and ex-post informational asymmetry.

Define mathematical regularities for the contracts C_T:

Definition 3.3 *Define the admissible set, denoted by $\mathcal{A}_2'$, of the contracts C_T that satisfy:*

(i) *$C_T(X, \tilde{X})$ is $\mathcal{F}_T^B$-measurable,*

(ii) *$E\left[|U_1(C_T)|^4 + e^{4U_1(C_T)}\right] < \infty$,*

(iii) *For $u \in \mathcal{A}_0$, $E^u\left[e^{U_1(C_T)}\,|U_2(X_T - C_T)|\right] < \infty$*

where E^u denotes the expectation operator under P^u.

Note that I will later restrict the contract space $\mathcal{A}_2'$ to a further particular set, as usual in financial contract theory.

Given the admissible sets, the insured's problem is formulated as: for any C_T that satisfies Definition 3.3 (i),(ii),

$$V_1 := \sup_{\tilde{u} \in \mathcal{A}_1} V(\tilde{u}) := \sup_{\tilde{u} \in \mathcal{A}_1} \sup_{u \in \mathcal{A}_0} E^u\left[U_1\left(C_T(X, \tilde{X})\right) - G_T\right].$$

Now, I characterize the insured's optimal effort $\hat{u}$ and $V(\tilde{u})$.

Proposition 3.1 *For any C_T that satisfies Definition 3.3 (i),(ii), and for any $\tilde{u} \in \mathcal{A}_1$, the optimal effort $\hat{u}$ is obtained by solving the backward stochastic differential equation (henceforth, BSDE):*

$$\tilde{Y}_t = e^{U_1(C_T)} - \int_t^T \hat{u}_s \tilde{Y}_s\,\mathrm{d}B_s \tag{3.1}$$

and is in $\mathcal{A}_0$. Moreover, the insured's expected utility before controlling $\tilde{u}$, namely $V(\tilde{u})$, is given by

$$V(\tilde{u}) = \log E\left[e^{U_1(C_T)}\right]. \tag{3.2}$$

Proof: See appendix.

As a direct result from Proposition 3.1, I can check whether the contract C_T induces the insured to make the optimal efforts $\hat{u}$. A contract C_T is said to be implementable for the associated optimal effort $\hat{u}$ (with $\tilde{u} \in \mathcal{A}_1$ given) if, with $\tilde{u} \in \mathcal{A}_1$ given, there exists a one-to-one function $J(\hat{u}; \tilde{u}) = C_T$.

Corollary 3.1 *Any C_T that satisfies Definition 3.3 (i),(ii) is implementable for the associated optimal effort $\hat{u} \in \mathcal{A}_0$, with $\tilde{u} \in \mathcal{A}_1$ given, such that*

$$M_T^{\hat{u}} = e^{-V(\tilde{u})} e^{U_1(C_T(X,\tilde{X}))} \tag{3.3}$$

Call Eq.(3.3) the implementability condition.

Proof: See appendix.

This corollary means that the choice of the probability measure associated with the optimal action $\hat{u}$ has an explicit functional relationship with the insured's allocation C_T.

3.2 Insurer's problem: optimal insurance design

Next, I move on to the insurer's problem. First, as usual in contract theory, I will impose incentive compatibility, participation constraint, and implementability on the contract space. Next, I will solve an optimal insurance design problem. Finally, I will characterize the optimal insurance contracts.

3.2.1 Incentive compatibility, participation constraint and implementability

As usual in financial contract theory, I restrict the above-defined contract space $\mathcal{A}_2'$ to a further particular set, denoted by $\mathcal{A}_2$, in the following two respects. First, I impose the incentive compatibility condition, i.e., the contracts are restricted to the ones that induce the insured to tell the truth. Following standard discussions of costly monitoring, I try debt-type contracts to be incentive compatible. Specifically, I divide the space of the reported $\tilde{X}$ into two strategically predetermined regions: the region that the monitoring action is triggered and its complement. On the one hand, in the region where the monitoring action is not triggered (call this the non-monitoring region), the insurer's share $X_T - C_T$ should be deterministic, independent of the report. This is because, otherwise, while monitoring does not take place, the insured would have an incentive to tell a lie leading to the minimum payment. Let F denote the insurer's deterministic share in the no-monitoring state. When the cash-flow level is high enough, the insured exploits an informational advantage by enjoying the residual cash flow that exceeds the deterministic share F. As usual in contract theory, I assume that, when the insured is indifferent between two actions, he will choose the one that is better to the insurer.

On the other hand, in the region where the monitoring action is triggered (call this the monitoring region), the insurer should monitor the insured's cash flow. If the truth is verified, she should provide the insured with some compensation to insure against the low income. Otherwise, the insured should be penalized with a very large penalty (such as a legal penalty, a reputation loss, etc.). For convenience, assume that, when the monitoring proves false reporting, the insured's allocation C_T is $-\infty$. This means a penalty for the false reporting, which leads to out-of-equilibrium.

Let $C_M(X)$ denote the time-T share (or, compensation) that the insured receives in case that the monitoring action verifies the truth. I impose three assumptions on the set of $C_M(X)$. First, $C_M(X)$ is assumed to be continuous The continuity could be justified if the contract needs to be renegotiation-proof. Second, $C_M(X)$ is assumed to be non-decreasing in X, in that, for any cash flows X, X' such that $X \leq X'$ (i.e., $X_t \leq X'_t \, \forall t$), $C_M(X) \leq C_M(X')$. The no-decreasing structure means a co-payment contractual relationship, which does not look quite restrictive in practice. Third, I assume that $C_M(X)$ satisfies $-\frac{U''_1}{U'_1} \leq U'_1$. This means that the insured is not quite risk-averse. If he is too risk-averse, the necessary, large compensation could negate the contracting opportunity. These assumption will be used below for obtaining incentive compatibility and a necessary and sufficient condition for optimality of C_M.

Furthermore, due to the standard characteristics of the utility functions, there exists a constant b, if feasible, such that, when $\tilde{X}_T$ is lower than b, a monitoring action is triggered, and the compensation C_M is paid. Call b the compensation trigger. In other words, if b is feasible, the insurer verifies the outcome only when the reported outcome is low (i.e., $\tilde{X}_T < b$). The compensation works as a put option. The infeasible b is categorized into the following three cases:

(1) $b = -\infty$ if only the non-monitoring region exists (i.e., the monitoring region does not exist).

(2) $b = +\infty$ if only the monitoring region exists (i.e., the non-monitoring region does not exist).

When $X_T < b$, the insured could have an incentive to tell a lie if he is better off behaving as if he would be in the no-monitoring state. It would not be incentive compatible. Also, when $X_T \geq b$, the insured should not request the monitoring action. Therefore, the compensation C_M should be less than $X_T - F$. Note that, as I will discuss below, $b - F$ may be negative, depending on the players' utility forms. Thus the incentive compatibility condition is written as:

$$\Big(C_M(X) - (X_T - F)\Big)\ (X_T - b) \leq 0 \tag{3.4}$$

Note that, in standard costly monitoring models (e.g., Gale and Hellwig (1985)), due to the assumption of risk neutrality, minimizing the probability of undertaking a monitoring action is equivalent to maximizing the principal's expected cash flows while providing the agent with no lower than his reservation utility. To be incentive compatible, everything should be confiscated from the agent when being monitored in those models. In contrast, in this present model, since the utility functions are non-linear, the insured obtains some positive compensation in optimal risk-sharing allocations when the monitoring action is made. The insured receives the compensation when X_T is larger than F. In other words, the monitoring action can be triggered when the insured is liquid.

In short, the incentive-compatible contract is characterized by a triplet $\{F, b, C_M(X)\}$:

$$C_T(X, \tilde{X}) = \begin{cases} X_T - F = (X_T - \tilde{X}_T) + (\tilde{X}_T - F) & \text{if } \tilde{X}_T \geq b, \\ C_M(X) \quad \text{if } \tilde{X}_T < b \text{ and if } X = \tilde{X}, \\ -\infty \quad \text{if } \tilde{X}_T < b \text{ and if } X \neq \tilde{X} \end{cases}$$

subject to Condition (3.4). In other words, F stands for an insurance premium.[4] When $X_T < b$, $C_M(X) - (X_T - F)$ is the transfer to the insured who possesses $X_T - F$. By the incentive compatibility condition, $\tilde{u} = \hat{u}$, i.e., the insured reports optimally the optimal effort.

Second, I look at the implementability condition and the participation constraint. As usual in hidden action problems, the principal gives the smallest possible utility to the agent, i.e., the participation condition of the insured holds binding: $V_1 = r$. By the incentive compatibility

[4]For example, I can consider the case that the insured could not commit to the contract at time T, even although that case is out of the scope of this paper. The insurer could then force the insured to pay F at time 0 in order to avoid failure to collect F at time T.

condition and Eq.(3.3) in Corollary 3.1,

$$M_T^{\hat{u}} = e^{-r} e^{U_1(C_T)}. \tag{3.5}$$

With the two constraints, define the admissible set $\mathcal{A}_2$ of the contracts C_T, which is characterized by $\{F, b, C_M\}$ as:

Definition 3.4 *The admissible set $\mathcal{A}_2$ of the contracts C_T is the subset of $\mathcal{A}_2'$ consisting of the contracts satisfying*

(i) *C_M is continuous and non-decreasing in X, and satisfies $-\frac{U_1''}{U_1'} \leq U_1'$.*

(ii) *C_T satisfies Conditions* (3.4),(3.5)*.*

Note that, for some X_T, the contract may be infeasible. This case corresponds to a policy limit or a deductible in practice. Note also that I do not impose the assumption of positive consumption for either player. Negative consumption may take place in equilibrium, depending on the utility forms.

3.2.2 Optimal insurance design

Now, I formulate the insurer's optimization problem:

$$\sup_{C_T \in \mathcal{A}_2} E^{\hat{u}(C_T)} \left[U_2(\tilde{X}_T - C_T(X, \tilde{X})) - K_M \mathbf{1}_M \right]$$

where $\mathbf{1}_M$ is an indicator function such that $\mathbf{1}_M = 1$ when a monitoring action is undertaken (otherwise, 0). By Definition 3.3 (iii), the integrability is ensured. Define the Lagrangian multipliers associated with Conditions (3.4) and (3.5) as μ and λ, respectively. Due to $V_1 = r$, $\lambda > 0$. The constrained optimization problem is rewritten into:

$$\sup_{\{F,b,C_M\}} \left\{ \begin{array}{l} e^{-r} E \left[e^{U_1(C_T)} \left(\left(U_2(\tilde{X}_T - C_T(X, \tilde{X})) - K_M \mathbf{1}_M \right) + \lambda \right) \right] \\ +\mu \left[\left(C_M - (X_T - F) \right) (b - X_T) \right] \end{array} \right\}. \tag{3.6}$$

From this equation, the incentive-compatible compensation C_M should depend only on X_T, not on the trajectory X, i.e.,

$$C_M(X) = C_M(X_T).$$

With regard to C_M, a necessary condition for optimality is:

$$e^{U_1(C_M)}U_1'(C_M)\left\{(\lambda - K_M) - \left(\frac{U_2'(X_T - C_M)}{U_1'(C_M)} - U_2(X_T - C_M)\right)\right\} + e^r \mu\,(b - X_T) = 0. \quad (3.7)$$

On the assumption that $C_M(X_T)$ is continuously non-decreasing in X_T, by the incentive compatibility condition (3.4), if b is feasible,

$$C_M(b) = b - F. \quad (3.8)$$

Accordingly, when $X_T \geq b$, the incentive compatibility condition is necessarily slack. Therefore, I can focus attention on the case of $X_T < b$ in the second term on the left-hand side of Eq.(3.7). For notational convenience, define

$$\begin{aligned} D(y) &:= e^{U_1(y)}\Big(U_2(X_T - y) - K_M + \lambda\Big) + e^r \mu\,\Big(y - (X_T - F)\Big)\,(b - X_T)\,, \\ H(y) &:= \frac{U_2'(X_T - y)}{U_1'(y)} - U_2(X_T - y). \end{aligned}$$

The necessary condition (3.7) for optimality is then rewritten as:

$$D'(y) = e^{U_1(y)}U_1'(y)\,\Big((\lambda - K_M) - H(y)\Big) + e^r \mu\,(b - X_T) = 0.$$

Furthermore,

$$\begin{aligned} D''(y) &= e^{U_1(y)}U_1'(y)\left\{U_1'(y)\,\Big((\lambda - K_M) - H(y)\Big)\left(1 + \frac{U_1''(y)}{(U_1'(y))^2}\right) - H'(y)\right\} \\ &= e^{U_1(y)}U_1'(y)\left\{-\frac{\mu\,(b - X_T)\,e^{r - U_1(y)}}{U_1'(y)}\left(U_1'(y) + \frac{U_1''(y)}{U_1'(y)}\right) - H'(y)\right\}. \end{aligned}$$

For $C_T \in \mathcal{A}_2$, $-\frac{U_1''}{U_1'} \leq U_1'$. Noting $H'(y) = \frac{-U_2''(X_T - y)U_1'(y) - U_2'(X_T - y)U_1''(y)}{(U_1'(y))^2} + U_2'(X_T - y) > 0$,

$$D''(y) < 0.$$

Therefore, Eq.(3.7) is the necessary and sufficient condition for optimality.

For the reference, I examine the case that there exists moral hazard without ex-post informational asymmetry. It means that the incentive compatibility condition is slack, i.e., $\mu = 0$. The insurer does not need to monitor. Thus C_M can be rewritten as C_T. Hence,

$$\frac{U_2'(X_T - C_T)}{U_1'(C_T)} - U_2(X_T - C_T) = \lambda. \tag{3.9}$$

And,

$$0 < \frac{\mathrm{d}C_T}{\mathrm{d}X_T} = 1 - \frac{U_2'U_1''}{U_2''U_1' + U_2'U_1'' - U_2'(U_1')^2} < 1. \tag{3.10}$$

Furthermore, I examine the case that there are no moral hazard and no ex-post informational asymmetry. The standard Borch rule is then obtained:

$$\frac{U_2'(X_T - C_T)}{U_1'(C_T)} = \lambda, \tag{3.11}$$

$$0 \leq \frac{\mathrm{d}C_T}{\mathrm{d}X_T} = 1 - \frac{U_2'U_1''}{U_2''U_1' + U_2'U_1''} < 1. \tag{3.12}$$

Accordingly, from Eq.(3.9) and Eq.(3.11), we see that the term $U_2(X_T - C_T)$ stands for the effect of moral hazard, whereas the term $\mu\,(b - X_T)$ stands for the effect of ex-post informational asymmetry. From Eq.(3.10) and Eq.(3.12), $\frac{\mathrm{d}C_T}{\mathrm{d}X_T}$ is less than one, either with or without moral hazard. In addition, it is higher in Eq.(3.10) than in Eq.(3.12). I.e., when X_T gets higher, larger compensation is required in the moral hazard case due to the necessity to induce the insured to make the optimal efforts. The effect of moral hazard on C_T is represented by the difference between Eq.(3.10) and Eq.(3.12):

$$\begin{aligned} \Delta\left(\frac{\mathrm{d}C_T}{\mathrm{d}X_T}\right) &:= \left(1 - \frac{U_2'U_1''}{U_2''U_1' + U_2'U_1'' - U_2'(U_1')^2}\right) - \left(1 - \frac{U_2'U_1''}{U_2''U_1' + U_2'U_1''}\right) \\ &= \frac{-U_1''(U_1'U_2')^2}{(U_2''U_1' + U_2'U_1'')\,(U_2''U_1' + U_2'U_1'' - U_2'(U_1')^2)} > 0. \end{aligned} \tag{3.13}$$

3.2.3 Characterization

I characterize the optimal C_M from Eq.(3.7). Assume that, in the monitoring region, it holds true that

$$\frac{U_2'(X_T - C_M(X_T))}{U_1'(C_M(X_T))} - U_2(X_T - C_M(X_T)) = \lambda - K_M. \tag{3.14}$$

Noting strict concavity of U_1,

$$0 < \frac{\mathrm{d}C_M}{\mathrm{d}X_T} = 1 - \frac{U_2' U_1''}{U_2'' U_1' + U_2' U_1'' - U_2' (U_1')^2} < 1. \tag{3.15}$$

By Eq.(3.8) and Eq.(3.15), $C_M \geq X_T - F$ in the monitoring region, i.e, the insured is better off telling the truth when the payment rule C_M as well as (F, b) are given. Also, by construction, there is no informational asymmetry in the non-monitoring region. Hence, $\mu = 0$. Accordingly, if optimal efforts are attained as in Eq.(3.14), the problem of moral hazard and the problem of ex-post informational asymmetry are differentiated in the monitoring region.

Implications are as follows. C_T is non-linear in X_T. Similarly to Cvitanić and Zhang [5], this result is more general than Holmström and Milgrom [10] and Schättler and Sung [16]. Furthermore, this present model draws richer implications of insurance than Cvitanić and Zhang [5]. First, if optimal efforts are attained as in Eq.(3.14), the insurer can write explicitly the insurance contract that differentiates the two information problems, by using the costly monitoring effectively. In contrast, Cvitanić and Zhang (2007) face difficulty with writing the optimal contract that differentiates the informational problems, even by using dynamic data, because it is very hard to compute the Lagrangian multipliers associated with the informational problems.

Second, I look at dynamic interaction between the problem of moral hazard and the problem of the ex-post informational asymmetry in the optimal insurance contract. I focus on the case that b is feasible. Especially, I compare this model with the case of moral hazard without ex-post informational asymmetry characterized by Eq.(3.9). The ex-post informational asymmetry problem is reduced, i.e., the insured can still enjoy an information advantage while in good shape. On the other hand, in the monitoring region, there is no μ, i.e., informational asymmetry is removed there. Eq.(3.14) looks equivalent to Eq.(3.9) except for K_M. Let us examine the effect of the monitoring on moral hazard, which is measured by the effect on $\frac{\mathrm{d}C_T}{\mathrm{d}X_T}$, like $\Delta\left(\frac{\mathrm{d}C_T}{\mathrm{d}X_T}\right)$ in Eq.(3.13). From Eq.(3.15), when $U_2'' < 0$, then

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}K_M}\left(\frac{\mathrm{d}C_M}{\mathrm{d}X_T}\right) &= \frac{\mathrm{d}}{\mathrm{d}C_M}\left(\frac{\mathrm{d}C_M}{\mathrm{d}X_T}\right)\cdot\frac{\mathrm{d}C_M}{\mathrm{d}(\lambda-K_M)}\cdot\frac{\mathrm{d}(\lambda-K_M)}{\mathrm{d}K_M} \\
&= \underbrace{\frac{U_2'U_1'U_2''U_1''}{\left(U_2''U_1'+U_2'U_1''-U_2'(U_1')^2\right)^2}}_{\geq 0}\left(\begin{array}{l}\left(\left(-\frac{U_2'''}{U_2''}\right)-\left(-\frac{U_2''}{U_2'}\right)\right) \\ +\left(\left(-\frac{U_1'''}{U_1''}\right)-\left(-\frac{U_1''}{U_1'}\right)\right) \\ +\frac{U_1'}{\left(-\frac{U_2''}{U_2'}\right)}\left(\left(-\frac{U_1'''}{U_1''}\right)-2\left(-\frac{U_1''}{U_1'}\right)\right)\end{array}\right) \\
&\quad\cdot\underbrace{\frac{1}{\frac{-U_2''U_1'-U_2'U_1''}{(U_1')^2}+U_2'}}_{>0}\cdot\frac{\mathrm{d}(\lambda-K_M)}{\mathrm{d}K_M}. \qquad (3.16)
\end{aligned}$$

To make clear some economic implications of insurance, I add two definitions as follows. For $i \in \{1,2\}$, player i is said to be not downward-risk averse highly if $-\frac{U_i'''}{U_i''} \leq -\frac{U_i''}{U_i'}$, in that $\frac{\mathrm{d}\left(-\frac{U_i''(x)}{U_i'(x)}\right)}{\mathrm{d}x} = \frac{U_i''}{U_i'}\left(\left(-\frac{U_i'''}{U_i''}\right)-\left(-\frac{U_i''}{U_i'}\right)\right) \geq 0$. Note that $-\frac{U_1'''}{U_1''} \leq -\frac{U_1''}{U_1'} < -2\frac{U_1''}{U_1'}$. Also, I can guess that higher K_M leads to higher λ due to the tighter participation constraint. However, we are not sure if $\frac{\mathrm{d}\lambda}{\mathrm{d}K_M} \gtrless 1$. The participation constraint is said to be not too tight (with respect to K_M) if $\frac{\mathrm{d}\lambda}{\mathrm{d}K_M} < 1$.

From Eq.(3.16), when $U_2'' < 0$, if the insured and the insurer are not downward-risk averse highly and if the participation constraint is not too tight, then a higher level of the monitoring technology (i.e., lower K_M) mitigates the problem of moral hazard. Furthermore, regardless of the insurer's downward-risk aversion, if the insured is not downward-risk averse enough and if the participation constraint is not too tight, then a higher level of the monitoring technology (i.e., lower K_M) can mitigate the problem of moral hazard.

A logic behind this scene is as follows. When the insured is not downward-risk averse highly, he does not demand excessively high compensation. Also, the tighter participation constraint is not so tight as to negate the effect of higher K_M on the right-hand side of Eq.(3.14). Under such environments, the necessity to induce the insured to make the optimal efforts is reduced when the monitoring technology is available. With higher (lower) K_M, the monitoring technology is a less (more) useful devise to verify the truth. A lower (higher) level of the monitoring technology (i.e., higher (lower) K_M) then leads to more (less) moral hazard.

When $U_2''=0$, then

$$\frac{\mathrm{d}}{\mathrm{d}K_M}\left(\frac{\mathrm{d}C_M}{\mathrm{d}X_T}\right) = \underbrace{\frac{-\left(U_2'\right)^2\left(U_1'\right)^2 U_1''}{\left(U_2'U_1''-U_2'(U_1')^2\right)^2}}_{\geq 0}\left(\left(-\frac{U_1'''}{U_1''}\right)-2\left(-\frac{U_1''}{U_1'}\right)\right)\cdot$$
$$\underbrace{\frac{1}{\frac{-U_2'U_1''}{\left(U_1'\right)^2}+U_2'}}_{>0}\cdot\frac{\mathrm{d}(\lambda-K_M)}{\mathrm{d}K_M} \tag{3.17}$$

From this, when $U_2''=0$, if the insured is not downward-risk averse highly and if the participation constraint is not too tight, then a lower level of the monitoring technology (i.e., higher K_M) mitigates the problem of moral hazard.

4 Numerical analysis of optimal insurance design

4.1 Numerical method

So far I have not obtained explicitly the effect of K_M on λ, although I imposed the condition regarding the tightness of the participation constraint for examining the effect of the monitoring on the problem of moral hazard. In fact, λ is an endogenous variable. Also, I imposed other high-level assumptions on the admissible set of the controls. Thus we are not certain whether there exist plausible solutions to the above optimal insurance design problem, being given relevant structural parameters To complete the study of the optimal insurance design, I will do numerical analyses and draw quantitative implications in this section.

Based on the results in Section 3, a derivation method for the optimal values of $b, F, C_M, \hat{u}, \lambda$ consists of the following five steps.

(1) $\hat{u}$ is characterized by Eq.(3.5),

$$M_T^{\hat{u}} = e^{-r}e^{U_1(C_T)}.$$

(2) Assume that Eq.(3.14) holds true. $C_M(X_T)$ can then be written as a function of λ, denoted by $C_M^{\lambda}(X_T)$.

(3) By Eq.(3.8), if b is feasible,

$$C_M^{\lambda}(b) = b - F. \tag{4.1}$$

Thus F can be written as a function of b and λ, denoted by $F(b, \lambda)$.

(4) By Eq.(3.5), λ can be written as a function of b, denoted by $\lambda(b)$, satisfying:

$$\begin{aligned} e^r &= E\left[e^{U_1(C_T)}\right] \\ &= \int_b^{+\infty} e^{U_1(X_T - F(b,\lambda))}\Phi(\mathrm{d}X_T) + \int_{-\infty}^{b} e^{U_1(C_M^{\lambda}(X_T))}\Phi(\mathrm{d}X_T) \end{aligned} \tag{4.2}$$

where Φ denotes the cumulative distribution function of X_T under P at time 0. Following standard finance textbooks (e.g., Musiela and Rutkowski [13]), it holds true that $\Phi(x) = N\left(\frac{x}{v\sqrt{T}}\right)$ where $N(\cdot)$ denotes the standard normal cumulative distribution function, i.e., for $x \in \mathbf{R}$, $N(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u$.

(5) By Eq.(3.6), with respect to b, the insurer optimizes his expected utility:

$$\begin{aligned} & e^{-r}E\left[e^{U_1(C_T)}\Big((U_2(X_T - C_T) - K_M\mathbf{1}_M) + \lambda(b)\Big)\right] \\ = \; & e^{-r}\left[\begin{array}{l} \Big(U_2(F(b,\lambda(b))) + \lambda(b)\Big)\int_b^{+\infty} e^{U_1(X_T - F(b,\lambda(b)))}\Phi(\mathrm{d}X_T) \\ + \int_{-\infty}^{b} e^{U_1(C_M^{\lambda(b)}(X_T))}\Big(U_2(X_T - C_M^{\lambda(b)}(X_T)) - K_M + \lambda(b)\Big)\,\Phi(\mathrm{d}X_T). \end{array}\right] \end{aligned} \tag{4.3}$$

Recall that b may be infeasible, i.e., $b = -\infty$, $+\infty$, or $\varnothing$.

4.2 Numerical results

As an example, let us see the case that $U_1(x) = \log(x)$ and $U_2(x) = x$. In fact, the financial institution is thought to be more risk-tolerant than the insured firm. Note that $-\frac{U_1''(x)}{U_1'(x)} = \frac{1}{x} = U_1'(x)$. By Eq.(3.14),

$$C_M = \frac{X_T + (\lambda - K_M)}{2}.$$

For the positivity of C_M, $X_T > -(\lambda - K_M)$. By Eq.(4.1), if b is feasible,

$$F = \frac{b - (\lambda - K_M)}{2}.$$

Note that the insurer's consumption takes negative values due to his risk-neutral utility form, while the insured's one is positive due to his log utility. By Eq.(4.2),

$$\begin{aligned}
e^r &= \int_b^{+\infty} \left(x - \frac{b - (\lambda - K_M)}{2} \right) \Phi(\mathrm{d}x) + \int_{-(\lambda - K_M)}^b \left(\frac{x + (\lambda - K_M)}{2} \right) \Phi(\mathrm{d}x) \\
&= \int_b^{+\infty} \left(x - \frac{b}{2} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2v^2T}} \,\mathrm{d}x + \int_{-(\lambda - K_M)}^b \left(\frac{x}{2} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2v^2T}} \,\mathrm{d}x \\
&\quad + \frac{(\lambda - K_M)}{2} \left[1 - N \left(\frac{-(\lambda - K_M)}{v\sqrt{T}} \right) \right]
\end{aligned} \tag{4.4}$$

Finally, by Eq.(4.3), the insurer optimizes his utility with respect to b and λ,

$$\begin{aligned}
&(F + \lambda) \int_b^{+\infty} (x - F) \Phi(\mathrm{d}x) + \int_{-(\lambda - K_M)}^b C_M \,\Big((x - C_M) + (\lambda - K_M) \Big)\, \Phi(\mathrm{d}x) \\
= &\left(\frac{b + \lambda + K_M}{2} \right) \int_b^{+\infty} \left(x - \frac{b - (\lambda - K_M)}{2} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2v^2T}} \,\mathrm{d}x \\
&+ \int_{-(\lambda - K_M)}^b \left(\frac{x + (\lambda - K_M)}{2} \right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2v^2T}} \,\mathrm{d}x
\end{aligned}$$

subject to Eq.(4.4).

Finally, let us look at optimal insurance properties in the cases of $b = -\infty$ and $b = +\infty$. First, when $b = -\infty$, the optimal b should degenerate to F in the above optimization procedure because the payment rule is continuous. Hence, $F = -(\lambda - K_M)$. Note that, regardless of the convergence, I am using the notation $b = -\infty$ in order to stress that there is no trigger to the compensation in this case. From Eq.(4.4),

$$e^r = \int_{-(\lambda - K_M)}^{+\infty} (x + (\lambda - K_M))\, \Phi(\mathrm{d}x) = \int_0^{+\infty} z\, \Phi \,\Big(\mathrm{d}(z - (\lambda - K_M)) \Big) .$$

Therefore, so long as $b = -\infty$ holds, $F = -(\lambda - K_M)$ remains constant as K_M is changed. Next, when $b = +\infty$, from Eq.(4.4),

$$e^r = \int_{-(\lambda - K_M)}^{+\infty} \frac{x + (\lambda - K_M)}{2}\, \Phi(\mathrm{d}x) = \int_0^{+\infty} \frac{z}{2}\, \Phi \,\Big(\mathrm{d}(z - (\lambda - K_M)) \Big) .$$

Therefore, similarly, so long as $b = +\infty$ holds, $-(\lambda - K_M)$ remains constant as K_M is changed.

I set the cash-flow volatility at a conventional level $v = 25\%$, based on previous empirical results. With regard to the drift, Goldstein, Ju, and Leland [8] calibrate a slightly negative μ, whereas Leland [11] chooses $\mu = 1\%$. I am taking a zero drift under P as in the above model. Finally, with regard to the estimates of monitoring cost K_M (relative to r), there is a controversy in the previous empirical literature.[5] Thus I cover a wide range of K_M (relative to the insured's utility level r) in this present numerical analysis.[6] To determine the scale of this model, set $r = 1$. Parameterization and numerical results are shown in Table 1.

Table 1: Parameterization and numerical results

Parameters								
Cash flows: Diffusion	v	25%	25%	25%	25%	25%	25%	25%
Reservation utility	r	1	1	1	1	1	1	1
Monitoring cost	K_M	$0.01 \sim 0.9$	1.0	1.6	2.2	2.8	5.0	10.0
Optimal results								
Insurance premium	F	n.a.	-1.09	-0.95	-0.94	-0.93	-1.06	-1.06
Compensation trigger	b	$+\infty$	-0.17	-0.32	-0.36	-0.45	$-\infty$	$-\infty$
Lagrangian multiplier	λ	$5.45 \sim 6.34$	3.00	3.19	3.72	4.21	6.06	11.06
Deductible	$-(\lambda - K_M), F$	-5.44	-2.00	-1.59	-1.52	-1.41	-1.06	-1.06
Monitoring region	$b + (\lambda - K_M)$	> -5.44	1.83	1.27	1.16	0.96	0.00	0.00

When the monitoring cost K_M is an immediate level, the monitoring action is undertaken only for poor cash flows $-(\lambda - K_M) < X_T < b$. The contract is deductible when the time-T cash flow is less than $-(\lambda - K_M)$. This is a typical insurance contract with deductibles. Since $0 < \frac{d\lambda}{dK_M} < 1$, the participation constraint is not too tight. However, $-\frac{U_1'''}{U_1''} = -2\frac{U_1''}{U_1'}$, i.e., the insured is downward-risk averse highly. From Eq.(3.17), the problem of moral hazard is unchanged by a higher level of the monitoring technology (i.e., lower K_M). As K_M gets higher, the monitoring region gets smaller. Also, the insurance premium F (the compensation trigger b, respectively) is slightly increasing (decreasing) in K_M. A logic behind this result is as follows. When K_M becomes higher, the whole

[5]There are very few empirical studies of monitoring costs in insurance. Instead, for the reference, let us look at empirical studies of default costs, including monitoring costs, under debt contracts. Warner [17] estimates a bankruptcy cost at approximately $1.0\% \sim 5.3\%$ of firm value, by using the data of US railroad firms in 1933-1955. However, in his paper, the costs are only direct bankruptcy costs such as legal fees. Altman [2] estimates the sum of direct and indirect bankruptcy costs at about $11\% \sim 17\%$ of firm value. Thus there is no agreement upon the empirical values. Furthermore, the monitoring costs may be bigger lately under tighter financial regulation on structured finance than before.

[6]In this present model, monitoring costs K_M are exogenous. On the contrary, based on the numerical analysis, I could find optimal monitoring costs K_M.

pie to be shared becomes smaller. Thus the probability of monitoring is reduced. As discussed above, since the insured is not downward-risk averse highly, he does not demand excessively high compensation. Also, the tighter participation constraint is not so tight as to negate the effect of higher K_M on the right-hand side of Eq.(3.14). Under such environments, the necessity to induce the insured to make the optimal efforts is reduced due to the existence of K_M. With a higher cost K_M, the monitoring technology is a less useful devise to verify the truth. A lower level of the monitoring technology (i.e., higher K_M) then leads to larger moral hazard. To make up for the high monitoring cost, the insurer demands the high insurance premium F.

For a very low K_M (i.e., $0.01 \leq K_M < 1$), the insurer necessarily prefers the truth, even by incurring the monitoring cost (i.e., $b = +\infty$). The monitoring action is then undertaken for all feasible X_T, and the allocations are state-dependent. As mentioned above, when $b = +\infty$ holds, the deductible $-(\lambda - K_M)$ remains constant as K_M is changed. On the other hand, for a very high K_M, the monitoring action would shrink largely the whole income to be shared. The monitoring action is necessarily avoided in equilibrium. The optimal contract is then of a state-independent debt-type for all feasible X_T, i.e., $b = -\infty$. The insured is given no compensation for low X_T. Note that, by construction, in this case, F stands for a deductible threshold as well. Again, as mentioned above, when $b = -\infty$ holds, the insurance premium and the deductible remain constant as K_M is changed.

5 Conclusion

I found the properties of optimal insurance in the model of continuous-time costly monitoring under moral hazard and ex-post informational asymmetry. For future work, I will extend the model to have (1) risk pooling in a model with multi-insureds and (2) securitized insurance contracts. Also, I will do empirical work based on the above analytical results.

References

[1] Abbring, Jaap H., James J. Heckman, Pierre-André Chiappori, and Jean Pinquet (2003): "Adverse Selection and Moral Hazard In Insurance: Can Dynamic Data Help to Distinguish?" *Journal of the European Economic Association.*, 1, 512-521.

[2] Altman, Edward I. (1984): "A Further Empirical Investigation of the Bankruptcy Cost Question," *Journal of Finance*, 39, 1067-1089.

[3] Bolton, P., and M. Dewatripont (2005): *Contract Theory*, MIT Press: Cambridge, MA.

[4] Chiappori, Pierre-André (2000): "Econometric Models of Insurance under Asymmetric Information," in *Handbook of Insurance*, ed. by Georges Dionne. Kluwer Academic Publishers: Norwell, MA, 365-392.

[5] Cvitanić, Jakša, and Jianfeng Zhang (2007): "Optimal Compensation with Adverse Selection and Dynamic Actions," *Mathematics and Financial Economics*, 1, 21-55.

[6] Dionne, Georges, Florence Giuliano, and Pierre Picard (2009): "Optimal Auditing with Scoring: Theory and Application to Insurance Fraud," HEC Montreal, mimeo.

[7] Gale, Douglas, and Martin F. Hellwig (1985): "Incentive-Compatible Debt Contracts: The One-Period Problem, " *Review of Economic Studies*, 52, 647-663.

[8] Goldstein R., N. Ju, and H. Leland (2001): "An EBIT-Based Model of Dynamic Capital Structure," *Journal of Business*, 74, 483-512.

[9] Harrington, Scott E., and Gregory R. Niehaus (2004): *Risk Management and Insurance*, Second Edition, McGraw-Hill/Irwin: NY.

[10] Holmström, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303-328.

[11] Leland, Haynes E. (1998): "Agency Costs, Risk Management, and Capital Structure," *Journal of Finance*, 53, 1213-1243.

[12] MacMinn, Ricard, and James Garven (2000): "On Corporate Insurance," in *Handbook of Insurance*, ed. by Georges Dionne. Kluwer Academic Publishers: Norwell, MA, 541-564.

[13] Musiela, Marek, and Marek Rutkowski (1997): *Martingale Methods in Financial Modelling*, New York, Springer-Verla.

[14] Picard, Pierre (2000): "Economic Analysis of Insurance Fraud," in *Handbook of Insurance*, ed. by Georges Dionne. Kluwer Academic Publishers: Norwell, MA, 315-362.

[15] Rogerson, William (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53, 1357-1368.

[16] Schättler, H., and J. Sung (1993): "The First-Order Approach to the Continuous-Time Principal-Agent Problem with Exponential Utility", *Journal of Economic Theory*, 61, 331-371.

[17] Warner, Jerold B. (1977): "Bankruptcy Costs: Some Evidence," *Journal of Finance*, Papers and Proceedings of the Thirty-Fifth Annual Meeting of the American Finance Association, Atlantic City, New Jersey, September 16-18, 32, 337-347.

Appendix

Upon request.